

# UNIVERSIDAD PÚBLICA DE EL ALTO

## CARRERA INGENIERÍA DE SISTEMAS



## TESIS DE GRADO

### “MODELO PREDICTIVO DE ÍNDICES DELICTIVOS BASADO EN MINERÍA DE DATOS APLICADO EN EL MUNICIPIO DE EL ALTO”

Para Optar al Título de Licenciatura en Ingeniería de Sistemas  
MENCIÓN: INFORMÁTICA Y COMUNICACIONES

Postulante: Adan Eddy Paredes Barrionuevo

Tutor Metodológico: M. Sc. Lic. Ing. Fanny Helen Perez Mamani

Tutor Especialista: Lic. Ing. Sergio Ramiro Rojas Saire

Tutor Revisor: Lic. Ing. Carmiña Argani Gutiérrez

EL ALTO - BOLIVIA

2024

# DECLARACIÓN JURADA DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, **Adan Eddy Paredes Barrionuevo** estudiante con **C.I. 8362144LP** mediante la presente **declaro** de manera pública que la propuesta del **TESIS DE GRADO** titulada **“MODELO PREDICTIVO DE ÍNDICES DELICTIVOS BASADO EN MINERÍA DE DATOS APLICADO EN EL MUNICIPIO DE EL ALTO”** es original, siendo resultado de mi trabajo personal y no constituye una copia o replica de trabajos similares elaborados.

Autorizo la publicación del resumen de mi propuesta en internet y me comprometo a responder a todos los cuestionamientos que se desprenden de su lectura.

Asimismo, me hago responsable ante la universidad o terceros, de cualquiera irregularidad o daño que pudiera ocasionar, por el incumplimiento de lo declarado.

De identificarse falsificación, plagio, fraude, o que la **TESIS DE GRADO** haya sido publicado anteriormente; asumo las consecuencias y sanciones que de mi acción se deriven, responsabilizándome por todas las cargas legales que se deriven de ello sometiéndome a las normas establecidas y vigentes de la Carrera de Ingeniería de Sistemas de la Universidad Pública de El Alto.

El Alto, junio del 2024.

**Adan Eddy Paredes Barrionuevo**  
**C.I. 8362144Lp**  
**paredes.barrionuevo.adan.eddy@gmail.com**

**DEDICATORIA**

*A Dios por haberme permitido alcanzar este hito en mi vida. Por guiarme por el camino correcto, por bendecirme con salud y fortaleza, y por infundirme la fuerza para seguir adelante.*

*A mis amados padres.*

*Felipe Paredes por su apoyo incondicional, por su paciencia y comprensión.*

*A mi madre Agustina Barrionuevo por brindarme esa gran motivación, por su fe inquebrantable en mí, por impulsarme a nunca rendirme.*

*A mis queridos hermanos Wily y Yhon, por su motivación y apoyo incondicional.*

*Adan Eddy Paredes Barrionuevo*

## **AGRADECIMIENTO**

*Agradezco a Dios por guiarme en el mi camino correcto, brindarme Salud, y sabiduría por permitirme estar presente día a día.*

*A mi Padre Felipe Paredes, por todo su apoyo y comprensión, por haberme apoyado en el transcurso de mi vida gracias por siempre creer en mí.*

*A mi Madre Agustina Barrionuevo por todo su apoyo incondicional, por el sacrificio de haberme brindado la educación y velar por mi salud, gracias por ser mi ángel guardián y por enseñarme el verdadero significado del amor incondicional.*

*A mi tutor Ing. Sergio Rojas Saire por compartir sus conocimientos y el apoyo que fueron importantes para la culminación de este trabajo.*

*A mi tutor Ing. Carmiña Argani Gutiérrez por su gran apoyo en mi camino profesional y colaboración en el cumplimiento de mi meta.*

*A mi tutor Ing. Fanny Helen Perez Mamani, por haberme brindado sus conocimientos y ser la guía que me complementaron para culminar el presente trabajo.*

## ÍNDICE GENERAL

|                                                    | Página |
|----------------------------------------------------|--------|
| CAPÍTULO I .....                                   | 1      |
| 1.1 MARCO PRELIMINAR .....                         | 1      |
| 1.2 INTRODUCCIÓN.....                              | 1      |
| 1.3 ANTECEDENTES .....                             | 2      |
| 1.3.1 Antecedentes Afines a la Investigación ..... | 2      |
| 1.3.1.1 Antecedentes Internacionales.....          | 2      |
| 1.3.1.2 Antecedentes Nacionales.....               | 3      |
| 1.3.1.3 Antecedentes Locales .....                 | 3      |
| 1.4 PLANTEAMIENTO DEL PROBLEMA.....                | 4      |
| 1.4.1 Problema Principal .....                     | 4      |
| 1.4.2 Problemas Secundarios .....                  | 5      |
| 1.4.3 Formulación del Problema .....               | 5      |
| 1.5 OBJETIVOS.....                                 | 5      |
| 1.5.1 Objetivo General.....                        | 5      |
| 1.5.2 Objetivos Específicos .....                  | 5      |
| 1.6 HIPÓTESIS.....                                 | 6      |
| 1.6.1 Identificación de Variables .....            | 6      |
| 1.6.2 Conceptualización de Variables.....          | 7      |
| 1.6.3 Operacionalización de Variables.....         | 7      |
| 1.7 JUSTIFICACIÓN .....                            | 8      |
| 1.7.1 Justificación Científica .....               | 8      |
| 1.7.2 Justificación Técnica .....                  | 8      |
| 1.7.3 Justificación Económica .....                | 8      |
| 1.7.4 Justificación Social .....                   | 9      |
| 1.8 METODOLOGÍA.....                               | 9      |
| 1.8.1 Método Científico.....                       | 9      |
| 1.8.2 Técnicas de Investigación .....              | 10     |
| 1.8.2.1 Investigación Cuantitativa .....           | 10     |
| 1.8.3 Metodología CRISP-DM.....                    | 11     |
| 1.8.4 Metodología SEMMA.....                       | 13     |
| 1.8.5 Metodología de desarrollo SCRUM .....        | 14     |
| 1.8.6 Método de Ingeniería.....                    | 14     |

|         |                                                    |    |
|---------|----------------------------------------------------|----|
| 1.8.6.1 | Métrica de Calidad.....                            | 14 |
| 1.8.6.2 | Estimación de Costos.....                          | 14 |
| 1.8.6.3 | Seguridad de La Información .....                  | 15 |
| 1.8.7   | Pruebas de Hipótesis .....                         | 16 |
| 1.8.7.1 | T-STUDENT .....                                    | 16 |
| 1.9     | HERRAMIENTAS .....                                 | 16 |
| 1.9.1   | Hardware .....                                     | 17 |
| 1.9.2   | Software.....                                      | 17 |
| 1.10    | LÍMITES Y ALCANCES.....                            | 18 |
| 1.10.1  | Límites .....                                      | 18 |
| 1.10.2  | Alcances .....                                     | 18 |
| 1.11    | Aportes.....                                       | 19 |
| 2       | CAPÍTULO II.....                                   | 20 |
| 2.1     | MARCO TEÓRICO.....                                 | 20 |
| 2.2     | DATO.....                                          | 20 |
| 2.3     | INFORMACIÓN .....                                  | 20 |
| 2.4     | CONOCIMIENTO .....                                 | 20 |
| 2.5     | MINERÍA DE DATOS .....                             | 21 |
| 2.5.1   | Historia.....                                      | 22 |
| 2.5.2   | Definición de la Minería de Datos .....            | 24 |
| 2.5.3   | Modelos de Minería de Datos .....                  | 27 |
| 2.5.3.1 | Modelos Descriptivos .....                         | 28 |
| 2.5.3.2 | Modelos Predictivos .....                          | 28 |
| 2.6     | ETAPAS DE LA MINERÍA DE DATOS .....                | 29 |
| 2.6.1   | Proceso de Extracción de Conocimiento .....        | 31 |
| 2.6.1.1 | Integración y Recopilación .....                   | 32 |
| 2.6.1.2 | Selección, Limpieza y Transformación .....         | 34 |
| 2.6.1.3 | Minería de Datos .....                             | 37 |
| 2.6.1.4 | Interpretación y Evaluación .....                  | 37 |
| 2.6.1.5 | Difusión, Uso y Monitorización.....                | 38 |
| 2.7     | MÉTODOS Y TÉCNICAS DE LA MINERÍA DE DATOS .....    | 39 |
| 2.7.1   | Métodos de la Minería de Datos .....               | 39 |
| 2.7.1.1 | Agrupamiento ("Clustering").....                   | 40 |
| 2.7.1.2 | Asociación (" Association Pattern Discovery")..... | 40 |

|          |                                                                                                             |    |
|----------|-------------------------------------------------------------------------------------------------------------|----|
| 2.7.1.3  | Secuenciamiento ("Sequential Pattern Discovery").....                                                       | 41 |
| 2.7.1.4  | Reconocimiento de Patrones ("Pattern Matching") .....                                                       | 41 |
| 2.7.1.5  | Previsión ("Forecasting").....                                                                              | 41 |
| 2.7.1.6  | Simulación .....                                                                                            | 42 |
| 2.7.1.7  | Optimización .....                                                                                          | 42 |
| 2.7.1.8  | Clasificación ("Classification", "Prediction" o "Scoring") .....                                            | 43 |
| 2.7.2    | Técnicas de la Minería de Datos.....                                                                        | 44 |
| 2.7.2.1  | Métodos Estadísticos .....                                                                                  | 46 |
| 2.7.2.2  | Métodos Basados en Árboles de Decisión .....                                                                | 48 |
| 2.7.2.3  | Reglas de Asociación.....                                                                                   | 48 |
| 2.7.2.4  | Redes Neuronales ("Neural Networks").....                                                                   | 49 |
| 2.7.2.5  | Algoritmos Genéticos ("Genetic Algorithms") .....                                                           | 50 |
| 2.7.2.6  | Lógica Difusa ("fuzzy logic").....                                                                          | 51 |
| 2.7.2.7  | Series Temporales .....                                                                                     | 51 |
| 2.7.2.8  | Redes Bayesianas.....                                                                                       | 51 |
| 2.7.2.9  | Inducción de Reglas.....                                                                                    | 52 |
| 2.7.2.10 | Sistemas Basados en el Conocimiento y Sistemas Expertos ("Knowledge Based Systems" & "Expert Systems")..... | 53 |
| 2.7.2.11 | Algoritmos Matemáticos .....                                                                                | 53 |
| 2.8      | APLICACIONES DE LA MINERÍA DE DATOS .....                                                                   | 53 |
| 2.9      | HERRAMIENTAS DE LA MINERÍA DE DATOS .....                                                                   | 55 |
| 2.9.1    | Weka.....                                                                                                   | 56 |
| 2.9.1.1  | Características de Weka .....                                                                               | 57 |
| 2.9.2    | Rapid Miner .....                                                                                           | 58 |
| 2.9.2.1  | Características de Rapid Miner.....                                                                         | 59 |
| 2.9.3    | R Development.....                                                                                          | 59 |
| 2.9.3.1  | Características del Lenguaje R .....                                                                        | 60 |
| 2.9.4    | Orange Data Minig .....                                                                                     | 61 |
| 2.9.5    | Knime.....                                                                                                  | 62 |
| 2.9.6    | Python.....                                                                                                 | 62 |
| 2.10     | MÉTODO CIENTÍFICO .....                                                                                     | 64 |
| 2.11     | METODOLOGÍAS PARA MINERÍA DE DATOS .....                                                                    | 67 |
| 2.11.1   | KDD .....                                                                                                   | 68 |
| 2.11.2   | SEMMA.....                                                                                                  | 71 |

|          |                                                             |     |
|----------|-------------------------------------------------------------|-----|
| 2.11.3   | CRISP-DM .....                                              | 73  |
| 2.12     | INGENIERÍA DE SOFTWARE .....                                | 77  |
| 2.12.1   | Historia de la Ingeniería de Software .....                 | 80  |
| 2.12.2   | Procesos de Software .....                                  | 80  |
| 2.13     | METODOLOGÍA SCRUM .....                                     | 82  |
| 2.14     | ÍNDICES DELICTIVOS.....                                     | 85  |
| 2.14.1   | Causas.....                                                 | 90  |
| 2.14.1.1 | La Familia .....                                            | 90  |
| 2.14.1.2 | Las Condiciones Socioeconómicas .....                       | 90  |
| 2.14.1.3 | Influencias Psicosociales .....                             | 91  |
| 2.14.1.4 | Influencia Cultural.....                                    | 92  |
| 2.14.1.5 | Corrupción y el Delito en las Instituciones .....           | 92  |
| 2.14.2   | Efectos .....                                               | 93  |
| 2.14.2.1 | Daños Físicos y Sicológicos a las Víctimas.....             | 93  |
| 2.14.2.2 | Pérdidas Económicas.....                                    | 94  |
| 2.14.2.3 | Deterioro del Tejido Social .....                           | 94  |
| 2.14.2.4 | Impacto en el Sistema Judicial.....                         | 94  |
| 2.14.2.5 | Estigmatización de las Víctimas y los grupos Sociales ..... | 95  |
| 2.15     | ÍNDICES DELICTIVOS EN EL MUNICIPIO DE EL ALTO .....         | 95  |
| 2.15.1   | Posibles factores delictivos .....                          | 96  |
| 2.15.2   | Datos sobre los Índices Delictivos .....                    | 99  |
| 2.16     | MÉTRICAS DE CALIDAD .....                                   | 100 |
| 2.16.1   | ISO 25010.....                                              | 100 |
| 2.17     | EVALUACIÓN DE COSTOS COCOMO II.....                         | 105 |
| 2.17.1   | Características.....                                        | 107 |
| 2.17.2   | Modelos de Estimación .....                                 | 108 |
| 2.17.3   | Modelos Básicos .....                                       | 111 |
| 2.17.4   | Modelos Intermedios .....                                   | 111 |
| 2.17.5   | Modelos Detallados .....                                    | 114 |
| 3        | CAPÍTULO III.....                                           | 116 |
| 3.1      | DISEÑO METODOLÓGICO.....                                    | 116 |
| 3.1.1    | Tipo de Investigación.....                                  | 116 |
| 3.1.1.1  | Investigación Correlacional .....                           | 116 |
| 3.1.2    | Diseño de la Investigación .....                            | 118 |



|         |                                                                  |     |
|---------|------------------------------------------------------------------|-----|
| 3.1.2.1 | Diseño no Experimental .....                                     | 118 |
| 3.1.2.2 | Tipos de Diseños no Experimentales .....                         | 118 |
| 3.1.3   | Variables de la Investigación.....                               | 120 |
| 3.1.4   | Ambiente de la Investigación .....                               | 120 |
| 3.1.4.1 | Muestreo .....                                                   | 121 |
| 3.1.5   | Descripción de la Metodología a usar.....                        | 121 |
| 3.1.5.1 | Método Científico.....                                           | 121 |
| 3.2     | HERRAMIENTAS .....                                               | 123 |
| 3.2.1   | Python.....                                                      | 123 |
| 3.2.2   | Postgres.....                                                    | 124 |
| 3.2.3   | Visual Studio Code .....                                         | 125 |
| 3.3     | HERRAMIENTAS A USAR .....                                        | 126 |
| 3.3.1   | Técnicas de Investigación e Instrumentos.....                    | 126 |
| 3.3.1.1 | Técnica de Investigación.....                                    | 126 |
| 3.3.2   | Instrumentos de investigación.....                               | 126 |
| 3.4     | APLICACIÓN DE CRISP-DM.....                                      | 128 |
| 3.4.1   | FASE I: Comprensión del negocio .....                            | 129 |
| 3.4.2   | FASE II: Comprensión de los datos .....                          | 129 |
| 3.4.2.1 | Recolección de los datos iniciales .....                         | 129 |
| 3.4.2.2 | Datos recolectados.....                                          | 129 |
| 3.4.2.3 | Descripción de los datos .....                                   | 130 |
| 3.4.3   | FASE III: Preparación de los datos .....                         | 131 |
| 3.4.3.1 | Importando a la Base de Datos.....                               | 133 |
| 3.4.4   | FASE IV: Modelado .....                                          | 133 |
| 3.4.5   | FASE V: Evaluación .....                                         | 133 |
| 3.5     | APLICANDO LA METODOLOGIA SCRUM.....                              | 135 |
| 3.5.1   | Planificación de la iteración (Sprint Planning) .....            | 135 |
| 3.5.2   | Ejecución de la iteración (Sprint) .....                         | 136 |
| 3.5.3   | Reunión diaria de sincronización del equipo (Daily meeting)..... | 136 |
| 3.5.4   | Demostración de requisitos completados (Sprint Review) .....     | 137 |
| 3.6     | METRICAS DE CALIDA.....                                          | 137 |
| 3.6.1   | Funcionalidad .....                                              | 137 |
| 3.6.2   | Usabilidad.....                                                  | 138 |
| 3.6.3   | Confiableidad.....                                               | 139 |

|         |                                     |     |
|---------|-------------------------------------|-----|
| 3.6.4   | Eficiencia.....                     | 141 |
| 3.6.5   | Mantenibilidad.....                 | 142 |
| 3.6.6   | Portabilidad.....                   | 143 |
| 3.7     | EVALUACION DE COSTOS.....           | 143 |
| 3.7.1   | Adecuación Funcional.....           | 143 |
| 3.7.2   | Aplicación de COCOMO II.....        | 148 |
| 3.7.2.1 | Costos del sistema.....             | 150 |
| 4       | CAPÍTULO IV.....                    | 153 |
| 4.1     | PRUEBAS Y RESULTADOS.....           | 153 |
| 4.2     | PRESENTACIÓN DEL MODELO.....        | 153 |
| 4.3     | DESARROLLO DEL MODELO.....          | 153 |
| 4.4     | DEMOSTRACIÓN DEL PROTOTIPO.....     | 154 |
| 4.5     | PRUEBAS DE HIPÓTESIS.....           | 164 |
| 5       | CAPÍTULO V.....                     | 169 |
| 5.1     | CONCLUSIONES Y RECOMENDACIONES..... | 169 |
| 5.1.1   | CONCLUSIONES.....                   | 169 |
| 5.1.2   | RECOMENDACIONES.....                | 170 |
|         | BIBLIOGRAFÍA.....                   | 180 |

## ÍNDICE DE TABLAS

|                                                                                 |     |
|---------------------------------------------------------------------------------|-----|
| <i>Tabla 1. Conceptualización de variables</i> .....                            | 7   |
| <i>Tabla 2. Operacionalización de variables</i> .....                           | 7   |
| <i>Tabla 3. Evolución de las tecnologías relacionadas con Data Mining</i> ..... | 24  |
| <i>Tabla 4. Clasificación de las técnicas de Minería de Datos</i> .....         | 45  |
| <i>Tabla 5. Técnicas del Data Mining</i> .....                                  | 46  |
| <i>Tabla 6. Presencia de Factores de Riesgo</i> .....                           | 87  |
| <i>Tabla 7. Modelo Básico</i> .....                                             | 111 |
| <i>Tabla 8. Modelo Intermedio</i> .....                                         | 114 |
| <i>Tabla 9. Datos de la Tabla</i> .....                                         | 131 |
| <i>Tabla 10. Total, datos delictivos</i> .....                                  | 132 |
| <i>Tabla 11. Sprint Planning</i> .....                                          | 135 |
| <i>Tabla 12. Puntuación de la Funcionalidad</i> .....                           | 138 |
| <i>Tabla 13. Cálculo de la Usabilidad</i> .....                                 | 138 |
| <i>Tabla 14. Evaluación de la Eficiencia</i> .....                              | 141 |
| <i>Tabla 15. Cálculo de Adecuación Funcional</i> .....                          | 144 |
| <i>Tabla 16. Cálculo de cuenta total</i> .....                                  | 145 |
| <i>Tabla 17. Cálculo factor de ajuste de complejidad</i> .....                  | 146 |
| <i>Tabla 18. Factor LCD/PF de lenguaje de Programación</i> .....                | 149 |
| <i>Tabla 19. Constantes A, B, C, D COCOMO</i> .....                             | 150 |
| <i>Tabla 20. Costo total del Modelo</i> .....                                   | 152 |
| <i>Tabla 21. Comparación Correlacional</i> .....                                | 164 |
| <i>Tabla 22. Índices Pasados, índices Futuros</i> .....                         | 165 |
| <i>Tabla 23. Media, Desviación estándar</i> .....                               | 166 |

## ÍNDICE DE FIGURAS

|                                                                                                                          |            |
|--------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Figura 1. La Minería de Datos frente al KD y KDD.....</i>                                                             | <i>27</i>  |
| <i>Figura 2. Proceso del KDD.....</i>                                                                                    | <i>30</i>  |
| <i>Figura 3. Etapas del proceso KDD.....</i>                                                                             | <i>31</i>  |
| <i>Figura 4. Integración en un almacén de datos .....</i>                                                                | <i>33</i>  |
| <i>Figura 5. Discretización del atributo.....</i>                                                                        | <i>36</i>  |
| <i>Figura 6. Herramientas más usadas de la Minería de Datos .....</i>                                                    | <i>55</i>  |
| <i>Figura 7. Proceso KDD.....</i>                                                                                        | <i>69</i>  |
| <i>Figura 8. Metodología SEMMA .....</i>                                                                                 | <i>72</i>  |
| <i>Figura 9. Fases en SEMMA .....</i>                                                                                    | <i>73</i>  |
| <i>Figura 10. Metodología CRISP-DM .....</i>                                                                             | <i>75</i>  |
| <i>Figura 11. Capas de la Ingeniería de Software .....</i>                                                               | <i>78</i>  |
| <i>Figura 12. Denuncias de Atracos.....</i>                                                                              | <i>88</i>  |
| <i>Figura 13. Problemas que posee Bolivia .....</i>                                                                      | <i>89</i>  |
| <i>Figura 14. Inseguridad en horarios nocturnos.....</i>                                                                 | <i>90</i>  |
| <i>Figura 15. Denuncias de delitos.....</i>                                                                              | <i>99</i>  |
| <i>Figura 16. Denuncias de Delitos.....</i>                                                                              | <i>100</i> |
| <i>Figura 17. Estructura de la ISO 25010Nota. Diseño de la estructura de la ISO 25010<br/>(Moreno, 2018, p. 20).....</i> | <i>101</i> |
| <i>Figura 18. Diseño no experimental.....</i>                                                                            | <i>119</i> |
| <i>Figura 19. Metodología CRISP-DM .....</i>                                                                             | <i>128</i> |
| <i>Figura 20. Datos de hechos Delictivos .....</i>                                                                       | <i>130</i> |
| <i>Figura 21. Base de Datos PostgreSQL.....</i>                                                                          | <i>133</i> |
| <i>Figura 22. Resultado de Arima .....</i>                                                                               | <i>134</i> |
| <i>Figura 23. Interfaz principal.....</i>                                                                                | <i>153</i> |

|                                                                                              |            |
|----------------------------------------------------------------------------------------------|------------|
| <i>Figura 24. Porcentaje de índices delictivos Gestión 2017 .....</i>                        | <i>154</i> |
| <i>Figura 25. Porcentaje de índices delictivos Gestión 2018 .....</i>                        | <i>155</i> |
| <i>Figura 26. Porcentaje de índices delictivos Gestión 2019 .....</i>                        | <i>155</i> |
| <i>Figura 27. Porcentaje de índices delictivos Gestión 2020 .....</i>                        | <i>156</i> |
| <i>Figura 28. Porcentaje de índices delictivos Gestión 2021 .....</i>                        | <i>156</i> |
| <i>Figura 29. Porcentaje de índices delictivos Gestión 2022 .....</i>                        | <i>157</i> |
| <i>Figura 30. Porcentaje de índices delictivos Gestión 2023 .....</i>                        | <i>157</i> |
| <i>Figura 31. Porcentaje de índices delictivos Gestiones 2017 al 2023.....</i>               | <i>158</i> |
| <i>Figura 32. Índices delictivos altos y bajos por distritos gestión 2017.....</i>           | <i>158</i> |
| <i>Figura 33. Índices delictivos altos y bajos por distritos gestión 2018.....</i>           | <i>159</i> |
| <i>Figura 34. Índices delictivos altos y bajos por distritos gestión 2019.....</i>           | <i>159</i> |
| <i>Figura 35. Índices delictivos altos y bajos por distritos gestión 2020.....</i>           | <i>160</i> |
| <i>Figura 36. Índices delictivos altos y bajos por distritos gestión 2021.....</i>           | <i>160</i> |
| <i>Figura 37. Índices delictivos altos y bajos por distritos gestión 2022.....</i>           | <i>161</i> |
| <i>Figura 38. Índices delictivos altos y bajos por distritos gestión 2023.....</i>           | <i>161</i> |
| <i>Figura 39. Índices delictivos altos y bajos por distritos gestiones 2017 al 2023.....</i> | <i>162</i> |
| <i>Figura 40. Información del Modelo Arima .....</i>                                         | <i>162</i> |
| <i>Figura 41. Resultados obtenidos .....</i>                                                 | <i>163</i> |
| <i>Figura 42. Predicción de 6 años futuros .....</i>                                         | <i>163</i> |

**ÍNDICE DE ECUACIONES**

|                                               |     |
|-----------------------------------------------|-----|
| ( 1 ) <i>Porcentaje de fallas</i> .....       | 140 |
| ( 2 ) <i>Confiabilidad del sistema</i> .....  | 141 |
| ( 3 ) <i>Valor de la eficiencia</i> .....     | 142 |
| ( 4 ) <i>Valor de la mantenibilidad</i> ..... | 143 |
| ( 5 ) <i>Valor de la portabilidad</i> .....   | 143 |
| ( 6 ) <i>Punto de función</i> .....           | 148 |
| ( 7 ) <i>Kilo líneas de código</i> .....      | 151 |
| ( 8 ) <i>Esfuerzo requerido</i> .....         | 151 |
| ( 9 ) <i>Tiempo requerido</i> .....           | 151 |
| ( 10 ) <i>Cálculo de programadores</i> .....  | 151 |
| ( 11 ) <i>Estimación de salario</i> .....     | 152 |
| ( 12 ) <i>Varianza común</i> .....            | 166 |
| ( 13 ) <i>Estadístico de prueba</i> .....     | 167 |
| ( 14 ) <i>Valor crítico</i> .....             | 167 |

## RESUMEN

En el Municipio de El Alto existe una gran inquietud constante acerca de los hechos delictivos, que es preocupante para todo el Municipio en general, no solo a cada distrito, sino a todos los distritos que conforman este Municipio de El Alto y su inexistencia sobre el incremento en el índice delictivo. En respuesta al desconocimiento acerca del índice delictivo, la presente investigación tiene como objetivo principal “Diseñar un modelo predictivo de índices delictivos basado en Minería de Datos, aplicado al Municipio de El Alto, utilizando datos recolectados de las diferentes instituciones relacionadas con la temática de Seguridad, con el fin de anticipar el incremento de los índices delictivos”, para hacer posible el objetivo propuesto se realizó la recolección de información sobre los hechos delictivos y mediante técnicas y herramientas de minería de datos se diseñó un modelo con el fin de proyectar el índice de crecimiento. Las pruebas se realizaron con los datos obtenidos que nos proporcionó el Comando General de la Policía Boliviana. Se utilizó una metodología para la predicción de datos, la metodología CRISP-DM que estructura el proceso de minería de datos, se utilizó algoritmos de series de tiempo, el modelo Arima. La presente Investigación contribuye a la toma de decisiones por parte de las autoridades encargadas de seguridad ciudadana, directamente al Comando General de la Policía Boliviana, y de esa manera lograr tener acciones en contra de los distritos que conforman El Municipio de El Alto, para contribución con la población que conforman este Municipio.

**Palabras clave:** CRISP-DM, Índices delictivos, Minería de Datos, Predicción.

## ABSTRACT

In the Municipality of El Alto there is a great constant concern about criminal acts, which is worrying for the entire Municipality in general, not only each district, but all the districts that make up this Municipality of El Alto and its nonexistence on the increase in the crime rate. In response to the lack of knowledge about the crime rate, the main objective of this research is "Design a predictive model of crime rates based on Data Mining, applied to the Municipality of El Alto, using data collected from different institutions related to the topic of Security. , in order to anticipate the increase in crime rates", to make the proposed objective possible, information on criminal events was collected and, using data mining techniques and tools, a model was designed in order to project the rate. of growth. The tests were carried out with the data obtained provided to us by the General Command of the Bolivian Police. A methodology was used for data prediction, the CRISP-DM methodology that structures the data mining process, time series algorithms, the Arima model, were used. This Investigation contributes to decision-making by the authorities in charge of citizen security, directly to the General Command of the Bolivian Police, and in this way achieve actions against the districts that make up the Municipality of El Alto, to contribute with the population that makes up this Municipality.

**Keywords:** CRISP-DM, Crime rates, Data Mining, Prediction.



## GLOSARIO DE ABREVIATURAS

**KDD:** Knowledge Discovery in Databases (Descubrimiento de Conocimiento en bases de datos).

**SEMMA:** Sample, Explore, Modify, Model and Assess (Muestrear, Explorar, Modificar, Modelar y evaluar).

**CRISP-DM:** Cross Industry Standard Process for Data Mining (Modelo de Proceso Estándar para Minería de Datos)

**KNIME:** Konstanz Information Miner (Plataforma analítica de desarrollo de modelos).

**WEKA:** Waikato Environment Knowledge Analysis (Entorno de Análisis De Conocimiento de la Universidad de Waikato).

**DM:** Minería de datos.



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# **CAPÍTULO I**

## MARCO PRELIMINAR



## CAPÍTULO I

### 1.1 MARCO PRELIMINAR

### 1.2 INTRODUCCIÓN

La definición de Minería de Datos puede variar entre los diferentes investigadores, ya sean estadísticos, analistas de datos, La Minería de Datos es el análisis de archivos y bitácoras de transacciones, trabaja a nivel del conocimiento con el fin de descubrir patrones, relaciones, reglas, asociaciones o incluso excepciones útiles para la toma de decisiones (Suarez Yuniel & Diaz Amador, 2009).

El uso de la Minería de Datos o Data Mining, como soporte a las decisiones en las actividades de negocio, requiere mucho más que la aplicación de sofisticadas técnicas como redes neuronales o árboles de decisión sobre las tablas de datos. Por un lado, como uno de los pasos del proceso de descubrimiento de conocimiento, se utilizan como apoyo, técnicas relacionadas con la estadística, el reconocimiento de patrones y algoritmos de aprendizaje, entre otras.

Todos estos estudios han incrementado el deseo desenfrenado por demandar un mayor control de los procesos u operaciones y servicios, visto como núcleo de una gestión global, fundamental para proporcionar servicios de calidad y lograr un rendimiento óptimo de las inversiones (Aular y otros, 2007).

En el Municipio de El Alto, hoy en día los hechos delictivos están en todas partes. Al ser una ciudad con gran población, los delitos de igual manera van creciendo día tras día, por lo cual, con ayuda de la Minería de Datos, se quiere lograr identificar cuál es el índice de crecimiento en los distintos distritos del Municipio de El Alto, para brindar el conocimiento a la población acerca del incremento de hechos delictivos con ayuda de herramientas que puedan facilitar el análisis y proyección de los datos.

Esta investigación propone demostrar mediante la recolección de datos usando las distintas herramientas con un enfoque predictivo que nos proporcione el índice del alto crecimiento de hechos delictivos en los distintos distritos.

La siguiente investigación estará planteada en el método científico, se empleará la metodología CRISP-DM. como herramienta de análisis y predicción se usará Minería de Datos. Herramientas de desarrollo para el prototipo: se empleará la metodología SCRUM, el lenguaje de programación Python, con un gestor de base de datos Postgres, editor de código fuente Visual Studio Code.

### **1.3 ANTECEDENTES**

#### **1.3.1 *Antecedentes Afines a la Investigación***

##### **1.3.1.1 Antecedentes Internacionales**

(Almora Castro, 2018), tiene como tema de investigación titulado “ANÁLISIS Y USOS DEL BIG DATA APLICADO EN LA UNIVERSIDAD NACIONAL ‘SAN LUIS GONZAGA’ DE ICA”, el objetivo que define con su investigación es entender con mayor amplitud al Big Data como nueva tecnología para grandes volúmenes de datos, qué herramientas están relacionadas con él, así como el alcance de los datos para que sean aprovechados. La tesis de grado se desarrolló en la Universidad Nacional “San Luis Gonzaga de Ica, Ica – Perú”.

(Linares Berrocal, 2019), elaboro como tema de investigación titulado “IMPLEMENTACIÓN DE UN SISTEMA DE BIG DATA APLICADO A LA MIGRACIÓN DE DATOS BAJO LA DISTRIBUCIÓN CLOUDERA CONAPACHE HADOOP, EN EL BANCO INTERBANK”, su objetivo propuesto se enfoca en Implementar un sistema de migración de datos que permita la optimización de tiempos en el procesamiento de los datos, un ahorro en costos a nivel de infraestructura, además de gestionar el crecimiento exponencial de los datos

con el fin de garantizar la escalabilidad y alta disponibilidad del sistema. Esta tesis se desarrolló en la Universidad Tecnología del Perú.

### **1.3.1.2 Antecedentes Nacionales**

(Alejandro Aquino, 2015), que desarrollo la investigación titulada “ANALÍTICA PREDICTIVA DE BIG DATA EN SISTEMAS DE BASE DE DATOS RELACIONALES”, plantea como objetivo Diseñar un modelo que realice un análisis predictivo en una base de datos relacional de un sistema, que permita procesar información para tomar decisiones de manera oportuna, esta investigación fue elaborada en la Universidad Mayor de San Andrés.

(Fernández Villaroel, 2016), desarrollo la investigación titulada “BIG DATA EN EL COMPORTAMIENTO DE DATOS CLIMATOLÓGICOS Y ESTRATEGIAS INTERNACIONALES DE REDUCCIÓN DE DESASTRES PARA LA GESTIÓN DE RIESGO AMBIENTAL”, su objetivo planteado menciona, realizar una evaluación y gestión de riesgo ambiental a partir del análisis del comportamiento de datos climatológicos, aplicando teorías de Big Data y la teoría de Estrategias Internacionales de Reducción de Desastres (EIRD), investigación elaborada en la Universidad Mayor de San Andrés.

### **1.3.1.3 Antecedentes Locales**

(Apaza Ajtona, 2022), desarrollo el tema de investigación titulada “MODELO DE PREDICCIÓN SOBRE EL ÍNDICE DECRECIMIENTO DEL CÁNCER DE MAMA EN LAS MUJERES DE EDADES ENTRE 20 A 40 AÑOS DE LA CIUDAD DE LA PAZ, BASADO EN MINERÍA DE DATOS”, Aplicar un Modelo de Predicción sobre el índice de crecimiento del cáncer de mama en las mujeres de edades entre 20 a 40 años de la ciudad de La Paz, basado en Minería de Datos, para la toma de decisiones para prevenir las muertes por esta patología. Investigación elaborada en la Universidad Pública de El Alto.

(Hidalgo Mamani, 2020), elaboro la investigación denominada “MODELO DE PREDICCIÓN APLICADO AL ÍNDICE CRECIMIENTO DE CÁNCER DE CUELLO UTERINO EN LA CIUDAD DE EL ALTO EN BASE A ALGORITMOS GENÉTICOS”, Desarrollar un Modelo de predicción en base a Algoritmos Genéticos aplicado al Índice de Crecimiento del Cáncer de Cuello Uterino en la Ciudad de El Alto, investigación desarrollada en la Universidad Pública de El Alto.

#### **1.4 PLANTEAMIENTO DEL PROBLEMA**

En el Municipio de El Alto, a lo largo de los años presenta un incremento en los índices delictivos, debido a este hecho es clasificada como una de las ciudades más inseguras. Existe una gran cantidad de delincuencia, robos, asaltos y delitos aún mayores. Debido a esta situación existe un temor en la población.

- Existe una carencia significativa de información detallada sobre estos hechos delictivos.
- Debido a la falta de la aplicación de las nuevas tecnologías, no se dispone de una proyección precisa de los índices delictivos, lo que dificulta la acción efectiva de las autoridades
- Se desconoce la magnitud del crecimiento de estos hechos.

##### **1.4.1 *Problema Principal***

La ausencia de un modelo predictivo nos impide conocer y precisar la magnitud de los delitos, así como disponer de datos exactos o estimados sobre los índices delictivos actuales. Esta falta de información nos impide prever o determinar el crecimiento del índice delictivo en el Municipio de El Alto en el futuro.

#### **1.4.2 Problemas Secundarios**

- Ausencia de un modelo predictivo basado en Minería de Datos para los índices delictivos.
- Desconocimiento sobre el índice delictivo.
- Carencia de herramientas informáticas que apoyen en el análisis de los índices delictivos.
- Inexistencia de aplicaciones técnicas y de diseño para el análisis de los índices delictivos.

#### **1.4.3 Formulación del Problema**

¿Podrá un modelo predictivo basado en Minería de Datos ayudar a predecir los futuros índices delictivos en el Municipio de El Alto?

### **1.5 OBJETIVOS**

#### **1.5.1 Objetivo General**

Diseñar un modelo predictivo de índices delictivos basado en Minería de Datos, aplicado al Municipio de El Alto, utilizando datos recolectados de las diferentes instituciones relacionadas con la temática de Seguridad, con el fin de anticipar el incremento de los índices delictivos.

#### **1.5.2 Objetivos Específicos**

- Analizar y relevar los requerimientos necesarios sobre el índice delictivo en el Municipio de El Alto.
- Diseñar un modelo de Minería de Datos para el análisis del índice delictivo.
- Identificar los requerimientos de un modelo predictivo, para los datos recolectados.
- Desarrollar el algoritmo del modelo predictivo para el índice delictivo.

- Analizar los índices delictivos mediante los datos obtenidos utilizando el modelo de Minería de Datos.
- Establecer la relación entre los datos y el prototipo diseñado.
- Verificar e interpretar los resultados obtenidos por el prototipo diseñado con el modelo de Minería de Datos.

## 1.6 HIPÓTESIS

Se obtendrá un modelo predictivo basado en la Minería de Datos, teniendo un prototipo con la capacidad de realizar predicciones a 3 años de los índices delictivos del Municipio de El Alto con una eficiencia del 85% para un mejor conocimiento y contribución a la toma de decisiones por parte de las autoridades pertinentes.

- **H<sub>1</sub>**

Se obtendrá un modelo predictivo basado en la Minería de datos, teniendo un prototipo con la capacidad de realizar predicciones a 1 años de los índices delictivos del Municipio de El Alto con una eficiencia del 55% para un mejor conocimiento y contribución a la toma de decisiones por parte de las autoridades pertinentes.

- **H<sub>0</sub>**

No se obtendrá un modelo predictivo basado en la Minería de Datos, teniendo un prototipo con la capacidad de realizar predicciones, no logrará la predicción de los índices delictivos del Municipio de El Alto.

### 1.6.1 *Identificación de Variables*

**Variable Independiente:** Modelo Predictivo.

**Variable Dependiente:** Índices delictivos.



### 1.6.2 Conceptualización de Variables

**Tabla 1.**

*Conceptualización de variables*

| <b>Variable</b>    | <b>Definición Conceptual</b>                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Modelo predictivo  | Un modelo predictivo es un mecanismo que predice el comportamiento de un individuo. Utiliza las características del individuo como entrada y proporciona una calificación predictiva como salida (Timón, 2017).                                                                                                                                                                                      |
| Índices delictivos | El concepto Índice se utiliza como una medida estadística que permite comparar una magnitud simple o compleja en dos situaciones diferentes respecto al tiempo o al espacio, tomando una de ellas como referencia (Fernández, 2022).<br>Delictivo Pertenece al delito o relativo a él. Condición de un hecho que, como punible, está previsto y sancionado en la ley penal positiva (Derecho, 2022). |

### 1.6.3 Operacionalización de Variables

**Tabla 2.**

*Operacionalización de variables*

| <b>VARIABLE</b>    | <b>TIPO DE VARIABLE</b> | <b>DIMENSIÓN</b>        | <b>INDICADORES</b>                |
|--------------------|-------------------------|-------------------------|-----------------------------------|
| Modelo predictivo  | Variable independiente  | Patrones<br>Análisis    | Número de Patrones<br>Exponencial |
| Índices delictivos | Variable dependiente    | Incremento en el índice | Factores en los hechos            |

## **1.7 JUSTIFICACIÓN**

Las justificaciones se desarrollarán de acuerdo a los cuatro aspectos fundamentales, científicos, técnicos, económicos y sociales.

### **1.7.1 *Justificación Científica***

Creación de una herramienta para la predicción de índices delictivos. En el enfoque científico, la Minería de Datos es una herramienta y que a medida que pasa el tiempo con la moderna tecnología es muy utilizada para distintas operaciones, se implementa en casos y problemas que ayudan a la sociedad a la toma de decisiones. El modelo planteado predicción de índices delictivos basado en Minería de Datos, aplicado en el Municipio de El Alto, es muy importante porque tendrá un gran aporte a la población.

### **1.7.2 *Justificación Técnica***

Se utilizará las siguientes herramientas, lenguaje de programación Python, editor de código Visual Studio Code. Las siguientes metodologías: Método científico, Metodología CRISP-DM (Cross Industry Standard Process for Data Mining), Minería de Datos, técnica predictiva, llegando a obtener un modelo predictivo de índices delictivos que nos brindara datos en diferentes intervalos de tiempo.

### **1.7.3 *Justificación Económica***

El modelo que se elaborará no tendrá ningún tipo de costo porque será desarrollado con herramientas de software libre y herramientas que no requieran licencia, con el fin de beneficiar a las autoridades competentes al tema con un modelo tecnológico de fácil uso para la toma de decisiones y contar con el conocimiento ante el índice de crecimiento de la delincuencia en El Municipio de El Alto.

#### **1.7.4 Justificación Social**

La presente tesis modelo predictivo de índices delictivos basado en Minería de Datos aplicado en el Municipio de El Alto será muy importante, ya que de alguna manera contribuirá a la alcaldía y entidades policiales como también a la población brindando la información del índice delictivo.

Por otro lado, también favorece, en el ámbito de la política, en el tema de seguridad ciudadana para los sectores más riesgosos con un alto índice delictivo.

### **1.8 METODOLOGÍA**

Para la presente investigación se utilizará el enfoque científico con un enfoque cuantitativo, para lo cual se divide en dos partes, la primera enfocada al desarrollo de la investigación y la segunda al desarrollo del modelo.

#### **1.8.1 Método Científico**

El trabajo de investigación a desarrollar utilizar el método científico bajo el enfoque cuantitativo establecidos por (Sampieri, 2006).

El método científico está basado en los preceptos de falsabilidad (indica que cualquier proposición de la ciencia debe resultar susceptible a ser falsada) y reproducibilidad (un experimento tiene poder repetirse en lugares indistintos y por un sujeto cualquiera).

Entre los pasos necesarios que conforman el método científico se hallan la observación (el investigador debe apelar a sus sentidos para estudiar el fenómeno de la misma manera en que este se muestra en la realidad), la inducción partiendo de las observaciones, el científico debe extraer los principios particulares de ellas, el planteo de una hipótesis (surgido de la propia observación), la demostración o refutación de la misma y la representación de la tesis (la teoría científica).

- Observación.
- Experimentación.
- Comparación.
- Abstracción.
- Generación.

## **1.8.2 Técnicas de Investigación**

### **1.8.2.1 Investigación Cuantitativa**

La investigación cuantitativa es aquella donde se recogen y analizan datos cuantitativos, por su parte la cualitativa evita la cuantificación; sin embargo, los registros se realizan mediante la narración, la observación participante y las entrevistas no estructuradas.

En los métodos cuantitativos los datos se pueden acumular y comparar para tener datos comunes, en tanto que en los métodos cualitativos pueden abarcar todo el espectro de una población cuya distribución sea considerada como normal y además son personalizados. En términos generales, la validez se refiere al grado en que un instrumento realmente mide la variable que se pretende medir (Cadena y otros, 2017).

Según (Fassler, 2020), la metodología de la investigación cuantitativa existe siete pasos que deben seguirse para realizar un análisis cuantitativo.

1. Plantear Hipótesis.
2. Recolectar datos.
3. Especificar el modelo cuantitativo.

4. Realizar estadística descriptiva.
5. Estimar los parámetros del modelo cuantitativo.
6. Probar hipótesis y analizar los resultados.
7. Realizar pronósticos.

### **1.8.3 Metodología CRISP-DM**

La metodología para Minería de Datos CRISP-DM, está descrita como un proceso jerárquico, que consiste en un conjunto de tareas descritas en cuatro niveles de abstracción, desde el general hasta el específico: fase, tareas generales, tareas específicas e instancias de proceso.

Al nivel más alto, el proceso está organizado en un número de fases. Cada fase consiste en varias tareas generales de segundo nivel. Este segundo nivel se denomina genérico porque se pretende que sea lo suficientemente general como para cubrir todas las posibles situaciones. Las tareas generales deben ser lo más completas y estables posibles. Se entenderán tareas completas a aquellas que cubran completamente el proceso de análisis y sus posibles aplicaciones. Por otro lado, se entiende como estables aquellas tareas que cubran incluso desarrollos aún no conocidos.

El tercer nivel, el nivel de tareas especializadas, es el lugar en el que se describe cómo las acciones de las tareas generales (nivel 2) se deberían desarrollar en ciertas situaciones específicas. Por ejemplo, En el segundo nivel puede existir una tarea general llamada “limpieza de datos”. El tercer nivel describe cómo difiere esta tarea de unas situaciones a otras, por ejemplo, la limpieza de valores numéricos y la limpieza de valores categóricos o si el tipo de problema es un clusterizado a un modelo predictivo.

La descripción de fases y tareas en pasos discretos desarrollados en un orden específico representa una secuencia idealizada de eventos. En la práctica, muchas de estas tareas pueden ser desarrolladas en un orden diferente y frecuentemente será necesario volver atrás a tareas previas y repetir ciertas acciones. El modelo de procedimiento no pretende abarcar todas estas posibles rutas a lo largo del proyecto porque esto requeriría un modelo enormemente complejo.

El cuarto nivel, el nivel de instancias de proceso, es un conjunto de acciones, decisiones y resultados sobre el proceso de data mining en curso. Una instancia de proceso se organiza de acuerdo con las tareas definidas en los niveles superiores, pero representa lo que pasa en realidad en un proceso particular, más que lo que pasa en general.

Horizontalmente, la metodología CRISP-DM distingue entre el modelo de referencia y la guía del usuario. El modelo de referencia presenta una vista rápida de las fases, tareas y sus salidas y describe lo que hay que hacer en un proyecto de data mining. La guía del usuario da consejos y trucos mucho más detallados para cada fase y para cada tarea dentro de una fase y describe cómo desarrollar un proyecto de análisis de datos (Martínez de Pisón Ascacibar, 2003, p. 54).

Se mencionará las 6 fases del ciclo de vida que contiene un proyecto de data mining.

1. Análisis de un problema.
2. Análisis de Datos.
3. Preparación de Datos.
4. Modelizado.
5. Evaluación.

## 6. Desarrollo.

### 1.8.4 Metodología SEMMA

SAS Institute desarrollador de esta metodología, la define como el proceso de selección, exploración y modelizado de grandes cantidades de datos para descubrir patrones de negocio desconocidos.

El nombre de esta terminología es el acrónimo correspondiente a los cinco pasos básicos del proceso “Sample, Explore, Modify, Model and Assess”. El esquema siguiente presenta la dinámica del sistema (Martínez de Pisón Ascacibar, 2003).

**Muestreo:** extracción de la población muestral sobre la que se va a aplicar el análisis. En ocasiones se trata de una muestra aleatoria, pero puede ser también un subconjunto de datos de la data warehouse que cumplan unas condiciones determinadas.

**Exploración:** el objetivo es simplificar en lo posible el problema con el fin de optimizar la eficiencia del modelo. En este paso se pueden emplear herramientas que permitan visualizar de forma gráfica la información, utilizando las variables explicativas como dimensiones.

**Manipulación:** tratamiento realizado sobre los datos de forma previa a la modelización, en base a la exploración realizada, selección de variables explicativas, agrupación de variables similares.

**Modelización:** permite establecer una relación entre las variables explicativas y las variables objeto del estudio, lo que posibilitan inferir el valor de las mismas con un nivel de confianza determinado.

### **1.8.5 Metodología de desarrollo SCRUM**

La metodología de trabajo de Scrum, tiene sus principios fundamentales en la década de 1980, la cual fue desarrollada por su necesidad en procesos de reingeniería por Goldratt, Takeuchi y Nonaka. El concepto de Scrum tiene su origen sobre los nuevos procesos de desarrollo utilizados en productos exitosos en Japón y los Estados. Los equipos que desarrollaron estos productos partían de requisitos muy generales, así como novedosos, y debían salir al mercado en mucho menos del tiempo del que se tardó en lanzar productos anteriores. Estos equipos seguían patrones de ejecución de proyecto muy similares. En este estudio se comparaba la forma de trabajo de estos equipos altamente productivos y multidisciplinarios con la colaboración entre los jugadores de Rugby y su formación de Scrum, de la cual se tomó su nombre (Maida & Pacienza, 2015, p. 74).

### **1.8.6 Método de Ingeniería**

#### **1.8.6.1 Métrica de Calidad**

La Norma ISO 25000, proporciona una guía para el uso de las series de estándares internacionales llamados requisitos y Evaluación de Calidad de Productos Software (SQuaRE). La norma establece criterios para la especificación de requisitos de calidad de productos software, sus métricas y su evaluación, e incluye un modelo de calidad para unificar las definiciones de calidad de los clientes con los atributos en el proceso de desarrollo. Pueden ser útiles no solamente para evaluar el producto software, sino también para definir los requerimientos de calidad (ISO 25000, 2022).

#### **1.8.6.2 Estimación de Costos**

El modelo Cocomo II permite realizar estimaciones en función del tamaño del software, y de un conjunto de factores de costo y de escala. Posee tres modelos: Composición de Aplicación, Diseño Temprano y Post-Arquitectura.



En los factores de coste se incluyen aspectos relacionados con la naturaleza del sistema, equipo, y características propias del proyecto.

Producida a medida que un proyecto de software incrementa su tamaño (Gómez y otros, 2010).

### **1.8.6.3 Seguridad de La Información**

La norma ISO/IEC 27001 es actualmente la norma internacional más reconocida para los sistemas de gestión de la seguridad.

Ayuda a las organizaciones a establecer la política y los objetivos de gestión de la seguridad de la información y a comprender cómo se pueden gestionar los aspectos importantes, aplicar los controles necesarios y establecer objetivos claros para mejorar la seguridad de la información.

Permite a una organización gestionar su obligación de cumplir con los requisitos legales aplicables, como el GDPR (junto con la norma ISO 27701) y comprobar periódicamente el estado de cumplimiento. Esto permite una mejora continua del sistema para garantizar la protección y abordar las vulnerabilidades.

Adopta un enfoque integral de la seguridad de la información. Los activos que necesitan protección van desde la información digital, los documentos en papel y los activos físicos (ordenadores y redes) hasta los conocimientos de los empleados individuales. Las cuestiones que hay que abordar van desde el desarrollo de la competencia del personal a la protección técnica contra el fraude informático.

La norma ISO 27001 está diseñada para ser compatible y armonizada con otras normas reconocidas de sistemas de gestión. Por lo tanto, es ideal para su integración en los sistemas y procesos de gestión existentes.

## **1.8.7 Pruebas de Hipótesis**

### **1.8.7.1 T-STUDENT**

Actualmente, se conoce a la distribución T de Student como la distribución usada para hacer inferencia sobre la media de una población normalmente distribuida cuando el tamaño muestral es pequeño. Esta distribución fue formalmente introducida por el químico William Sealy Gosset (1876-1937) en el año 1908. Este trabajaba para la famosa marca de cerveza Guinness (de Dublín) con el fin de lograr avances para el proceso industrial y la comercialización de la marca. El artículo donde se introducía al mundo esta distribución fue publicado en la revista inglesa Biometrika en el año 1908. Gosset firmó con un pseudónimo, Student, para que así se mantuvieran en secreto los procesos y avances industriales de la cervecera.

Los grados de libertad son el grado de información dada por el conjunto de observaciones de la muestra que se emplean en estimar los valores de parámetros desconocidos o en estimar la variabilidad de los tests. Se determinan dependiendo del número de parámetros del modelo y del número de observaciones. Por un lado, es obvio que, si aumenta el número de observaciones, aumenta la cantidad de información, lo que se traduce como un incremento del número de los grados de libertad. Por otra parte, si añadimos parámetros al modelo estamos “gastando” información, lo que reducirá el número de los grados de libertad. Se calculan mediante la fórmula  $(n - r)$  donde  $n$  es el tamaño de la muestra y  $r$  el número parámetros que es necesario estimar (Crespo, 2021).

## **1.9 HERRAMIENTAS**

Con la tecnología de hoy en día se pueden contar con distintas herramientas para la realización y manejo de Minería de Datos, se detallará lo requerido para la propuesta mencionada tanto en Hardware como en Software en los siguientes puntos.

### 1.9.1 *Hardware*

- Laptop.
- Microprocesador Intel Core i3-1125G4 / i3-1115G4 de 2.0 GHz / 3.0 GHz- onceava generación Equipos de computación para el desarrollo.
- RAM de 4 GB DDR4-2400 ampliable.
- SATA de 1 TB 5400 RPM.

### 1.9.2 *Software*

**Windows 10:** Viene siendo la anteúltima versión desarrollada por Microsoft como parte de la familia de Windows NT.5. Esta versión ha presentado innovaciones en comparación con las anteriores, ya que Microsoft ofrece gratuitamente este sistema operativo para aquellos usuarios que cuenten con copias originales de Windows 7 y Windows 8.1 update. Es una edición supercompleta diseñada para toda la familia de los productos Microsoft, tales como: laptops, tabletas, teléfonos inteligentes, Xbox One, entre otros. Esto se da a partir de su código casi idéntico que le permite tener tal compatibilidad en las diferentes herramientas tecnológicas (ConceptoDefinition, 2022).

**PostgreSQL:** Es un servidor de Bases de Datos relacionales Orientadas a Objetos, de software libre bajo licencia BSD.

Se opta por PostgreSQL, es un magnífico gestor de bases de datos. Tiene Prácticamente todo lo que tienen los gestores comerciales, haciendo De él una muy buena alternativa GPL (eni, 2022).

**Python:** Es un lenguaje de programación con la ventaja de ser relativamente fácil de usar para usuarios que no estén familiarizados con la informática de manera profesional, pero

que necesitan trabajar con análisis de datos para Minería de Datos muy eficiente, en parte debido a la gran comunidad existente, por lo que Python dispone de muchas librerías ya hechas por otros usuarios (profile, 2022).

## **1.10 LÍMITES Y ALCANCES**

### **1.10.1 Límites**

- Esta investigación es limitada a los datos del Municipio de El Alto y no se está tomando en cuenta a los otros municipios.
- En el presente modelo de predicción solo tomará las áreas de índices de delincuencia y no así la predicción poblacional.
- El modelo planteado no podrá resolver otro problema fuera del entorno mencionado.
- Le modelo no incluirá los tipos de delitos.

### **1.10.2 Alcances**

- Se diseñará un Modelo de Minería de Datos a base de los datos adquiridos sobre los índices delictivos en el Municipio de El Alto.
- Se desarrollará el modelo predictivo basándonos en Minería de Datos.
- La presente investigación utilizara técnicas apoyadas en Minería de Datos, lo cual nos ayudara a predecir así también conocer el índice delictivo basado en Minería de Datos aplicado en el Municipio de El Alto.

### 1.11 Aportes

Los aportes que ofrecerá el modelo serán los siguientes:

- La implementación del modelo basado en Minería de Datos será beneficioso para la población en general del Municipio de El Alto, con la finalidad de proporcionar la información de los índices delictivos.
- La ejecución del modelo predictivo de los datos obtenidos nos brindará la verificación del aumento y disminución del índice delictivo.
- El modelo permitirá una mejor planificación y toma de decisiones en el ámbito de la seguridad pública mediante predicciones futuras.
- La investigación también fomentará la innovación hacia futuras investigaciones, al mostrar cómo los datos pueden ser enfocados y adaptados a otros problemas sociales.



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# CAPÍTULO II

## MARCO TEÓRICO



## 2 CAPÍTULO II

### 2.1 MARCO TEÓRICO

### 2.2 DATO

Según Davenport y Prusak en 1999, dato es un conjunto discreto de factores objetivos sobre un hecho real. Un dato no dice nada sobre el porqué de las cosas y, por sí mismo, tiene poca o ninguna relevancia o propósito.

Los datos describen únicamente una parte de lo que pasa en la realidad y no proporcionan juicios de valor o interpretaciones. Por lo tanto, no son orientativos para la acción. La toma de decisiones se basará en datos, pero estos nunca indicarán qué hacer.

Los datos no dicen nada acerca de lo que es importante o no. A pesar de todo, los datos son importantes para las organizaciones, pues son la base para la creación de información. (Rojas, 2022).

### 2.3 INFORMACIÓN

Idalberto Chiavenato afirmaba que la información consiste en un conjunto de datos que poseen un significado, de modo tal que reducen la incertidumbre y aumentan el conocimiento de quien se acerca a contemplarlos.

Estos datos se encuentran disponibles para su uso inmediato y sirven para clarificar incertidumbres sobre determinados temas (Thompson, 2022).

### 2.4 CONOCIMIENTO

Para Davenport y Prusak en 1999, el conocimiento es una mezcla de experiencia, valores, información y “saber hacer”, que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción.

El conocimiento deriva de la información, así como la información deriva de los datos. Estas actividades de creación de conocimiento tienen lugar dentro y entre personas. Al igual que encontramos datos en registros, e información en mensajes, obtenemos conocimiento de individuos, grupos de conocimiento, o incluso de rutinas organizativas (Rojas, 2022).

## **2.5 MINERÍA DE DATOS**

La Minería de Datos busca el procesamiento de información de forma clara para el usuario o cliente, de tal forma que pueda clasificar la información de acuerdo a parámetros inicialmente establecidos y de acuerdo a las necesidades que se buscan en el caso. La Minería de Datos produce cinco tipos de información, entre estas tenemos. Asociaciones, Secuencias, Clasificaciones, Agrupaciones, Pronósticos (Vizcanio, 2008).

La Minería de Datos es un intento de buscarle sentido a la información que actualmente puede ser almacenada en grandes volúmenes. La fase de minar los datos es la representación del tipo de modelo obtenido. Se concentra en la búsqueda, que tendrán unas varias formas de representación en dependencia del tipo de modelo obtenido.

El análisis de los datos puede proporcionar en conjunto un verdadero conocimiento que ayude en la toma de decisiones. Puede definirse como el uso consistente de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre procesados, que sean de utilidad para la toma de decisiones. Se relacionan de manera estrecha con la estadística, usando técnicas de muestreo y visualización de datos. (Gutiérrez cruz y otros, 2019).

Características especiales de los datos:

Aparte del gran volumen, ¿por qué el aprendizaje del sistema y las estrategias estadísticas no son inmediatamente relevantes?



- Los datos residen en el disco. No se pueden escanear un par de veces.
- Algunas técnicas de muestreo no son compatibles con los algoritmos no incrementales.
- Dimensionalidad muy excesiva (muchos campos).
- Prueba positiva.
- Datos imperfectos

Aunque algunos se aplican de forma casi inmediata, la afición en los estudios de Minería de la información está en adaptarlos. Patrones a localizar:

- Una vez acumulada la información de interés, un explorador puede decidir qué tipo de estilos desea averiguar.
- El tipo de conocimiento a extraer marcará absolutamente el enfoque de minería de información a utilizar.
- Dependiendo del tipo de conocimiento que se busque, se puede distinguir entre:
  - Minería de información dirigida: se sabe realmente lo que se busca, normalmente se esperan datos positivos o instrucciones.
  - Minería de estadísticas no dirigida: ahora no se reconoce lo que se busca, se figura con los hechos.

En el primer caso, las propias estructuras de minería de registros son normalmente responsables de decidir el máximo conjunto apropiado de reglas entre las disponibles para un determinado tipo de patrón a buscar. (Beltrán, 2018).

### **2.5.1 Historia**

Es correcto que los inicios de la Minería de Datos (DM) pueden rastrearse hasta los años 50, aunque la disciplina en sí misma se consolidó y se hizo más prominente en las décadas siguientes. La descripción que ha proporcionado refleja la evolución de los primeros

sistemas de información y cómo se desarrollaron con el objetivo de resumir la información contenida en los ficheros de los ordenadores centrales para facilitar la toma de decisiones. Aquí hay algunos puntos clave sobre el desarrollo de la Minería de Datos y los sistemas de información. En los 60 nacen los sistemas gestores de base de datos que aún se mostraban rígidos y carecían de flexibilidad para realizar consultas. Luego aparecieron los motores relacionales resolviendo estos problemas, aunque los informes resultaban muy laboriosos de preparar y depurar, perdiéndose relevancia por su bajo nivel de actualización. Otro grave problema era la diversidad de bases de datos no integradas establecidas por los diferentes departamentos de una organización. Nadie reparaba en la posible utilidad futura de un sistema interdependiente. (Beltrán, 2018)

El Data Warehouse (DW) viene a solucionar este problema en los finales de los 80. La existencia de DW ha estimulado el desarrollo de los enfoques de DM, en los que las tareas de análisis se automatizan y dan un paso más al posibilitar la extracción de conocimiento inductivo.

(Beltrán, 2018) indica que, aunque los componentes clave del Data Mining o Minería de datos (DM) existen desde hace décadas en la investigación en áreas como la inteligencia artificial, la estadística o el aprendizaje automático, se puede afirmar que ahora estamos asistiendo al reconocimiento de la madurez de estas técnicas, lo que, junto al espectacular desarrollo de los motores de bases de datos y las herramientas para integración de información justifican su introducción en la esfera empresarial.

**Tabla 3.***Evolución de las tecnologías relacionadas con Data Mining*

| <b>Etapa</b>                                              | <b>Cuestión Planteada</b>                                              | <b>Tecnologías</b>                                                           | <b>Características</b>                                          |
|-----------------------------------------------------------|------------------------------------------------------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Recolección de datos en el año 60                         | Dime mis beneficios totales en los últimos 4 años                      | Ordenadores, cintas, discos.                                                 | Retrospectivo, datos estáticos.                                 |
| Acceso a los datos (años 80)                              | Ventas en Cataluña durante las últimas Navidades                       | Bases de Datos Relacionales (SQL) ODBC                                       | Retrospectivo, datos dinámicos a nivel de registro.             |
| Data Warehouse y soporte a la toma de decisiones. Años 90 | Ventas en Andalucía detalle por delegación y descender a nivel tienda. | (OLAP), bases de datos multidimensionales, data warehouse                    | Retrospectivo, obtención dinámica de datos a múltiples niveles. |
| Data Mining                                               | Justifica la tendencia de venta en Castilla para el próximo año        | Algoritmos avanzados, ordenadores, multiprocesadores, bases de datos masivas | Prospectivo, obtención proactiva de información.                |

*Nota.* Cuadro descriptivo de la evolución de las tecnologías relacionadas con Data Mining (Beltrán, 2018).

### **2.5.2 Definición de la Minería de Datos**

El nombre de Data Mining, o Minería de Datos, deriva de la similitud que se encuentra entre buscar valiosa información de negocios en grandes bases de datos con la búsqueda de vetas de metales preciosos dentro de una montaña. Ambos procesos requieren examinar

inteligentemente una inmensidad de material hasta encontrar algo que pueda resultarnos útil y valiosa.

La definición del concepto de Data Mining (DM) puede variar entre unos investigadores y otros. Por ejemplo, los estadísticos, analistas de datos y la comunidad de sistemas de gestión de la información adoptan mayoritariamente este término para referirse al proceso genérico correspondiente a las técnicas y herramientas de investigación usadas para extraer información útil de una base de datos. Dentro de estas técnicas podemos considerar todos aquellos métodos matemáticos y técnicas, software para el análisis inteligente de los datos y búsqueda de patrones o tendencias en los mismos, aplicados de forma iterativa e interactiva.

Dentro de las definiciones que se pueden encontrar en la literatura relacionada se muestran algunas de las más significativas:

- Data Mining es la exploración y análisis, mediante métodos automáticos o semiautomáticos, de grandes cantidades de datos para descubrir reglas o patrones significativos. (Berry & Linoff, 1997).
- Data Mining es el proceso de plantear varias preguntas y extraer información útil, patrones y tendencias de grandes cantidades de datos generalmente almacenados en bases de datos. (Rodríguez Suárez & Anolandy Díaz, 2009).
- Data Mining es el conjunto de técnicas y herramientas aplicadas al proceso trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con el objeto de predecir de forma automatizada tendencias y comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos. (Shapiro & WJ, 1991).

En cambio, en el ámbito del Knowledge Discovery in Databases (KDD) o descubrimiento de conocimiento en bases de datos, el Data Mining tiene otro significado. Efectivamente, el término KDD se empezó a utilizar en 1989

(Shapiro & WJ, 1991) popularizándose por los expertos en inteligencia artificial (IA) y aprendizaje de ordenadores (Machine Learning) para referirse al amplio proceso de búsqueda de conocimiento en bases de datos y para enfatizar de que este “conocimiento” es el producto final del proceso del KDD. La definición de KDD más representativa surge de diversos autores especialistas en ese campo:

- Descubrimiento en bases de datos (KDD): es el proceso no trivial de identificar patrones en datos que sean válidos, novedosos, potencialmente útiles y por último comprensibles.

Las fases en que se divide el KDD según son: exploración del dominio, recolección de los datos, extracción de patrones en los datos, inducir generalizaciones, verificación del conocimiento, transformación del conocimiento.

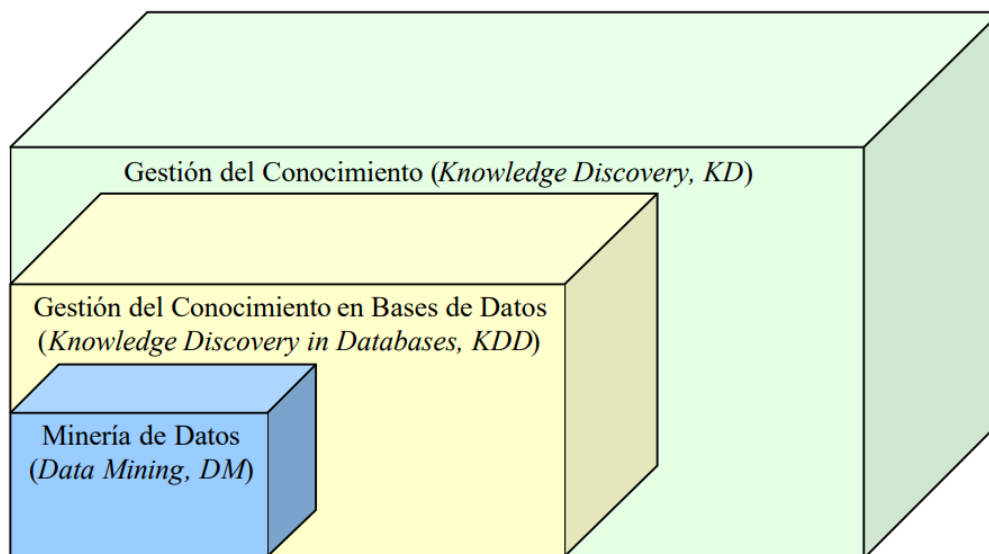
- De esta forma, y según los autores provenientes del campo de la IA o del Machine Learning (ML), el Data Mining corresponde a un paso del KDD y se define en la literatura especialista de las siguientes formas:
- Data Mining consiste en obtener modelos comprensibles o patrones de una base de datos.
- Data Mining: búsqueda de patrones de interés mediante árboles o reglas de clasificación, técnicas de regresión, clusterizado, modelizado secuencial, dependencias.

- El proceso de extraer patrones o modelos a partir de los datos.

En esta última definición coinciden la mayor parte de los autores que se dedican a la Data Mining, el KDD, la IA o el ML, aunque también otros autores, como se ve en definiciones anteriores, describen el DM como el proceso completo (Martínez de Pisón Ascacibar, 2003).

### Figura 1.

*La Minería de Datos frente al KD y KDD*



*Nota:* Se resume cómo el DM está incluido en el KDD, y cómo este se incluye a su vez en el KD.

### 2.5.3 Modelos de Minería de Datos

Los modelos de Minería de Datos se refieren a estructuras o algoritmos matemáticos utilizados en el proceso de descubrimiento de patrones, relaciones y conocimientos ocultos en conjuntos de datos extensos. Estos modelos pueden abordar una variedad de datos, como históricos, transaccionales o comportamentales, y emplear diversas técnicas, como árboles de decisión, regresión, redes neuronales, entre otros. Su propósito principal es proporcionar una

representación estructurada de los datos que permita la extracción de información valiosa y predictiva para mejorar la toma de decisiones y los procesos en diversas áreas.

### **2.5.3.1 Modelos Descriptivos**

Los métodos Descriptivos o aprendizaje no supervisado permiten formar grupos de datos rápidamente, también son conocidos como métodos simétricos, no supervisados o indirectos.

Las observaciones son generalmente clasificadas en grupos que no son conocidos con anterioridad, los elementos de las variables pueden estar conectados entre sí de acuerdo a vínculos desconocidos de antemano, de esta manera, todas las variables disponibles son tratados en el mismo nivel y no hay hipótesis de causalidad. (Hernández Avalos, 2015).

El aprendizaje no supervisado es no dirigido. No hay distinción entre atributos dependientes e independientes. Es decir, no hay un resultado previamente conocido que guie al algoritmo en la construcción del modelo. Por ejemplo, la Minería de Datos no supervisada puede ser usada para propósitos descriptivos.

Aunque también puede ser usada para hacer predicciones. (Hernández Cedano, 2015).

### **2.5.3.2 Modelos Predictivos**

Los métodos predictivos o de aprendizaje supervisado se basan en entrenar a un modelo o método por medio de diferentes datos para poder predecir una variable partiendo de estos mismos datos. Con lo que el método ya aprendido, su respuesta será que eso es un cuadrado, porque ya lo aprendió con los datos anteriores, es por ello que se llama aprendizaje supervisado o modelo predictivo porque queremos predecir ¿Qué es? Esa nueva figura. (Hernández Avalos, 2015).

El aprendizaje supervisado es también conocido como aprendizaje dirigido. El proceso de aprendizaje es dirigido por un atributo u objetivo dependiente previamente conocido. El aprendizaje supervisado generalmente resulta en modelos predictivos. Siendo este el contraste para el aprendizaje no supervisado, donde la meta es la detección de patrones.

La construcción de un modelo supervisado involucra el entrenamiento, un proceso mediante el cual el software analiza muchos casos donde el valor objetivo ya es conocido. En el proceso de entrenamiento, el modelo “aprende” la lógica de hacer la predicción. Por ejemplo, un modelo que busca identificar a los clientes que probablemente respondan a una promoción, debe ser entrenado para que analice las características de muchos clientes que ya se sabe que respondieron o no respondieron a una promoción en el pasado. (Hernández Cedano, 2015).

## **2.6 ETAPAS DE LA MINERÍA DE DATOS**

Existen términos que se utilizan frecuentemente como sinónimos de la Minería de Datos. Uno de ellos se conoce como análisis (inteligente) de datos, que suele hacer un mayor hincapié en las técnicas de análisis estadístico. Otro término muy utilizado, y el más relacionado con la Minería de Datos, es la extracción o “descubrimiento de conocimiento en bases de datos” (Knowledge Discovery in Databases, KDD). De hecho, en muchas ocasiones ambos términos se han utilizado indistintamente, aunque existen claras diferencias entre los dos. Así, últimamente se ha usado el término KDD para referirse a un proceso que consta de una serie de fases, mientras que la Minería de Datos es solo una de estas fases. se define el KDD como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”.

En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

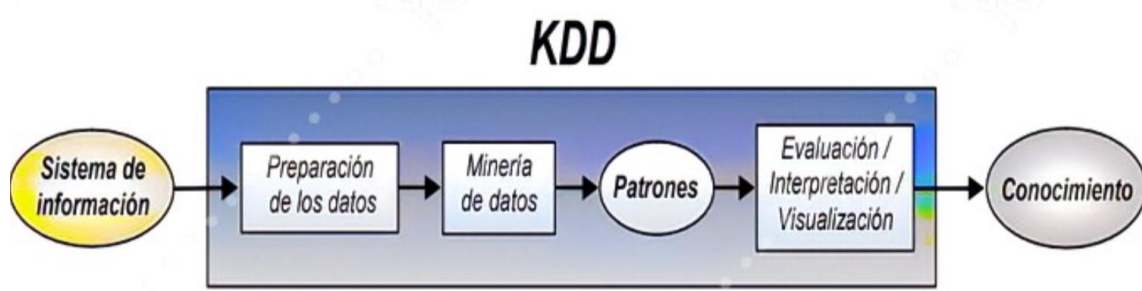


- Válido: hace referencia a que los patrones deben seguir siendo precisos para datos nuevos (con un cierto grado de certidumbre), y no solo para aquellos que han sido usados en su obtención.
- Novedoso: que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.
- Potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- Comprensible: la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de su utilidad).

Como se deduce de la anterior definición, el KDD es un proceso complejo que incluye no solo la obtención de los modelos o patrones (el objetivo de la Minería de Datos), sino también la evaluación y posible interpretación de los mismos, tal y como se refleja en la siguiente figura.

**Figura 2.**

*Proceso del KDD*

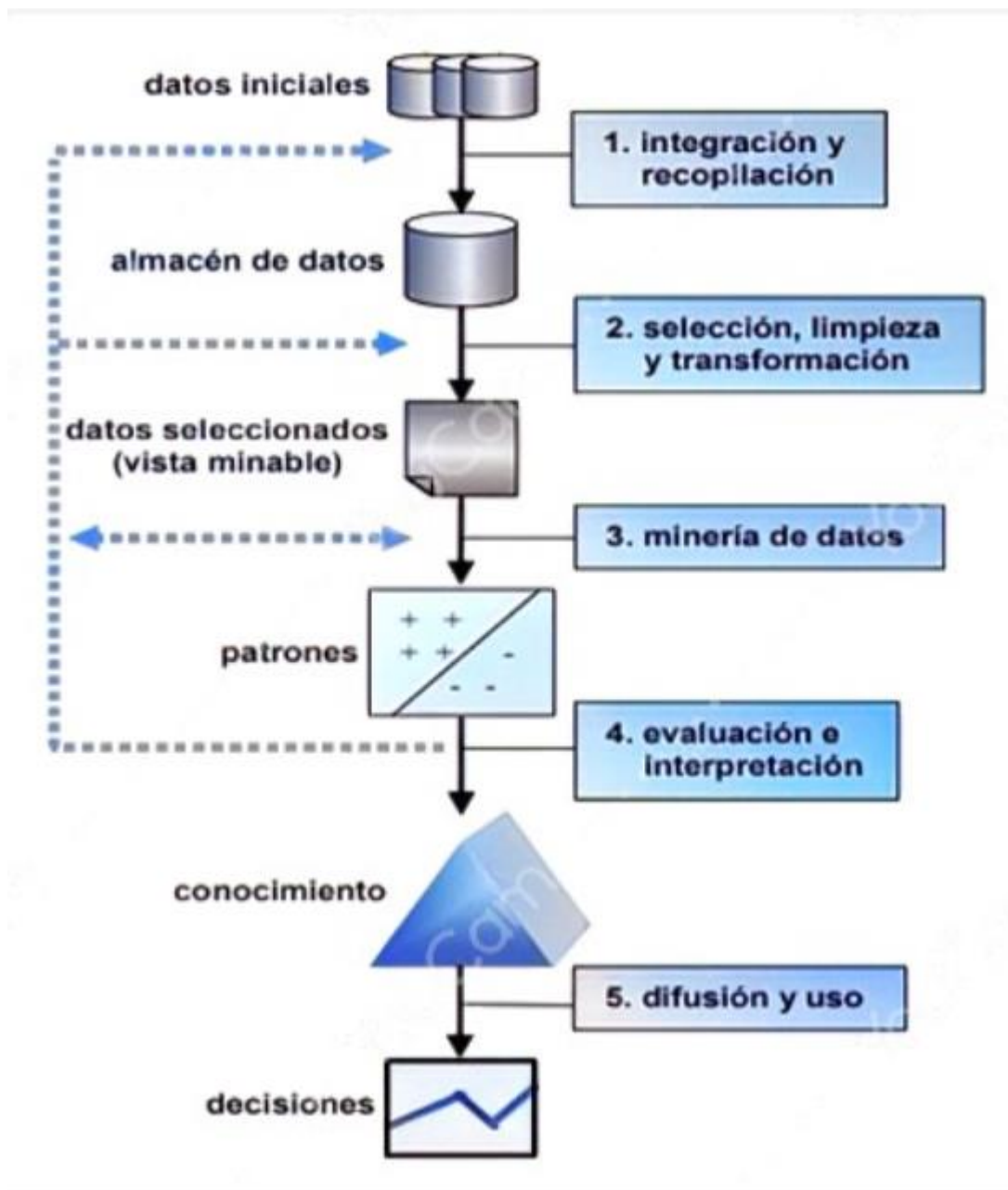


*Nota.* Introducción a la Minería de Datos proceso KDD (Ramírez y otros, 2004).

### 2.6.1 Proceso de Extracción de Conocimiento

Figura 3.

Etapas del proceso KDD



Nota. Fases del proceso de descubrimiento de conocimiento en bases de datos, KDD.

(Ramírez y otros, 2004).

### **2.6.1.1 Integración y Recopilación**

Se señala que las bases de datos y aplicaciones centradas en el procesamiento de datos convencional, conocido como procesamiento transaccional en línea (OLTP, On-Line Transaction Processing), son adecuadas para satisfacer las necesidades diarias de una organización, como facturación, control de inventario y nóminas. Sin embargo, estas resultan insuficientes para funciones más complejas como el análisis, la planificación y la predicción, especialmente en la toma de decisiones estratégicas a largo plazo. En estos casos, los datos necesarios para llevar a cabo un proceso de KDD pueden provenir de diversas organizaciones o departamentos dentro de una entidad. Incluso puede suceder que algunos datos esenciales para el análisis nunca hayan sido recopilados internamente debido a que no son necesarios para las aplicaciones cotidianas. Frecuentemente, será necesario adquirir datos externos, ya sea de bases de datos públicas (censo, datos demográficos, climatológicos) o privadas (datos de compañías de pagos, bancos, eléctricas), siempre y cuando sea a un nivel agregado para cumplir con la legalidad.

Este proceso presenta desafíos, ya que cada fuente de datos utiliza diferentes formatos de registro, grados de agregación, claves primarias y tipos de errores. La integración de estos datos se vuelve prioritaria. La noción de integrar múltiples bases de datos ha dado origen a la tecnología de almacenes de datos (data warehousing), una tendencia actual en empresas e instituciones que consiste en recopilar datos de bases de datos transaccionales y diversas fuentes para hacerlos accesibles en análisis y toma de decisiones. (Ramírez y otros, 2004).

**Figura 4.***Integración en un almacén de datos*

*Nota.* Integración en un almacén de datos (Ramírez y otros, 2004).

Un almacén de datos es un repositorio de información coleccionada desde varias fuentes, almacenada bajo un esquema unificado que normalmente reside en un único emplazamiento.

Existen varias formas de mezclar las distintas bases de datos para crear el repositorio. Una posibilidad es simplemente hacer una copia de las bases de datos integrantes (probablemente eliminando inconsistencias y redundancias). Obviamente, esta aproximación limita las ventajas para acceder a bases de datos heterogéneas. Por ello, generalmente los almacenes de datos se construyen vía un proceso de integración y almacenamiento en un nuevo esquema integrado. Se muestra este proceso de integración de un almacén de datos para tres fuentes de datos originales (A, B y C).

Fundamentalmente, los almacenes de datos se emplean para la eficiente consolidación y correlación de la información de manera avanzada. En este sentido, los datos se estructuran mediante una base de datos multidimensional, donde cada dimensión corresponde a un atributo o conjunto de atributos en el esquema que se centra alrededor de "hechos", los cuales almacenan el valor de alguna medida agregada, como, por ejemplo, la cantidad de un producto vendido en un día específico en una tienda. Esta perspectiva multidimensional hace que los

almacenes de datos sean apropiados para el procesamiento analítico en línea (OLAP). Las operaciones OLAP permiten un análisis multidimensional de los datos, superando al SQL en la generación de resúmenes y desgloses en diversas dimensiones. Además, pueden utilizar el conocimiento previo sobre el dominio de los datos para presentarlos a diferentes niveles de abstracción, adaptándose así a distintos puntos de vista de los usuarios.

Es importante destacar la diferencia entre Minería de Datos y OLAP para los profesionales del procesamiento de datos. Los usuarios de herramientas OLAP emplean la herramienta para obtener información agregada a partir de datos detallados, combinando información de manera flexible para obtener informes y vistas avanzadas en tiempo real. Las herramientas OLAP también pueden utilizarse para verificar rápidamente patrones y pautas hipotéticas propuestas por el usuario, siendo un proceso esencialmente deductivo. En contraste, la Minería de Datos, en lugar de verificar patrones hipotéticos, utiliza los datos para descubrir estos patrones, siendo un proceso inductivo. Ambos tipos de herramientas se complementan: se puede utilizar OLAP al principio del proceso de KDD para explorar los datos, identificar variables importantes, excepciones y encontrar interacciones. Aunque un almacén de datos es muy recomendable para la Minería de Datos, no es imprescindible. En algunos casos, especialmente cuando el volumen de datos no es muy grande, es posible trabajar con los datos originales o en formatos heterogéneos, como archivos de texto u hojas de cálculo. (Ramírez y otros, 2004).

#### **2.6.1.2 Selección, Limpieza y Transformación**

La calidad del conocimiento descubierto no solo depende del algoritmo de minería utilizado, sino también de la calidad de los datos minados. Por ello, después de la recopilación, el siguiente paso en el proceso de KDD es seleccionar y preparar el subconjunto de datos que se va a minar, los cuales constituyen lo que se conoce como vista minable. Este paso es

necesario, ya que algunos datos coleccionados en la etapa anterior son irrelevantes o innecesarios para la tarea de minería que se desea realizar.

Además de la falta de relevancia, hay otros problemas que impactan la calidad de los datos. Uno de estos inconvenientes es la presencia de valores que no siguen el patrón general de los datos, conocidos como valores atípicos. Estos datos anómalos pueden indicar errores o simplemente ser valores legítimos que se diferencian del resto. Algunos algoritmos de Minería de Datos optan por ignorar o descartar estos datos, considerándolos ruido o excepciones, pero otros son muy sensibles y esto afecta claramente los resultados. No obstante, eliminarlos no siempre es la opción más adecuada, especialmente en aplicaciones como la detección de compras fraudulentas con tarjetas de crédito o la predicción de inundaciones, donde eventos inusuales pueden resultar más relevantes que los eventos regulares, como compras significativamente mayores al promedio o días con lluvias notablemente superiores a la media.

La selección de atributos relevantes es uno de los preprocesamientos más importantes, ya que es crucial que los atributos utilizados sean relevantes para la tarea de Minería de Datos. Por ejemplo, supongamos que los jueces del torneo de Wimbledon desean determinar a partir de las condiciones climatológicas (nubosidad, humedad, temperatura, etc.) si se puede jugar o no al tenis. Para ello se cuenta con los datos recogidos de experiencias anteriores. Probablemente, la base de datos contenga un atributo que identifica cada uno de los días considerados (por ejemplo, la fecha). Si consideramos este atributo en el proceso de minería, un algoritmo de generación de reglas podría obtener reglas como:

- SI (fecha=10/06/2003) ENTONCES (jugar\_tenis=sí)

que, aunque correcta, es inútil para realizar predicciones futuras. Idealmente, uno podría usar todas las variables y dejar que la herramienta de Minería de Datos fuera probando

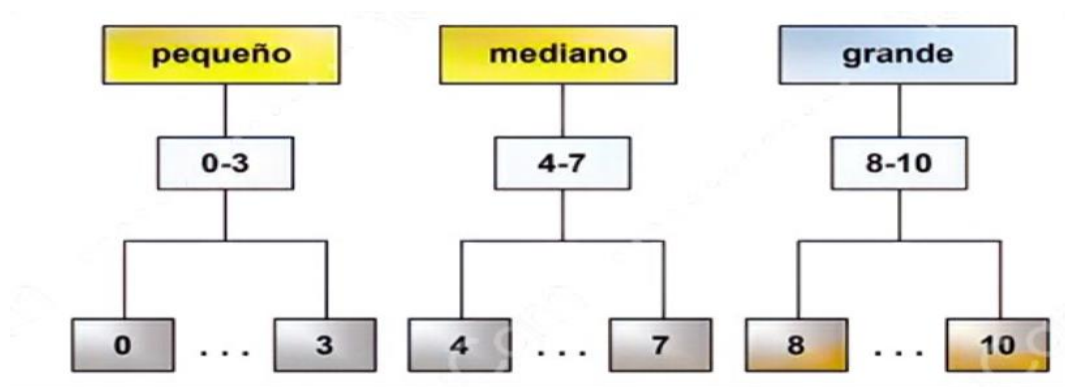
hasta elegir las mejores variables predictoras. Obviamente, esta forma de trabajar no funciona bien, entre otras cosas porque el tiempo requerido para construir un modelo crece con el número de variables. Aunque en principio algunos algoritmos de Minería de Datos automáticamente ignoran las variables irrelevantes, en la práctica nuestro conocimiento sobre el dominio del problema puede permitirnos hacer correctamente muchas de esas selecciones.

El tipo de los datos puede también modificarse para facilitar el uso de técnicas que requieren tipos de datos específicos. Así, algunos atributos se pueden numerizar, lo que reduce el espacio y permite usar técnicas numéricas. Por ejemplo, podemos reemplazar los valores del atributo “tipo de vivienda” por enteros.

El proceso inverso consiste en discretizar los atributos continuos, es decir, transformar valores numéricos en atributos discretos o nominales. Los atributos discretizados pueden tratarse como atributos categóricos con un número más pequeño de valores. La idea básica es partir los valores de un atributo continuo en una pequeña lista de intervalos, tal que cada intervalo es visto como un valor discreto del atributo. (Ramírez y otros, 2004).

**Figura 5.**

*Discretización del atributo*



*Nota.* Ejemplo de discretización del atributo tamaño (Ramírez y otros, 2004).

### **2.6.1.3 Minería de Datos**

La fase de Minería de Datos es la más característica del KDD y, por esta razón, muchas veces se utiliza esta fase para nombrar todo el proceso. El objetivo de esta fase es producir nuevo conocimiento que pueda utilizar el usuario. Esto se realiza construyendo un modelo basado en los datos recopilados para este efecto. El modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas. Para ello es necesario tomar una serie de decisiones antes de empezar el proceso:

- Determinar qué tipo de tarea de minería es el más apropiado. Por ejemplo, podríamos usar la clasificación para predecir en una entidad bancaria los clientes que dejarán de serlo.
- Elegir el tipo de modelo. Por ejemplo, para una tarea de clasificación podríamos usar un árbol de decisión, porque queremos obtener un modelo en forma de reglas.
- Elegir el algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo que estamos buscando. Esta elección es pertinente porque existen muchos métodos para construir los modelos. Por ejemplo, para crear árboles de decisión para clasificación podríamos usar CART o C5.0, entre otros. En los capítulos siguientes se presentarán los métodos más importantes para cada tipo de modelo. En lo que resta de esta sección, describimos las tareas y modelos más utilizados, así como algunos conceptos relacionados con la construcción del modelo. (Ramírez y otros, 2004).

### **2.6.1.4 Interpretación y Evaluación**

Medir la calidad de los patrones descubiertos por un algoritmo de Minería de Datos no es un problema trivial, ya que esta medida puede atañer a varios criterios, algunos de ellos



bastante subjetivos. Idealmente, los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles (es decir, inteligibles) e interesantes (útiles y novedosos). Según las aplicaciones puede interesar mejorar algún criterio y sacrificar ligeramente otro, como en el caso del diagnóstico médico que prefiere patrones comprensibles, aunque su precisión no sea muy buena. (Ramírez y otros, 2004).

#### **2.6.1.5 Difusión, Uso y Monitorización**

Una vez construido y validado, el modelo puede usarse principalmente con dos finalidades: para que un analista recomiende acciones basándose en el modelo y en sus resultados, o bien para aplicar el modelo a diferentes conjuntos de datos. También puede incorporarse a otras aplicaciones, como por ejemplo a un sistema de análisis de créditos bancarios, que asista al empleado bancario a la hora de evaluar a los solicitantes de los créditos, o incluso automáticamente, como los filtros de spam o la detección de compras con tarjetas de crédito fraudulentas. Tanto en el caso de una aplicación manual o automática del modelo, es necesario su difusión, es decir, que se distribuya y se comunique a los posibles usuarios, ya sea por cauces habituales dentro de la organización, reuniones, intranet.

El nuevo conocimiento extraído debe integrar el know-how de la organización. También es importante medir lo bien que el modelo evoluciona. Aun cuando el modelo funcione bien, debemos continuamente comprobar las prestaciones del mismo. Esto se debe principalmente a que los patrones pueden cambiar. Por ejemplo, todos los vendedores saben que las ventas se ven afectadas por factores externos como la tasa de inflación, la cual altera el comportamiento de compra de la gente. Por lo tanto, el modelo deberá ser monitorizado, lo que significa que de tiempo en tiempo el modelo tendrá que ser reevaluado, reentrenado y posiblemente reconstruido completamente. (Ramírez y otros, 2004).

## 2.7 MÉTODOS Y TÉCNICAS DE LA MINERÍA DE DATOS

### 2.7.1 *Métodos de la Minería de Datos*

Los métodos de Minería de Datos tienen como metas primarias (en un alto nivel) la predicción de datos desconocidos y la descripción de patrones.

Pueden emplearse diferentes criterios para clasificar los sistemas de Minería de Datos y, en general, los sistemas de aprendizaje inductivo en computadoras:

- Dependiendo del objetivo para el que se realiza el aprendizaje, pueden distinguirse sistemas para: clasificación (clasificar datos en clases predefinidas), regresión (función que convierte datos en valores de una función de predicción), agrupamiento de conceptos (búsqueda de conjuntos en los que agrupar los datos), compactación (búsqueda de descripciones más compactas de los datos), modelado de dependencias (dependencias entre las variables de los datos), detección de desviaciones (búsqueda de desviaciones importantes de los datos respecto de valores anteriores o medios).
- Dependiendo de la tendencia con que se aborde el problema, se pueden distinguir tres grandes líneas de investigación o paradigmas: sistemas conexionistas (redes neuronales), sistemas evolucionistas (algoritmos genéticos) y sistemas simbólicos.
- Dependiendo del lenguaje utilizado para representar del conocimiento, se pueden distinguir: representaciones basadas en la lógica de proposiciones, representaciones basadas en lógica de predicados de primer orden, representaciones estructuradas, representaciones a través de ejemplos y representaciones no simbólicas como las redes neuronales.

A continuación, describiremos con más detalle los diferentes métodos de representación del conocimiento que se emplean en la Minería de Datos, dado que el lenguaje de representación es uno de los aspectos importantes para el proceso de KDD (Beltrán, 2018).

#### **2.7.1.1 Agrupamiento ("Clustering")**

También llamada Segmentación, esta herramienta permite la identificación de tipologías o grupos donde los elementos guardan similitud entre sí y diferencias con aquellos de otros grupos. Para alcanzar las distintas tipologías o grupos existentes en una base de datos, estas herramientas requieren, como entrada, información sobre el colectivo a segmentar. Esta información corresponderá a los valores concretos, para cada elemento en un momento del tiempo, de una serie de variables ("Segmentación estática") o a través del comportamiento en el tiempo de cada uno de los elementos del colectivo ("Segmentación dinámica").

Como resultado del tratamiento de la información, estas herramientas presentan los distintos grupos detectados junto con los valores característicos de las variables. Este tipo de herramientas se basan en técnicas de carácter estadístico, de empleo de algoritmos matemáticos, de generación de reglas y de redes neuronales para el tratamiento de registros. Para otro tipo de elementos a agrupar o segmentar, como texto y documentos, se usan técnicas de reconocimiento de conceptos (Beltrán, 2018).

#### **2.7.1.2 Asociación (" Association Pattern Discovery")**

Este tipo de herramientas establece las posibles relaciones o correlaciones entre distintas acciones o sucesos aparentemente independientes, pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros.

Normalmente, este tipo de herramientas se fundamenta en técnicas estadísticas como los análisis de correlación y de variación (Beltrán, 2018).

### **2.7.1.3 Secuenciamiento ("Sequential Pattern Discovery")**

Esta herramienta permite identificar como, en el tiempo, la ocurrencia de una acción desencadena otras posteriormente. Es muy similar a la anteriormente analizada, si bien, en este caso, el tiempo es una variable crítica e imprescindible a introducir en la información a analizar (Beltrán, 2018).

### **2.7.1.4 Reconocimiento de Patrones ("Pattern Matching")**

Estas herramientas permiten la asociación de una señal, información de entrada con aquella o aquellas con las que guarda mayor similitud y que están catalogadas en el sistema.

Estas herramientas son usadas por elementos que son tan habituales como un procesador de texto o un despertador. Los patrones pueden ser cualquier elemento de información que deseemos.

En el ámbito particular del DM estas herramientas pueden ayudarnos en la identificación de problemas e incidencias y de sus posibles soluciones, toda vez que dispongamos de la base de información necesaria en la cual buscar

Estas herramientas se sustentan en las técnicas de Redes Neuronales y Algoritmos Matemáticos (Beltrán, 2018).

### **2.7.1.5 Previsión ("Forecasting")**

La Previsión establece el comportamiento futuro más probable dependiendo de la evolución pasada y presente. Esta herramienta tiene su uso fundamental en el tratamiento de Series Temporales y las técnicas asociadas disponen de una importante madurez. Las herramientas de Previsión utilizan bien la propia información histórica, o bien, la información histórica relativa a otras variables de las cuales la primera depende (Beltrán, 2018).

### **2.7.1.6 Simulación**

Las herramientas de Simulación forman parte también del conjunto de herramientas veteranas de la investigación científica. Como ejemplo están las herramientas de diseño y producción asistidas por ordenador, "CAD" - "CAM", en las cuales se revisan los diseños sometiéndoles a una amplísima serie de condiciones reales normales y extremas.

Ello permite no solo ajustar y adaptar el diseño, sino posteriormente establecer márgenes y límites de funcionamiento.

La simulación se puede definir como la generación de múltiples escenarios o posibilidades sujetas, normalmente, a unas reglas o esquemas con el objeto de analizar la idoneidad y comportamiento de una decisión o prototipo en un marco de posibles condiciones futuras o para analizar todas las posibles variaciones o alternativas a una decisión o situación y también se usa para el cálculo numérico (Beltrán, 2018).

### **2.7.1.7 Optimización**

Al igual que la Previsión y la Simulación, las herramientas de Optimización tienen una amplia tradición de uso.

La optimización ha sido y es extensivamente usada en la resolución de los problemas asociados a la logística de distribución y a la gestión de "Stocks" en los negocios y en la determinación de parámetros teóricos a partir de los experimentos en la investigación científica.

La optimización resuelve el problema de la minimización o maximización de una función que depende de una serie de variables, encontrando los valores de estas que satisfacen esa condición de máximo, típicamente beneficios, o mínimo, normalmente costes.

Habitualmente estos problemas conllevan, adicionalmente, una serie de "ligaduras" o restricciones de forma que no todas las posibles soluciones son aceptables, ello se traduce en que debemos reducir nuestro universo de búsqueda a aquellas soluciones que satisfagan tales restricciones (Beltrán, 2018).

#### **2.7.1.8 Clasificación (“Clasification”, “Prediction” o “Scoring”)**

La clasificación agrupa todas aquellas herramientas que permiten asignar a un elemento la pertenencia a un grupo o clase. Ello se instrumenta a través de la dependencia de la pertenencia a las clases en los valores de una serie de atributos o variables.

A través del análisis de un colectivo de elementos, o casos de los cuales conocemos la clase a la que pertenecen, se establece un mecanismo que establece la pertenencia a tales clases en función de los valores de las distintas variables y nos permite establecer el grado de discriminación o influencia de estas.

También se utiliza para estas herramientas la denominación de Predicción o Evaluación para aquellos casos donde se aplican técnicas, normalmente numéricas, que establecen para cada elemento un valor dependiente de los valores que tengan las variables en tal elemento.

Las herramientas de Clasificación hacen uso de técnicas como algoritmos matemáticos, análisis discriminante y de variaciones, sistemas expertos y sistemas de conocimiento e inducción de reglas.

Como se ha podido apreciar, normalmente es necesaria la conjunción e integración de varios tipos de herramientas a efectos de brindar una solución completa a nuestros problemas (Beltrán, 2018).

### Métodos apropiados

- No estructurados:
  - Métodos bayesianos.
  - Otros métodos estadísticos
  - Métodos relacionales
- Semiestructurados
  - Gramaticales.
  - Métodos relacionales con constructores.

### Métodos no apropiados

Sin una profunda transformación de los datos, muchas técnicas de aprendizaje automático son útiles para muchas aplicaciones.

- Métodos de clasificación (árboles de decisión) están basados. En una clase dependiente de un número de atributos predeterminados.
- Métodos numéricos (regresión, redes neuronales) los datos son simbólicos, no numéricos.
- Métodos por casos (KNN, CBR) tiempos de respuesta serían muy altos.

#### **2.7.2 Técnicas de la Minería de Datos**

La Minería de Datos ha dado lugar a una paulatina sustitución del análisis de datos dirigido a la verificación por un enfoque de análisis de datos dirigido al descubrimiento del conocimiento. La principal diferencia entre ambos se encuentra en que en el último se descubre información sin necesidad de formular previamente una hipótesis. La aplicación automatizada de algoritmos de Minería de Datos permite detectar fácilmente patrones en los datos, razón por la cual esta técnica es mucho más eficiente que el análisis dirigido a la

verificación cuando se intenta explorar datos procedentes de repositorios de gran tamaño y complejidad elevada. Dichas técnicas emergentes se encuentran en continua evolución como resultado de la colaboración entre campos de investigación tales como bases de datos, reconocimiento de patrones, inteligencia artificial, sistemas expertos, estadística, visualización, recuperación de información, y computación de altas prestaciones.

Como mencionamos al principio, los algoritmos de Minería de Datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o de descubrimiento del conocimiento.

En la tabla siguiente se muestran algunas de las técnicas de Minería de Datos en ambas categorías:

**Tabla 4.**

*Clasificación de las técnicas de Minería de Datos*

| <b>SUPERVISADOS</b> | <b>NO SUPERVISADOS</b>    |
|---------------------|---------------------------|
| Árboles de decisión | Detección de Desviaciones |
| Inducción neuronal  | Segmentación              |
| Regresión           | Agrupamiento (Clustering) |
| Series temporales   | Reglas de Asociación      |
|                     | Patrones Secuenciales     |

*Nota.* Técnicas de la Minería de Datos (Beltrán, 2018).

La aplicación de los algoritmos de Minería de Datos requiere la realización de una serie de actividades previas encaminadas a preparar los datos de entrada debido a que, en muchas ocasiones, dichos datos proceden de fuentes heterogéneas, no tienen el formato adecuado o contienen ruido. Por otra parte, es necesario interpretar y evaluar los resultados obtenidos.



La siguiente tabla muestra algunas de las técnicas más comunes de Minería de Datos y a continuación describiremos cada una de ellas:

**Tabla 5.**

*Técnicas del Data Mining*

|                                  |                                      |
|----------------------------------|--------------------------------------|
|                                  | ANOVA                                |
|                                  | Prueba Ji cuadrado.                  |
|                                  | Análisis de componentes principales. |
| Métodos Estadísticos             | Análisis de Clusters.                |
|                                  | Análisis discriminante.              |
|                                  | Regresión Lineal.                    |
|                                  | Regresión Logística.                 |
| Árboles de decisión.             | CHAID.                               |
|                                  | CART.                                |
| Reglas de Asociación.            |                                      |
| Reglas de neuronas artificiales. |                                      |
| Algoritmos genéticos.            |                                      |
| Otros.                           | Lógica Difusa.                       |
|                                  | Series Temporales.                   |

*Nota.* Listado de Técnicas de la Minería de Datos (Beltrán, 2018).

### **2.7.2.1 Métodos Estadísticos**

La estadística según (Beltrán, 2018), es tradicionalmente la técnica que se ha usado para el tratamiento de grandes volúmenes de datos numéricos y nadie pone en duda su efectividad al poseer un amplísimo conjunto de modelos de análisis para cubrir el tratamiento de todo tipo de poblaciones y series de datos. Estos son algunos de los métodos estadísticos más utilizados:

- **ANOVA:** análisis de la Varianza, contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupos de población distintos.
- **Ji cuadrado:** contrasta la hipótesis de independencia entre variables.
- **Componentes principales:** Permite reducir el número de variables observadas a un menor número de variables artificiales, conservando la mayor parte de la información sobre la varianza de las variables.
- **Análisis de clusters:** permite clasificar una población en un número determinado de grupos, sobre la base de semejanzas y diferencias de perfiles existentes entre los diferentes componentes de dicha población.
- **Análisis discriminante:** método de clasificación de individuos en grupos que previamente se han establecido, y que permite encontrar la regla de clasificación de los elementos de estos grupos, por tanto, identificar cuáles son las variables que mejor definan la pertenencia al grupo.
- **Regresión Lineal:** técnica más básica del Data Mining. Un modelo de regresión lineal se implementa identificando una variable dependiente ( $y$ ) y todas las variables independientes ( $X_1, X_2, \dots$ ). Se asume que la relación entre estas y aquella es lineal. Todas las variables han de ser continuas. El resultado es la ecuación de la recta que mejor se ajusta al juego de datos y esta ecuación se interpreta o se usa para predicción.
- **Regresión Logística:** puede trabajar con variables discretas. También requiere que todas las variables sean lineales.

### **2.7.2.2 Métodos Basados en Árboles de Decisión**

Son herramientas analíticas empleadas para el descubrimiento de reglas y relaciones mediante la ruptura y subdivisión sistemática de la información contenida en el conjunto de datos. El árbol de decisión se construye partiendo el conjunto de datos en dos (CART) o más (CHAID) subconjuntos de observaciones a partir de los valores que toman las variables predictoras. Cada uno de estos subconjuntos vuelve después a ser particionado utilizando el mismo algoritmo.

Este proceso continúa hasta que no se encuentran diferencias significativas en la influencia de las variables de predicción de uno de estos grupos hacia el valor de la variable de respuesta.

La raíz del árbol es el conjunto de datos íntegro, los subconjuntos y los subconjuntos conforman las ramas del árbol. Un conjunto en el que se hace una partición se llama nodo.

El método CHAID (Chi Squared Automatic Interaction Detector) es útil en aquellas situaciones en las que el objetivo es dividir una población en distintos segmentos basándose en algún criterio de decisión (Beltrán, 2018).

### **2.7.2.3 Reglas de Asociación**

Derivan de un tipo de análisis que extrae información por coincidencias. Este análisis, a veces llamado "cesta de la compra" permite descubrir correlaciones en los sucesos de la base de datos a analizar y se formaliza en la obtención de reglas de tipo; SI ... ENTONCES... (Beltrán, 2018).

#### **2.7.2.4 Redes Neuronales ("Neural Networks")**

Las Redes Neuronales constituyen una técnica inspirada en los trabajos de investigación, iniciados en 1930, que pretendían modelar computacionalmente el aprendizaje humano llevado a cabo a través de las neuronas en el cerebro.

Las redes neuronales son una nueva forma de analizar la información con una diferencia fundamental con respecto a las técnicas tradicionales: son capaces de detectar y aprender patrones y características dentro de los datos.

Se comportan de forma parecida a nuestro cerebro, aprendiendo de la experiencia y el pasado y aplicando tal conocimiento a la resolución de problemas nuevos.

Una vez adiestradas, las redes neuronales pueden hacer previsiones, clasificaciones y segmentación.

Las redes neuronales se construyen estructurando en una serie de niveles o capas compuesta por nodos o "neuronas". Poseen dos formas de aprendizaje derivadas del tipo de paradigma que usan: el supervisado y el no supervisado.

Son métodos de proceso numérico en paralelo que tratan de modelizar el funcionamiento del cerebro. La red asigna pesos al azar a cada variable independiente y determina si existe algún patrón, predictivo en los datos. Una vez que encuentra un patrón la red lo optimiza reforzando los pesos de las variables y comparando con los datos del grupo de validación. Luego prosigue el proceso y aprende de los resultados una y otra vez. Finalmente, se puede aplicar el modelo aprendido a cualquier nuevo conjunto de datos de entrada. Pueden manejar datos continuos y discretos, lineales y no-lineales simultáneamente. El único inconveniente que presentan es que no genera una ecuación o modelo que explique el

comportamiento del sistema, siendo muy difícil determinar la influencia de cada variable en el comportamiento global del sistema (Beltrán, 2018).

#### **2.7.2.5 Algoritmos Genéticos (“Genetic Algorithms”)**

Los Algoritmos Genéticos son otra técnica que debe su inspiración, de nuevo, a la Biología como las Redes Neuronales. Estos algoritmos representan la modelización matemática de como los cromosomas en un marco evolucionista alcanzan la estructura y composición más óptima en aras de la supervivencia. Entendiendo la evolución como un proceso de búsqueda y optimización de la adaptación de las especies que se plasma en mutaciones y cambios en los genes o cromosomas. Los Algoritmos Genéticos hacen uso de las técnicas biológicas de reproducción (mutación y cruce) para ser utilizadas en todo tipo de problemas de búsqueda y optimización. Esta aproximación está enfocada a problemas de optimización. Se comienza con una población de partida y se va alterando y optimizando su composición para la solución de un problema particular mediante mecanismos tomados de la teoría de la evolución (introducir elementos aleatorios para la modificación de las variables o mutaciones).

El material genético o información de los individuos puede ser transmitido a las siguientes generaciones, de diferentes formas, que van optimizando el proceso. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización. Esta herramienta se usa en las primeras fases del Data Mining, para seleccionar las variables que luego se emplearán con otra técnica, como las redes de neuronas o la regresión logística (Beltrán, 2018).

### **2.7.2.6 Lógica Difusa ("fuzzy logic")**

La Lógica Difusa surge de la necesidad de modelizar la realidad de una forma más exacta, evitando precisamente el determinismo o la exactitud. La Lógica permite el tratamiento probabilístico de la categorización de un colectivo. La Lógica Difusa es aquella técnica que permite y trata la existencia de barreras difusas o suaves entre los distintos grupos en los que categorizamos un colectivo o entre los distintos elementos, factores o proporciones que concurren en una situación o solución (Beltrán, 2018).

### **2.7.2.7 Series Temporales**

Consisten en el estudio de una variable a través del tiempo para, a partir de ese conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones. Suelen basarse en un estudio de la serie en ciclos, tendencias y estacionalidades, que se diferencian por el ámbito de tiempo abarcado, para, por composición, obtener la serie original. Se pueden aplicar enfoques híbridos con los métodos anteriores, en los que la serie se puede explicar no solo en función del tiempo, sino como combinación de otras variables de entorno más estables y, por lo tanto, más fácilmente predecibles (Beltrán, 2018).

### **2.7.2.8 Redes Bayesianas**

Las redes bayesianas son una alternativa para Minería de Datos, la cual tiene varias ventajas:

- Permiten aprender sobre relaciones de dependencia y causalidad.
- Permiten combinar conocimiento con datos.
- Evitan el sobre ajuste de los datos.
- Pueden manejar bases de datos incompletos.

El obtener una red bayesiana a partir de datos es un proceso de aprendizaje, el cual se divide, naturalmente, en dos aspectos:

1. **Aprendizaje paramétrico:** dada una estructura, obtener las probabilidades a priori y condicionales requeridas
2. **Aprendizaje estructural:** obtener la estructura de la red Bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas.

Las técnicas de aprendizaje estructural dependen del tipo de estructura de red: árboles, poliárboles y redes multiconectadas. Otra alternativa es combinar conocimiento subjetivo del experto con aprendizaje. Para ello se parte de la estructura dada por el experto, la cual se valida y mejora utilizando datos estadísticos (Beltrán, 2018).

### **2.7.2.9 Inducción de Reglas**

Las técnicas de Inducción de Reglas surgieron hace dos décadas y permiten la generación y contraste de árboles de decisión o reglas y patrones a partir de los datos de entrada.

Como información de entrada, tendremos un conjunto de casos donde se ha asociado una clasificación o evaluación a un conjunto de variables o atributos.

Con tal información, estas técnicas obtienen el árbol de decisión o conjunto de reglas que soportan la evaluación o clasificación.

En los casos en que la información de entrada posee algún tipo de "ruido" o defecto, estas técnicas pueden habilitar métodos estadísticos de tipo probabilístico para generar, en estos casos, árboles de decisión podados o recortados (Beltrán, 2018).

### **2.7.2.10 Sistemas Basados en el Conocimiento y Sistemas Expertos**

#### **("Knowledge Based Systems" & "Expert Systems")**

Estos sistemas son un clásico de la Inteligencia Artificial. Estas técnicas permiten la formalización de árboles y reglas de decisión extraídas de la formalización del conocimiento de los expertos.

Poseen motores llamados "Motores de Inferencia" que se encargan de gestionar las distintas preguntas al ser realizadas de forma que el proceso de decisión sea lo más eficiente y rápido posible. (Beltrán, 2018).

### **2.7.2.11 Algoritmos Matemáticos**

Sin llegar a ser técnicas que den soporte a unas necesidades concretas como las anteriores, existe una amplia gama de algoritmos matemáticos que son especialmente útiles y eficaces en la resolución y tratamiento de problemas muy específicos y puntuales y que, normalmente, son incorporados en alguna de aquellas técnicas con el objeto de mejorarlas (Beltrán, 2018).

## **2.8 APLICACIONES DE LA MINERÍA DE DATOS**

Algunas de las tareas importantes de la Minería de datos incluyen la identificación de aplicaciones para las técnicas existentes, y desarrollar nuevas técnicas para dominios tradicionales o de nueva aplicación, como el comercio electrónico y la bioinformática.

Existen numerosas áreas donde la Minería de Datos se puede aplicar, prácticamente en todas las actividades humanas que generen datos:

- Comercio y banca: segmentación de clientes, previsión de ventas, análisis de riesgo.



- Medicina y Farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos.
- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Astronomía: identificación de nuevas estrellas y galaxias.
- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites.
- Ciencias Ambientales: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales, plantas depuradoras de aguas residuales para mejorar su observación, gestión y/o control.
- Ciencias Sociales: estudio de los flujos de la opinión pública. Planificación de ciudades: identificar barrios con conflicto en función de valores sociodemográficos.

En la actualidad se puede afirmar que la MD ha demostrado la validez de una primera generación de algoritmos mediante diferentes aplicaciones al mundo real. Sin embargo, estas técnicas todavía están limitadas por bases de datos simples, donde los datos se describen mediante atributos numéricos o simbólicos, no conteniendo atributos de tipo texto o imágenes, y los datos se preparan con una tarea concreta en mente. Sobrepasar este límite será un reto a conseguir (Riquelme y otros, 2006).

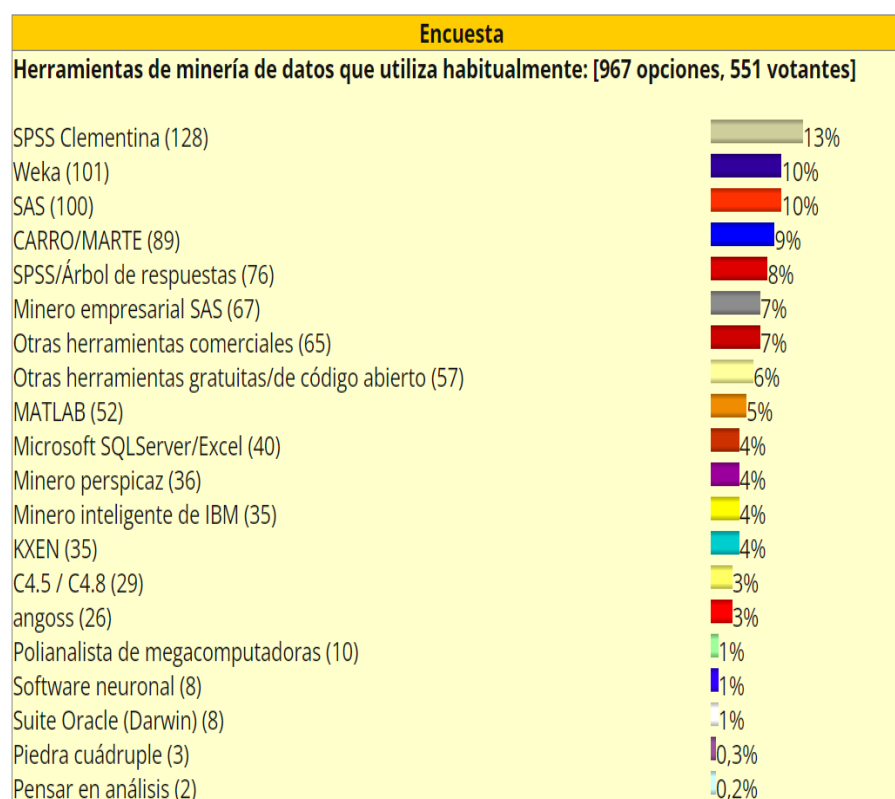
## 2.9 HERRAMIENTAS DE LA MINERÍA DE DATOS

En la figura que se mostrara de la investigación que realizo (Martínez de Pisón Ascacibar, 2003), podemos apreciar, el resultado de una encuesta hecha en el conocido portal sobre Minería de Datos y Gestión del Conocimiento, donde se pregunta al encuestado sobre la herramienta de Data Mining que habitualmente usa.

Este tipo de encuesta es particularmente importante, porque nos da una idea de las aplicaciones que más están usando los profesionales y nos puede ayudar a decidir correctamente cuando tengamos que adquirir uno de estos programas.

### Figura 6.

*Herramientas más usadas de la Minería de Datos*



*Nota.* Encuesta realizada en junio del 2002 Herramientas de la Minería de Datos (Martínez de Pisón Ascacibar, 2003, p. 68).

### 2.9.1 *Weka*

En 1993, la Universidad de Waikato en Nueva Zelanda, inició el desarrollo de la versión original de WEKA, y no fue hasta 1997 que se decidió reescribir su código en java, incluyendo la implementación de nuevos algoritmos de modelado. En el 2005, WEKA recibe el galardón “Data Mining and Knowledge Discovery Services” (Servicios de Minería de Datos y Descubrimiento del Conocimiento).

Waikato Environment for Knowledge Analysis (WEKA), es un entorno para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos. Para ello únicamente se requiere que los datos a analizar se almacenen con un cierto formato, conocido como Attribute-Relation File Format (ARFF).

WEKA es un software de libre distribución desarrollado en Java. Está constituido por una serie de paquetes de código abierto con diferentes tareas como clasificación, agrupamiento, asociación, visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. Estos paquetes pueden ser integrados en cualquier proyecto de análisis de datos, e incluso pueden extenderse con contribuciones de los usuarios que desarrollen nuevos algoritmos. Con objeto de facilitar su uso por un mayor número de usuarios, WEKA además incluye una interfaz gráfica de usuario para acceder y configurar las diferentes herramientas integradas.

WEKA se distribuye como un archivo ejecutable comprimido de java (archivo ".jar"), que se invoca directamente sobre la máquina virtual JVM. En las primeras versiones de WEKA se requería la máquina virtual Java 1.2 para invocar a la interfaz gráfica, desarrollada con el paquete gráfico de Java Swing. En el caso de la última versión, WEKA 3-4, se requiere Java 1.3 o superior. WEKA se distribuye como software de libre distribución desarrollado en Java.

Está constituido por una serie de paquetes de código abierto con diferentes tareas como clasificación, agrupamiento, asociación, y visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. Estos paquetes pueden ser integrados en cualquier proyecto de análisis de datos, e incluso pueden extenderse con contribuciones de los usuarios que desarrollen nuevos algoritmos. Con objeto de facilitar su uso por un mayor número de usuarios, WEKA además incluye una interfaz gráfica de usuario para acceder y configurar las diferentes herramientas integradas (Sagrario, 2011, p. 37).

### **2.9.1.1 Características de Weka**

WEKA es una extensa colección de algoritmos de Máquinas de conocimiento desarrollados por la universidad de Waikato (Nueva Zelanda) implementados en Java, útiles para ser aplicados sobre datos mediante las interfaces que ofrece o para embeberlos dentro de cualquier aplicación. Además, WEKA contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. WEKA está diseñado como una herramienta orientada a la extensibilidad, por lo que añadir nuevas funcionalidades es una tarea sencilla.

Sin embargo, y pese a todas las cualidades que WEKA posee, tiene un gran defecto y este es la escasa documentación orientada al usuario que conlleva una utilización bastante pobre, lo que la hace una herramienta difícil de comprender y manejar sin información adicional.

La licencia de WEKA es GPL\*, lo que significa que este programa es de libre distribución y difusión. Además, ya que WEKA está programado en Java, es independiente de la arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible (Sagrario, 2011, p. 38).

### **2.9.2 Rapid Miner**

La versión inicial fue desarrollada por el departamento de inteligencia artificial de la Universidad de Dortmund en 2001. Se distribuye bajo licencia GPL y está hospedado en SourceForge desde el 2004.

RapidMiner proporciona más de 500 operadores orientados al análisis de datos, incluyendo los necesarios para realizar operaciones de entrada y salida, preprocesamiento de datos y visualización.

RapidMiner es un software de distribución libre usado más comúnmente por las empresas, a diferencia de WEKA que es un software para aprendizaje sobre Minería de Datos, por lo cual es considerado por algunos especialistas como el líder mundial de código abierto en Minería de Datos; debido a la combinación de tecnología de primera calidad y su rango de funcionalidad. Además de ser una herramienta flexible para aprender y explorar la Minería de Datos, la interfaz gráfica de usuario tiene como objetivo simplificar el uso para las tareas complejas de esta área.

RapidMiner (anteriormente, YALE, Yet Another Learning Environment) es un programa informático para máquinas de aprendizaje y Minería de Datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. El concepto modular de Rapidminer permite que el operador diseñe y anide un gran número de cadenas para complejos problemas de aprendizaje, debido a que el manejo interno de datos es transparente para el usuario; RapidMiner introduce nuevos conceptos de transparencia del manejo de datos y modelado de procesos, en donde la configuración final del proceso está dada por el usuario (Sagrario, 2011, p. 38).

### 2.9.2.1 Características de Rapid Miner

Según (Sagrario, 2011, p. 39), su aplicación más común es en la investigación, y algunas de sus características principales son las siguientes:

- Se encuentra desarrollado en Java
- Es un sistema multiplataforma
- Representación interna de los procesos de análisis de datos en archivos XML
- Permite el desarrollo de programas a través de un lenguaje de script
- Puede usarse de diversas maneras:
  - A través de una GUI
  - En línea de comandos
  - En batch o Desde otros programas a través de llamadas a sus librerías
- Es extensible
- Incluye gráficos y herramientas de visualización de datos.

### 2.9.3 *R Development*

R es un conjunto integrado de programas para manipulación de datos, cálculo y gráficos. Entre otras características dispone de:

- Almacenamiento y manipulación efectiva de datos,
- Operadores para cálculo sobre variables indexadas (Arrays), en particular matrices,
- Una amplia, coherente e integrada colección de herramientas para análisis de datos,

- Posibilidades gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora, y
- Un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas. (Debe destacarse que muchas de las funciones suministradas con el sistema están escritas en el lenguaje R).

El término “entorno” lo caracteriza como un sistema completamente diseñado y coherente, antes que como una agregación incremental de herramientas muy específicas e inflexibles, como ocurre frecuentemente con otros programas de análisis de datos. R es en gran parte un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos. Como tal es muy dinámico y las diferentes versiones no siempre son totalmente compatibles con las anteriores. Algunos usuarios prefieren los cambios debido a los nuevos métodos y tecnología que los acompañan, a otros, sin embargo, les molesta porque algún código anterior deja de funcionar. Aunque R puede entenderse como un lenguaje de programación, los programas escritos en R deben considerarse esencialmente efímeros. (González & González, 2000, p. 8).

### **2.9.3.1 Características del Lenguaje R**

Se trata de un proyecto de software libre, resultado de la implementación del Lenguaje S. Probablemente, R es el lenguaje más utilizado en investigación por la comunidad estadística. A esto contribuye la posibilidad de cargar diferentes Librerías o Paquetes con finalidades específicas de cálculo o gráfico. (Ximénes & Revuelta, 2022, p. 4).

Sus principales características son las siguientes:

- R es un lenguaje de programación eficaz que permite hacer todo tipo de análisis estadísticos.
- Aunque en un principio R puede resultar poco amigable, ya que requiere redactar códigos, su sintaxis es relativamente fácil de aprender.
- R permite combinar fácilmente librerías de código R y de otros programas y ofrece gráficos de alta calidad.
- R es gratuito. Por tanto, se utiliza sin ningún coste. Esto hace que su alcance sea mucho mayor que el de cualquier otro programa comercial.
- R se está actualizando permanentemente, con las aportaciones de la comunidad científica de expertos en Estadística. Esto obliga al usuario a actualizar el programa de forma permanente también en su equipo.
- R tiene un Menú de Ayuda muy completo. Además, en la web pueden encontrarse múltiples tutoriales y videos donde se explica el manejo de muchas de las funcionalidades y librerías del programa.
- R no dispone de un soporte comercial, pero se actualiza permanentemente por miles de usuarios en todo el mundo, lo que permite contactar online y consultar directamente con el autor o autora de la función o librería en cuestión, aquellas dudas que tengamos.

#### **2.9.4 Orange Data Minig**

Orange nació de un proyecto como proyecto de la Universidad de Liubiana hace más de 20 años, desarrollado inicialmente en C++, para después ampliarse con Python. Orange ofrece aplicaciones de gran utilidad para el análisis de datos y de texto, así como



características de aprendizaje automático, trabajando con operadores para la clasificación, regresión y clustering a través de una programación visual la cual facilita a los usuarios en el proceso de datamining para tomar decisiones rápidamente en el ámbito profesional (Gutierrez Alvarez, 2019, p. 65).

### **2.9.5 *Knime***

El software KNIME (Konstanz Information Miner), desarrollado por la universidad de Constanza, está basado en Java y preparado con Eclipse, y se destaca por contar con una amplia gama de módulos y paquetes los cuales permiten descubrir estructuras ocultas de datos. Dentro de las bondades de KNIME se tiene: la integración de numerosos procedimientos de aprendizaje automático y de datamining, y la eficiencia en el tratamiento previo de los datos (Gutierrez Alvarez, 2019, p. 65).

### **2.9.6 *Python***

Python es un lenguaje de programación creado por Guido Van Rossum a principios de los años noventa, es un lenguaje que posee una sintaxis clara y estructurada que favorece a generar un código legible. Es administrado por la Python Software Foundation, posee una licencia de código abierto, denominada Python Software Foundation License, que es compatible con la Licencia pública general de GNU a partir de la versión 2.1.1. Se trata de un lenguaje interpretado o de script, con tipado dinámico, multiplataforma y orientado a objetos, además cuenta con una gran cantidad de librerías y frameworks disponibles. La sintaxis de Python es sencilla y cercana al lenguaje natural. Por estas razones se trata de uno de los mejores lenguajes para comenzar a programar, ya que es como programar en pseudocódigo, y la curva de aprendizaje es bastante alta y rápida. Es una tecnología muy utilizada en la actualidad, por su simplicidad y potencia permite realizar aplicaciones de una manera rápida, sencilla y óptima para el despliegue en distintas plataformas, además permite utilizar algoritmos de machine learning e inteligencia artificial en producción.

Existen muchos casos de grandes empresas que utilizan Python en sus aplicaciones con gran éxito, tal es el caso de Google, la NASA, Netflix, Mercado libre, Spotify, Dropbox, Instagram, Pinterest, Globant, Sattelogic.

Un lenguaje interpretado o de script es aquel que se ejecuta utilizando un programa intermedio llamado intérprete, en lugar de compilar el código a lenguaje máquina que pueda comprender y ejecutar directamente una computadora. La ventaja de los lenguajes compilados es que su ejecución es más rápida. Sin embargo, los lenguajes interpretados son más flexibles y más portables.

Python tiene, no obstante, muchas de las características de los lenguajes compilados, por lo que se podría decir que es semi interpretado. En Python, como en Java, Delphi y muchos otros lenguajes, el código fuente se traduce la primera vez que se ejecuta a un pseudocódigo máquina intermedio llamado bytecode, generando archivos “.pyc” o “.pyo”, que son los que se ejecutarán en sucesivas ocasiones.

La característica de tipado dinámico se refiere a que no es necesario declarar el tipo de dato que va a contener una variable, sino que su tipo se determinará en tiempo de ejecución según el tipo del valor al que se asigne, y el tipo de esta variable puede cambiar si se le asigna un valor de otro tipo, lo cual le otorga una gran flexibilidad a la hora de programar.

En los lenguajes con esta característica no se permite tratar a una variable como si fuera de un tipo distinto al que tiene, es necesario convertir de forma explícita dicha variable al nuevo tipo previamente. Por ejemplo, si tenemos una variable que contiene un texto (variable de tipo cadena o string) no podremos tratarla como una variable numérica, sin previamente realizar una conversión.

La programación orientada a objetos es un paradigma en el que los conceptos del mundo real relevantes para nuestro problema se trasladan a clases y objetos en nuestro programa. La ejecución del programa consiste en una serie de interacciones e intercambio de mensajes entre los objetos. No obstante, el lenguaje soporta también los paradigmas de programación imperativa y funcional (este último le permite al lenguaje añadir características avanzadas muy interesantes).

El intérprete de Python está disponible en muchas plataformas de las más comunes (Linux, UNIX, Mac OS, Solaris, DOS, Windows, etc.) por lo que si no utilizamos librerías específicas de cada plataforma nuestro programa podrá correr en todos estos sistemas sin tener que realizar ningún tipo de cambio. Esto nos permite desarrollar programas que sean portables de un sistema operativo a otro.

Los creadores de Python entienden que la gran sencillez, flexibilidad y potencia del lenguaje debe ir acompañada de una correcta formación del programador que lo utiliza para poder realizar las actividades de desarrollo con criterio, dado que la inexperiencia de los nuevos programadores puede conducirlos a programar de manera incorrecta, aunque lleguen a un resultado (Bel, 2020, p. 16).

## **2.10 MÉTODO CIENTÍFICO**

El método científico, según (Bernal Torres, 2010, p. 58), se entiende como el conjunto de postulados, reglas y normas para el estudio y la solución de los problemas de investigación, institucionalizados por la denominada comunidad científica reconocida. En un sentido más global, el método científico se refiere al conjunto de procedimientos que, valiéndose de los instrumentos o técnicas necesarias, examina y soluciona un problema o conjunto de problemas de investigación.

En la investigación científica han predominado, a lo largo de la historia, tres métodos científicos básicos:

1. El baconiano, que postula el desarrollo de la inducción.
2. El galileano, que postula la experimentación.
3. El cartesiano, que postula la duda fundamentada en el análisis y la síntesis de los problemas.

En la actualidad, sin embargo, dada la diversidad de escuelas y paradigmas investigativos, estos métodos se han complementado y es frecuente reconocer, entre otros, métodos como los siguientes:

**Método deductivo:** este método de razonamiento consiste en tomar conclusiones generales para obtener explicaciones particulares. El método se inicia con el análisis de los postulados, teoremas, leyes, principios, etcétera, de aplicación universal y de comprobada validez, para aplicarlos a soluciones o hechos particulares.

**Método inductivo:** este método utiliza el razonamiento para obtener conclusiones que parten de hechos particulares aceptados como válidos, para llegar a conclusiones cuya aplicación sea de carácter general. El método se inicia con un estudio individual de los hechos y se formulan conclusiones universales que se postulan como leyes, principios o fundamentos de una teoría.

**Método inductivo-deductivo:** este método de inferencia se basa en la lógica y estudia hechos particulares, aunque es deductivo en un sentido (parte de lo general a lo particular) e inductivo en sentido contrario (va de lo particular a lo general).

**Método hipotético-deductivo:** consiste en un procedimiento que parte de unas aseveraciones en calidad de hipótesis y busca refutar o falsear tales hipótesis, deduciendo de ellas conclusiones que deben confrontarse con los hechos.

**Método analítico:** este proceso cognoscitivo consiste en descomponer un objeto de estudio, separando cada una de las partes del todo para estudiarlas en forma individual.

**Método sintético:** integra los componentes dispersos de un objeto de estudio para estudiarlos en su totalidad.

**Método analítico-sintético:** estudia los hechos, partiendo de la descomposición del objeto de estudio en cada una de sus partes para estudiarlas en forma individual (análisis), y luego se integran esas partes para estudiarlas de manera holística e integral (síntesis).

**Método histórico-comparativo:** procedimiento de investigación y esclarecimiento de los fenómenos culturales que consiste en establecer la semejanza de esos fenómenos, infiriendo una conclusión acerca de su parentesco genético, es decir, de su origen común.

**Métodos de investigación cualitativa y cuantitativa:** otra forma reciente de caracterizar métodos de investigación es la concepción de métodos cimentada en las distintas concepciones de la realidad social, en el modo de conocerla científicamente y en el uso de herramientas metodológicas que se emplean para analizarla. Según esta concepción, el método de investigación suele dividirse en los métodos cuantitativo o investigación cuantitativa, y cualitativo o investigación cualitativa.

**Método cuantitativo o método tradicional:** se fundamenta en la medición de las características de los fenómenos sociales, lo cual supone derivar de un marco conceptual pertinente al problema analizado, una serie de postulados que expresen relaciones entre las

variables estudiadas de forma deductiva. Este método tiende a generalizar y normalizar resultados.

**El método cualitativo o método no tradicional:** se orienta a profundizar casos específicos y no a generalizar. Su preocupación no es prioritariamente medir, sino cualificar y describir el fenómeno social a partir de rasgos determinantes, según sean percibidos por los elementos mismos que están dentro de la situación estudiada.

Los investigadores que utilizan el método cualitativo buscan entender una situación social como un todo, teniendo en cuenta sus propiedades y su dinámica. En su forma general, la investigación cuantitativa parte de cuerpos teóricos aceptados por la comunidad científica, en tanto que la investigación cualitativa pretende conceptuar sobre la realidad, con base en la información obtenida de la población o las personas estudiadas.

Aunque durante años estos métodos asumieron posiciones antagónicas, en tanto que los cualitativos aún no tienen el amplio reconocimiento general de la comunidad científica clásica, con mayor frecuencia se reconoce que uno y otro tienen puntos fuertes y débiles. Algunos expertos consideran que en una investigación lo indicado es darles prioridad a las técnicas e instrumentos que ofrezcan mayor capacidad para generar conocimiento válido. A menudo se corrobora que, para el estudio de los aspectos sociales, ninguno de los dos métodos tiene validez universal para resolver satisfactoriamente los problemas de investigación (Bernal Torres, 2010, p. 61).

## **2.11 METODOLOGÍAS PARA MINERÍA DE DATOS**

Formalmente, (Moine, 2013, p. 11) indica que, una metodología consiste en un conjunto de actividades organizadas que tienen por objetivo la realización de un trabajo. Para cada actividad se define, además de las entradas y salidas, la forma en la que debe llevarse a cabo. En este capítulo se analizarán los distintos enfoques para la gestión de proyectos de Minería

de Datos más difundidos en la comunidad científica (KDD, SEMMA, CRISP-DM y Catalyst), aunque existen otros experimentales y de menor difusión que crean híbridos con estándares de Ingeniería de Software.

### **2.11.1 KDD**

El Descubrimiento de Conocimiento en Bases de Datos (KDD Knowledge Discovery in Databases) constituye el primer modelo que define el descubrimiento de conocimiento en bases de datos como un “proceso”, compuesto por distintas etapas y fases que van desde la preparación de los datos hasta la interpretación y difusión de los resultados.

En el año 1996, Fayyad define a KDD como el “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia entendibles en los datos”. El término proceso se refiere a la secuencia iterativa de etapas o fases que lo componen. Los patrones deberían ser válidos para nuevos datos, novedosos en el sentido que deberían aportar nuevo conocimiento al dominio de aplicación y potencialmente útiles para el usuario final o tomador de decisiones.

KDD es un proceso iterativo e interactivo. Iterativo ya que la salida de alguna de las fases puede retroceder a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Es interactivo porque el usuario, o más generalmente un experto en el dominio del problema, debe ayudar a la preparación de los datos y validación del conocimiento extraído (Moine, 2013, p. 11).

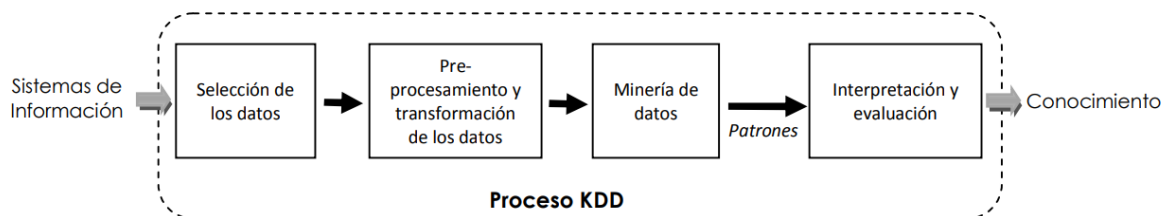
El modelo de proceso KDD se resume en las siguientes cinco fases.

- Selección de los datos sobre los que se trabajará.
- Preprocesamiento de los datos, donde se realiza un tratamiento de los datos incorrectos y ausentes.

- Transformación de los datos y reducción de la dimensionalidad.
- Minería de datos, donde se obtienen los patrones de interés según la tarea de minería que llevemos a cabo (descriptiva o predictiva).
- Interpretación y evaluación del nuevo conocimiento en el dominio de aplicación.

**Figura 7.**

*Proceso KDD*



*Nota.* Esquema resumido del proceso KDD (Moine, 2013, p. 12).

Si bien el proceso KDD puede resumirse en las cinco fases mencionadas anteriormente, Fayyad puntualiza nueve etapas para llevarlo a cabo:

1. **Comprensión del dominio de aplicación:** en esta primera etapa, se debería recolectar todo el conocimiento disponible y relevante sobre el dominio de aplicación e identificar los objetivos del proceso KDD desde el punto de vista del usuario.
2. **Creación del conjunto de datos:** esta etapa consiste en la elección de las fuentes de datos que se utilizarán, la integración de las mismas y la selección de las observaciones/atributos que conformarán la vista minable. Aunque no es estrictamente necesario, en este paso podría requerirse la construcción de un almacén de datos.



3. **Limpieza y preprocesamiento de los datos:** en esta fase se deberían llevar a cabo tareas como limpieza de ruido o datos anómalos (outliers) y tratamiento de datos faltantes (missing values).
4. **Reducción y proyección de los datos:** en este paso se detectan características útiles de representación de los datos dependiendo del objetivo de la tarea de minería (descripción o predicción). Se incluye la utilización de técnicas de reducción de la dimensionalidad y métodos de transformación de los datos para reducir la cantidad de variables en discusión o para encontrar representaciones invariantes de los datos. En esta etapa es frecuente la transformación de los datos, calculando nuevos atributos o bien redefiniendo los existentes con otro formato.
5. **Determinar la tarea de Minería de Datos:** en esta fase, se deberá determinar la tarea de minería con la que se abordará el estudio (como agrupamiento, regresión, clasificación, o asociación) teniendo en cuenta los objetivos definidos en la etapa 1.
6. **Determinar el algoritmo de minería:** de acuerdo a la tarea de minería establecida en el punto anterior, en esta etapa se define el algoritmo (o algoritmos) que se aplicarán para la búsqueda de patrones sobre los datos. Incluye la determinación de qué modelos y parámetros son los más adecuados según la naturaleza del problema y de los datos disponibles.
7. **Minería de Datos:** etapa en la que se aplican los algoritmos y técnicas seleccionadas al conjunto de datos en búsqueda de los patrones de interés.
8. **Interpretación:** comprende la interpretación de los patrones encontrados, visualizando y traduciendo los mismos en términos comprensibles por el usuario.

9. **Utilización del nuevo conocimiento:** en esta fase se implementa el conocimiento descubierto, apoyando con el mismo la toma de decisiones o bien reportándolo a las partes interesadas. Incluye la verificación y resolución de potenciales conflictos con conocimiento descubierto previamente.

Si bien KDD define las fases generales del proceso de Minería de Datos, no especifica qué actividades puntuales hay que realizar en cada una, quedando la definición de las mismas a criterio del equipo de trabajo (Moine, 2013, p. 13).

### 2.11.2 SEMMA

SEMMA, creada por SAS Institute, fue propuesta especialmente para trabajar con el software SAS Enterprise Miner. Si bien en la comunidad científica se conoce a SEMMA como una metodología, en el sitio de la empresa SAS se aclara que este no es el objetivo de la misma, sino más bien la propuesta de una organización lógica de las tareas más importantes del proceso de Minería de Datos. SEMMA establece un conjunto de cinco fases para llevar a cabo el proceso de minería: Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado), Assess (Evaluación).

Está especialmente enfocada al desarrollo del modelo de minería, y quedan fuera de su alcance otros aspectos del proyecto como el conocimiento del problema en estudio o la planificación de la implementación. SAS Enterprise Miner organiza sus herramientas (llamadas “nodos”) en base a las distintas fases que componen la metodología. Es decir, el software proporciona un conjunto de herramientas especiales para la etapa de muestreo, otras para la etapa de exploración, y así sucesivamente. Sin embargo, el usuario podría hacer uso del mismo siguiendo cualquier otra metodología de Minería de Datos (Moine, 2013, p. 14).

**Figura 8.***Metodología SEMMA*

*Nota.* Las cinco fases de la Metodología SEMMA (Moine, 2013, p. 14).

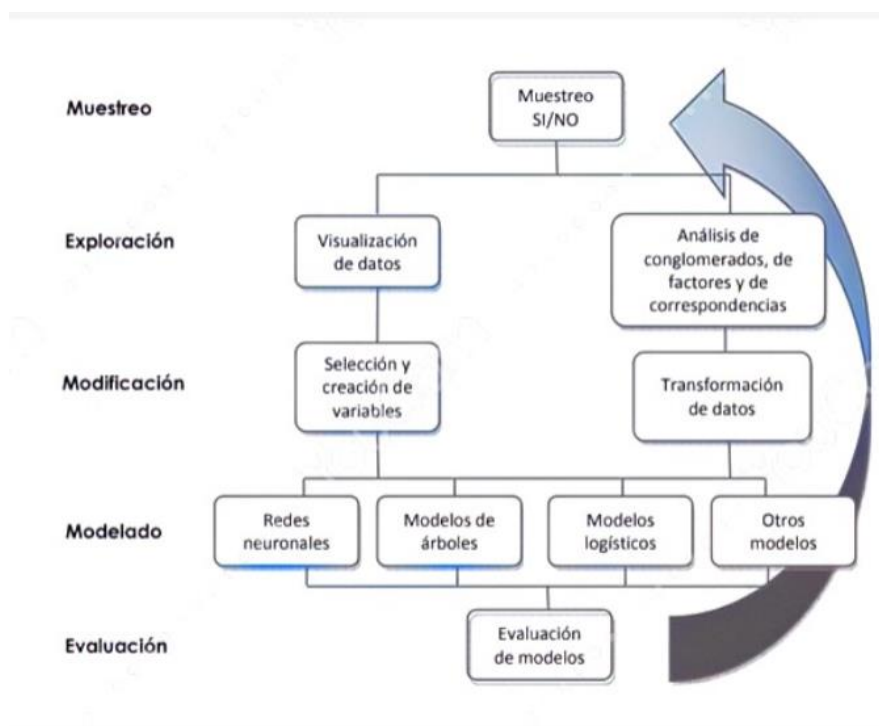
1. **Sample (Muestreo):** en esta etapa se toma una muestra del conjunto de datos disponible, que debe ser lo suficientemente grande para contener la información relevante, y lo suficientemente pequeña como para correr el proceso rápidamente. La etapa de muestreo es opcional, aconsejable cuando el tamaño del conjunto de datos es demasiado extenso.
2. **Explore (Exploración):** consiste en explorar los datos en búsqueda de relaciones y tendencias desconocidas. Es una etapa especial para familiarizarse con los datos, y formular nuevas hipótesis a partir de su análisis.
3. **Modify (Modificación):** consiste en una etapa de preparación de los datos, donde se limpian los valores anómalos, se realiza un tratamiento de los datos faltantes, y se seleccionan, crean y modifican las variables con las que se trabajarán.
4. **Model (Modelado):** consiste en la creación del modelo que permitirá predecir las variables de respuesta a partir de las variables explicativas, utilizando algunas de las técnicas predictivas como árboles de decisión, redes neuronales, análisis discriminante o análisis de regresión.

5. **Assess (Evaluación):** en esta fase se evalúa la utilidad y la exactitud de los modelos obtenidos en el proceso de Minería de Datos, por ejemplo, analizando la capacidad predictiva de los mismos.

SEMMA propone que luego de la fase de evaluación, se generan nuevas hipótesis que llevan a repetir el proceso iterativamente.

**Figura 9.**

*Fases en SEMMA*



*Nota.* Iteración de la metodología SEMMA (Moine, 2013, p. 15).

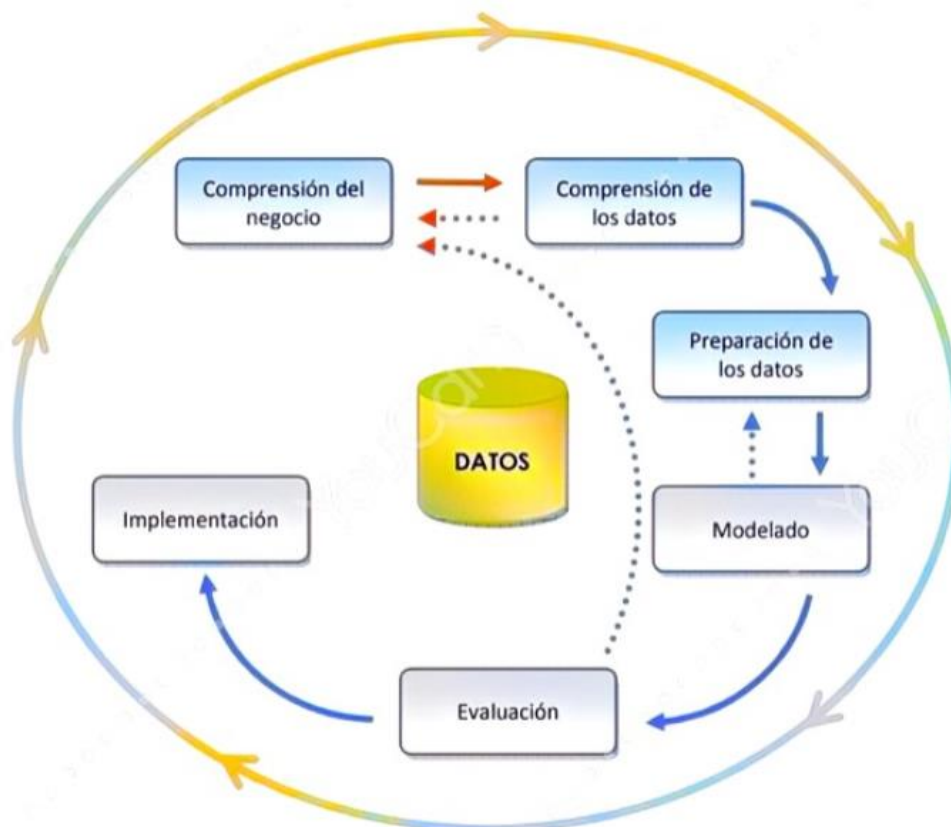
### 2.11.3 CRISP-DM

CRISP-DM (CRoss Industry Standard Process for Data Mining) fue presentada en el año 1999 por las empresas SPSS, Daimler Chrysler y NCR. Es una metodología abierta, no está ligada a ningún producto comercial, y fue construida en base a la experiencia de sus creadores, es decir, desde un enfoque práctico.

La metodología está estructurada en un proceso jerárquico, compuesto por tareas descritas en cuatro niveles diferentes de abstracción, que van desde lo general a lo específico.

CRISP-DM, propone en el nivel más alto seis fases para el proceso de minería de datos: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación e implementación. La sucesión de fases, no es necesariamente rígida.

Cada fase se descompone en un conjunto de tareas genéricas (o generales) de segundo nivel. Estas tareas son genéricas, ya que tratan de abarcar la mayoría de las situaciones posibles en Minería de Datos. A partir del tercer nivel de abstracción, se realiza un “mapeo” de las tareas genéricas definidas en el modelo a situaciones específicas. De esta forma, las tareas genéricas se traducen en tareas específicas para casos y proyectos concretos. En el cuarto nivel, encontramos las instancias de proceso, donde se describen las acciones, decisiones y resultados de un proyecto particular de Minería de Datos (Moine, 2013, p. 16).

**Figura 10.***Metodología CRISP-DM*

*Nota.* Se puede observar las fases de la metodología CRISP-DM (Moine, 2013, p. 16).

La metodología proporciona un modelo de referencia y una guía de usuario. El modelo de referencia presenta un resumen de las fases y tareas a llevar a cabo en cada una (junto con sus salidas). Es decir, describe “que” debería hacerse en un proyecto de Minería de Datos. La guía de usuario proporciona sugerencias para la ejecución de cada tarea del modelo de referencia.

Analizando el nivel más alto de abstracción del modelo, las seis fases que componen el proceso de Minería de Datos son:

1. **Comprensión del negocio:** en esta fase se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo.
2. **Comprensión de los datos:** fase que consiste en la recolección de datos que se utilizarán en el proyecto y la familiarización con los mismos. En esta etapa es posible el surgimiento de las primeras hipótesis acerca de la información que podría estar oculta.
3. **Preparación de los datos:** comprende aquellas actividades de tratamiento de los datos para construir la vista minable o conjunto de datos final sobre el cual se aplicarán las técnicas de minería.
4. **Modelado:** en esta etapa se aplican las diversas técnicas y algoritmos de minería sobre el conjunto de datos para obtener la información oculta y los patrones implícitos en ellos.
5. **Evaluación:** fase en la que se analizan los patrones obtenidos en función de los objetivos organizacionales. En esta etapa se debería determinar si se ha omitido algún objetivo importante del negocio y si el nuevo conocimiento será implementado, es decir, si se pasará a la próxima etapa.
6. **Implementación:** consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible para el usuario.

Cada una de estas fases generales se compone de un conjunto de tareas en las que se definen las salidas o entregables que se generan.

Si bien la metodología no especifica detalladamente cómo llevar a cabo cada tarea, los consejos de la sección “Guía de Usuario” resultan de mucha utilidad y orientación al momento de ejecutarlas (Moine, 2013, p. 17).

## 2.12 INGENIERÍA DE SOFTWARE

En la actualidad, la mayoría de profesionales y muchos usuarios tienen la fuerte sensación de que entienden el software. Pero, ¿es así? La descripción que daría un libro de texto sobre software sería más o menos así.

(Pressman, 2010, p. 4), No hay duda de que podrían darse definiciones más completas. Pero es probable que una definición más formal no mejore de manera apreciable nuestra comprensión. Para asimilar lo anterior, es importante examinar las características del software que lo hacen diferente de otros objetos que construyen los seres humanos. El software es elemento de un sistema lógico y no de uno físico. Por tanto, tiene características que difieren considerablemente de las del hardware:

1. El software se desarrolla o modifica con intelecto.
2. El software no se “desgasta”.
3. Aunque la industria se mueve hacia la construcción basada en componentes, la mayor parte del software se construye para un uso individualizado.

La ingeniería de software es una tecnología con varias capas. Cualquier enfoque de ingeniería (incluso la de software) debe basarse en un compromiso organizacional con la calidad. La administración total de la calidad, Six Sigma y otras filosofías similares que alimentan la cultura de mejora continua, y es esta cultura la que lleva en última instancia al desarrollo de enfoques cada vez más eficaces de la ingeniería de software. El fundamento en el que se apoya la ingeniería de software es el compromiso con la calidad. El fundamento para



la ingeniería de software es la capa proceso. El proceso de ingeniería de software es el aglutinante que une las capas de la tecnología y permite el desarrollo racional y oportuno del software de cómputo.

**Figura 11.**

*Capas de la Ingeniería de Software*



*Nota.* Capas de la Ingeniería de Software (Pressman, 2010, p. 12).

El proceso define una estructura que debe establecerse para la obtención eficaz de tecnología de ingeniería de software. El proceso de software forma la base para el control de la administración de proyectos de software, y establece el contexto en el que se aplican métodos técnicos, se generan productos del trabajo (modelos, documentos, datos, reportes, formatos, etc.), se establecen puntos de referencia, se asegura la calidad y se administra el cambio de manera apropiada.

Los métodos de la ingeniería de software proporcionan la experiencia técnica para elaborar software. Incluyen un conjunto amplio de tareas, como comunicación, análisis de los requerimientos, modelación del diseño, construcción del programa, pruebas y apoyo. Los métodos de la ingeniería de software se basan en un conjunto de principios fundamentales que gobiernan cada área de la tecnología e incluyen actividades de modelación y otras técnicas descriptivas.

Las herramientas de la ingeniería de software proporcionan un apoyo automatizado o semiautomatizado para el proceso y los métodos. Cuando se integran las herramientas de modo que la información creada por una pueda ser utilizada por otra, queda establecido un sistema llamado ingeniería de software asistido por computadora que apoya el desarrollo de software (Pressman, 2010, p. 12).

Es imposible operar el mundo moderno sin software. Las infraestructuras nacionales y los servicios públicos se controlan mediante sistemas basados en computadoras, y la mayoría de los productos eléctricos incluyen una computadora y un software de control. La fabricación y la distribución industrial están completamente computarizadas, como el sistema financiero. El entretenimiento, incluida la industria musical, los juegos por computadora, el cine y la televisión, usan software de manera intensiva. Por lo tanto, la ingeniería de software es esencial para el funcionamiento de las sociedades, tanto a nivel nacional como internacional (Sommerville, 2011, p. 4).

Los sistemas de software son abstractos e intangibles. No están restringidos por las propiedades de los materiales, regidos por leyes físicas ni por procesos de fabricación. Esto simplifica la ingeniería de software, pues no existen límites naturales a su potencial. Sin embargo, debido a la falta de restricciones físicas, los sistemas de software pueden volverse rápidamente muy complejos, difíciles de entender y costosos de cambiar.

Hay muchos tipos diferentes de sistemas de software, desde los simples sistemas embebidos, hasta los complejos sistemas de información mundial. No tiene sentido buscar notaciones, métodos o técnicas universales para la ingeniería de software, ya que diferentes tipos de software requieren distintos enfoques. Desarrollar un sistema organizacional de información es completamente diferente de un controlador para un instrumento científico. Ninguno de estos sistemas tiene mucho en común con un juego por computadora de gráficos

intensivos. Aunque todas estas aplicaciones necesitan ingeniería de software, no todas requieren las mismas técnicas de ingeniería de software.

Sin ingeniería de software, no se habría explorado el espacio, ni se tendría Internet o las telecomunicaciones modernas. Todas las formas de viaje serían más peligrosas y caras. La ingeniería de software ha contribuido en gran medida, y sus aportaciones en el siglo XXI serán aún mayores. (Sommerville, 2011, p. 4).

### **2.12.1 Historia de la Ingeniería de Software**

El concepto “ingeniería de software” se propuso originalmente en 1968, en una conferencia realizada para discutir lo que entonces se llamaba la “crisis del software” (Naur y Randell, 1969). Se volvió claro que los enfoques individuales al desarrollo de programas no escalaban hacia los grandes y complejos sistemas de software. Estos no eran confiables, costaban más de lo esperado y se distribuían con demora. A lo largo de las décadas de 1970 y 1980 se desarrolló una variedad de nuevas técnicas y métodos de ingeniería de software, tales como la programación estructurada, el encubrimiento de información y el desarrollo orientado a objetos. Se perfeccionaron herramientas y notaciones estándar y ahora se usan de manera extensa (Sommerville, 2011, p. 5).

### **2.12.2 Procesos de Software**

Un proceso de software es una serie de actividades relacionadas que conduce a la elaboración de un producto de software. Estas actividades pueden incluir el desarrollo de software desde cero en un lenguaje de programación estándar como Java o C. Sin embargo, las aplicaciones de negocios no se desarrollan precisamente de esta forma. El nuevo software empresarial con frecuencia ahora se desarrolla extendiendo y modificando los sistemas existentes, o configurando e integrando el software comercial o componentes del sistema.

Existen muchos diferentes procesos de software, pero todos deben incluir cuatro actividades que son fundamentales para la ingeniería de software:

1. **Especificación del software:** tienen que definirse tanto la funcionalidad del software como las restricciones de su operación.
2. **Diseño e implementación del software:** debe desarrollarse el software para cumplir con las especificaciones.
3. **Validación del software:** hay que validar el software para asegurarse de que cumple lo que el cliente quiere.
4. **Evolución del software:** el software tiene que evolucionar para satisfacer las necesidades cambiantes del cliente.

Los procesos de software son complejos y, como todos los procesos intelectuales y creativos, se apoyan en personas con capacidad de juzgar y tomar decisiones. No hay un proceso ideal; además, la mayoría de las organizaciones han diseñado sus propios procesos de desarrollo de software. Los procesos han evolucionado para beneficiarse de las capacidades de la gente en una organización y de las características específicas de los sistemas que se están desarrollando. Para algunos sistemas, como los sistemas críticos, se requiere de un proceso de desarrollo muy estructurado. Para los sistemas empresariales, con requerimientos rápidamente cambiantes, es probable que sea más efectivo un proceso menos formal y flexible (Sommerville, 2011, p. 28).

Los procesos de software pueden mejorarse con la estandarización de los procesos, donde se reduce la diversidad en los procesos de software en una organización. Esto conduce a mejorar la comunicación, a reducir el tiempo de capacitación, y a que el soporte de los procesos automatizados sea más económico (Sommerville, 2011, p. 29).

### 2.13 METODOLOGÍA SCRUM

Scrum es un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto. Estas prácticas se apoyan unas a otras y su selección tiene origen en un estudio de la manera de trabajar de equipos altamente productivos, Scrum se basa en los siguientes puntos nombre (Maida & Pacienza, 2015, p. 87).

- El desarrollo incremental de los requisitos del proyecto en bloques temporales cortos y fijos (iteraciones de un mes natural y hasta de dos semanas, si así se necesita). Las iteraciones se pueden entender como mini proyectos, en todas las iteraciones se repite un proceso de trabajo similar (de ahí el nombre “iterativo”) para proporcionar un resultado completo sobre el producto final, de manera que el cliente pueda obtener los beneficios del proyecto de forma incremental. Para ello, cada requisito se debe completar en una única iteración. El equipo debe realizar todas las tareas necesarias para completarlo (incluyendo pruebas y documentación) y que esté preparado para ser entregado al cliente con el mínimo esfuerzo necesario. De esta manera no se deja para el final del proyecto ninguna actividad arriesgada relacionada con la entrega de requisitos.
- La priorización de los requisitos por valor para el cliente y coste de desarrollo en cada iteración. Para que un proyecto proporcione el mejor resultado posible, y como soporte fundamental al control empírico del proyecto, es necesario priorizar los requisitos de manera regular, en cada iteración, según el valor que proporcionan al cliente en ese momento y su coste estimado de desarrollo. Como resultado de esta priorización se actualiza la lista de requisitos priorizada (Product Backlog).

- El control empírico del proyecto. Por un lado, al final de cada iteración se demuestra al cliente el resultado real obtenido, de manera que pueda tomar las decisiones necesarias en función de lo que observa y del contexto del proyecto en ese momento. Por otro lado, el equipo se sincroniza diariamente y realiza las adaptaciones necesarias.
- La potenciación del equipo, que se compromete a entregar unos requisitos y para ello se le otorga la autoridad necesaria para organizar su trabajo.
- La sistematización de la colaboración y la comunicación tanto entre el equipo y como con el cliente.
- El timeboxing de las actividades del proyecto, para ayudar a la toma de decisiones y conseguir resultados. La técnica del timebox consiste en fijar el tiempo máximo para conseguir ciertos objetivos, tomar una decisión o realizar unas tareas, y hacer lo mejor que podamos en ese.
- intervalo. De esta manera, en lugar de ponerse a trabajar en algo hasta que esté hecho, de antemano se acuerda que solo se dedica un tiempo limitado. La consciencia de esta limitación temporal favorece la priorización de objetivos/tareas y fuerza la toma de decisiones.

Herramientas que se utilizan en Scrum son:

- **Lista de objetivos / requisitos priorizada (Product Backlog):** la lista de objetivos/requisitos priorizada representa la visión y expectativas del cliente respecto a los objetivos y entregas del producto o proyecto. El cliente es el responsable de crear y gestionar la lista con la ayuda del Facilitador y del equipo, quien proporciona el coste estimado de completar cada requisito.

- **Lista de tareas de la iteración (Sprint Backlog):** es una lista de tareas que el equipo elabora en la reunión de planificación de la iteración (Sprint planning) como plan para completar los objetivos/requisitos seleccionados para la iteración y que se compromete a demostrar al cliente al finalizar la iteración, en forma de incremento de producto preparado para ser entregado. A cada tarea se la suele llamar Historia (History).
- **Gráficos de trabajo pendiente (Burndown charts):** consiste en un gráfico que indica el trabajo pendiente a lo largo del tiempo y muestra la velocidad a la que se están completando los objetivos/requisitos. Permite predecir si el equipo podrá completar el trabajo en el tiempo estimado

Las ventajas de utilizar Scrum son:

- Entregas parciales a corto plazo de resultados.
- Gestión regular de las expectativas del cliente y basada en resultados tangibles.
- Resultados anticipados.
- Flexibilidad y adaptación respecto a las necesidades del cliente, cambios en el mercado.
- Gestión sistemática del Retorno de Inversión (ROI).
- Mitigación sistemática de los riesgos del proyecto.
- Productividad y calidad.
- Alineamiento entre el cliente y el equipo de desarrollo.
- Equipo motivado.

## 2.14 ÍNDICES DELICTIVOS

La violencia urbana se expande, cada vez con mayor fuerza, en las ciudades de la Región. El incremento real de los actos delictivos y la nueva percepción de la población, provocan cambios manifiestos en las urbes latinoamericanas. Allí están las transformaciones en el urbanismo (amurallamiento de la ciudad, nuevas formas de segregación residencial), en los comportamientos de la población (angustia, desamparo), en la interacción social (reducción de la ciudadanía, nuevas formas de socialización) y en la militarización de las ciudades; amén de la reducción de la calidad de vida de la población.

Hoy se ve la necesidad de tener un mayor conocimiento de la problemática y una nueva óptica para enfrentar el problema. Ya no es suficiente actuar con el sentido común y con la transferencia de recursos hacia la represión; se requiere tener verdaderos observatorios de la violencia, nuevos conceptos y metodologías para entenderla y novedosas concepciones para enfrentarla. Esto será posible solo si se incorporan a nuevos actores sociales y no se convierte en un tema exclusivo de un sector del Estado. En definitiva, se trata de un problema crucial de la sociedad latinoamericana actual, porque en su enfrentamiento priman enfoques que tienden a incrementarla en vez de mitigarla. De allí que nos encontremos en los albores de un cambio en el escenario de la violencia (o de la guerra): del campo y las naciones a la ciudad. Sin embargo, desgraciadamente, es aún un tema poco conocido y poco explorado (Carrión, 2002, p. 14).

En América Latina se encuentra enfrentada a diversas fuentes de inseguridad. Desde la perspectiva económica, las fuentes son multifacéticas y entre las principales se han destacado las provenientes de los procesos de globalización que han sido acompañados por la declinación de la protección del empleo, por resultados imprevistos de las crisis foráneas y por la volatilidad de los flujos de capital, e inestabilidades macroeconómicas, además de la debilidad de las instituciones para enfrentar esos riesgos. Sin duda, entre las causas más



directas de esa inseguridad se encuentra el aumento de la violencia de diversos tipos en la Región y su expresión en la inseguridad creciente de la vida cotidiana de los ciudadanos. La seguridad ciudadana constituye un ámbito en que la ciudadanía y la cohesión social se vinculan estrechamente. En la medida en que los grupos interiorizan patrones de comportamiento que no obedecen a las normas legales y sociales, la sociedad se expone cada vez más al uso de la violencia como una forma de resolver conflictos y obtener recursos. Las dificultades en la aplicación de justicia que las instituciones experimentan para garantizar protección a sus ciudadanos, acentúan la percepción de inseguridad.

Las definiciones de seguridad ciudadana han enfatizado distintas dimensiones y niveles, en especial su intangibilidad y su carácter subjetivo. La seguridad ciudadana se define, de una manera amplia, como la preocupación por la calidad de vida y la dignidad humana en términos de libertad, acceso al mercado y oportunidades sociales. La pobreza y la falta de oportunidades, el desempleo, el hambre, el deterioro ambiental, la represión política, la violencia, la criminalidad y la drogadicción pueden constituir amenazas a la seguridad ciudadana (Carrión, 2002, p. 109).

Los factores de riesgo se pueden clasificar en tres grandes conjuntos:

- a) **factores relacionados con la posición y situación familiar y social de las personas:** sexo, edad, educación, socialización en la violencia, consumo de alcohol y drogas.
- b) **factores sociales, económicos y culturales:** desempleo, pobreza, hacinamiento, desigualdad social, violencia en los medios de comunicación, cultura de la violencia.
- c) **factores contextuales e institucionales:** guerra, tráfico de drogas, corrupción, disponibilidad de armas de fuego, festividades, entre otros.

**Tabla 6.***Presencia de Factores de Riesgo*

| <b>Factores de riesgo</b>                                                                                    | <b>Países con alta presencia</b>                                                                        | <b>Países con presencia media</b>                                                                                    | <b>Países con baja presencia</b>                                    |
|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| 1. Desigualdad del ingreso urbano.<br>Consiente entre 10% más rico y 40% más pobre                           | Sobre 11 veces Brasil, Colombia, Chile, Guatemala y Panamá                                              | Entre 8 y 11 veces Argentina, Bolivia, Honduras, Nicaragua, Paraguay, República Dominicana y Venezuela.              | Hasta 8 veces Costa Rica, Ecuador, El Salvador, México y Uruguay.   |
| 2. Pobreza de los hogares urbanos.                                                                           | 40% y más de los hogares Bolivia, Ecuador, Guatemala, Haití, Honduras, Nicaragua, Paraguay y Venezuela. | 20% y 39% de los hogares Brasil, Colombia, El Salvador, México, Panamá, Perú y República Dominicana.                 | Menos de 20% de los hogares Argentina, Chile, Costa Rica y Uruguay. |
| 3. Tasas de desempleo abierto urbano                                                                         | Más de 10% Argentina, Colombia, Panamá, Uruguay y Venezuela.                                            | Entre 6% y 10% Brasil, Chile, Costa Rica, Ecuador, Honduras, México, Paraguay y Perú.                                | Menos de 6% Bolivia, Honduras y México                              |
| 4. Porcentaje de jóvenes urbanos de 13 a 17 años que no estudian ni trabajan.                                | Sobre 15% Honduras y Uruguay.                                                                           | Entre 8% y 15% Argentina, Brasil, Colombia, Costa Rica, El Salvador, México, Nicaragua, Panamá, Paraguay y Venezuela | Menos de 8% Bolivia, Chile, Ecuador y República Dominicana          |
| 5. Déficit educacional, porcentaje de niños urbanos de 14 o 15 años que no han completado 6 años de estudio. | Sobre 20% Brasil y Honduras.                                                                            | Entre 10% y 20% Colombia, Costa Rica, Ecuador, México, Paraguay y Venezuela.                                         | Menos de 10% Argentina, Chile, Panamá y Uruguay.                    |
| 6. Porcentaje de jóvenes urbanos de 13 a 17 años que trabajan.                                               | Sobre 15% Bolivia, Brasil, Ecuador, Honduras, México y Paraguay                                         | Entre 5% y 15% Argentina, Colombia, Costa Rica, Venezuela y Uruguay.                                                 | Hasta 5% Chile y Panamá.                                            |

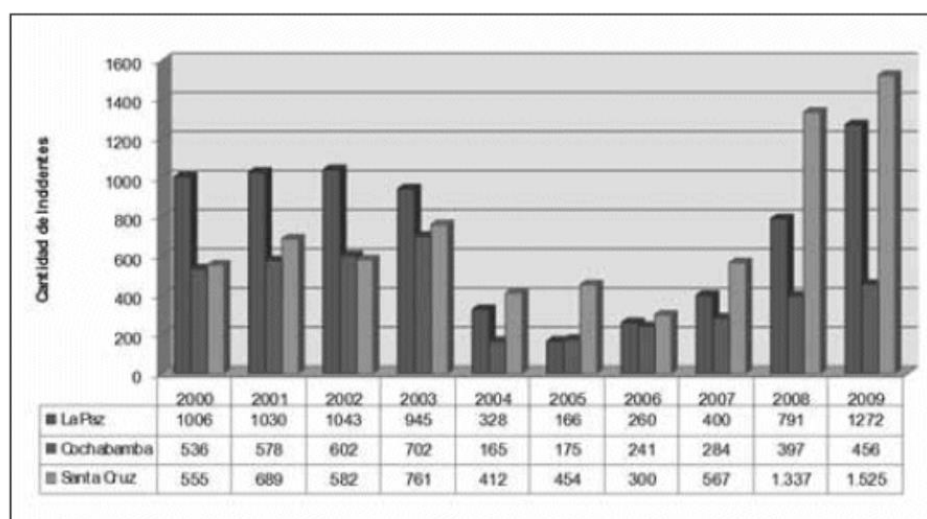
*Nota.* Factores de riesgo de violencia urbana (Carrión, 2002, p. 113).

En la última década, la violencia se ha convertido en uno de los temas más importantes de las ciudades que conforman el territorio boliviano, esto a causa de un notorio incremento en cantidad y modalidad de los actos delictivos secuestros, atracos organizados y a los impactos sociales y económicos que estos conllevan.

La violencia se ha extendido principalmente en los grandes centros urbanos del país, con peculiaridades propias en cada región, y provocando varios cambios en la modalidad del urbanismo y el comportamiento de la población. (Rubin de Celis y otros, 2012).

**Figura 12.**

*Denuncias de Atracos*



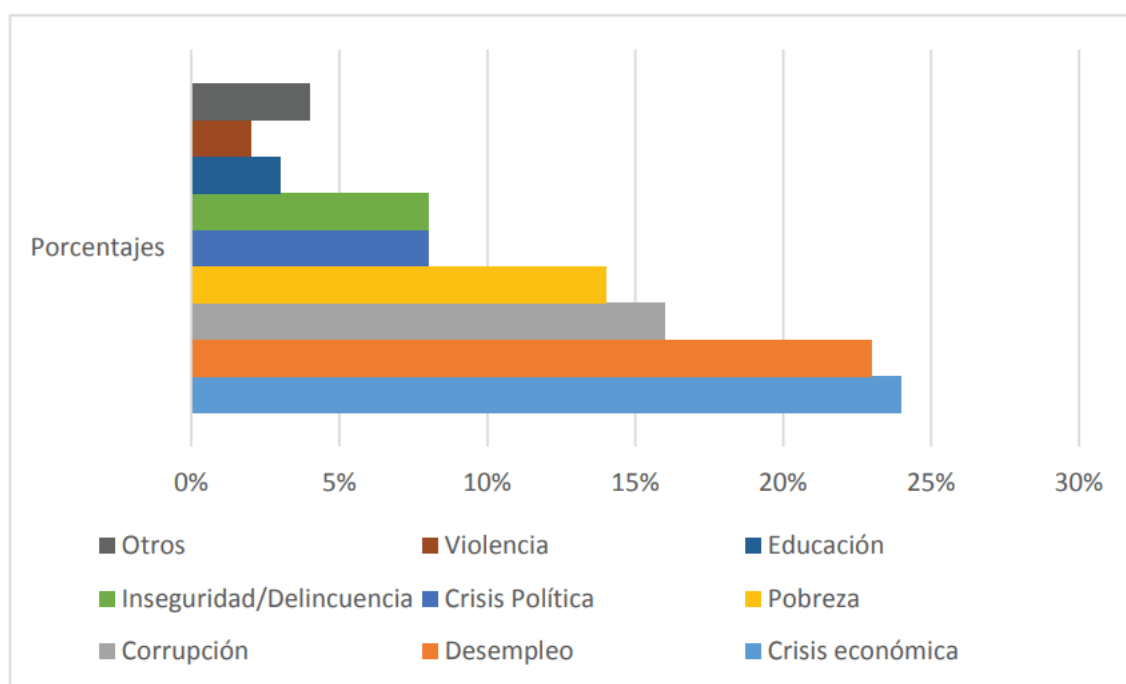
*Nota.* Denuncias de atracos en diferentes años (Rubin de Celis y otros, 2012).

La violencia en el mundo afecta mucho más las áreas urbanas que las rurales, y dentro de las primeras, mucho más las ciudades grandes que a las pequeñas. Esta conexión, aunque raras veces cuantificada, es parte de nuestro subconsciente: las actividades delictivas ya no ocurren en desolados pasajes en el campo sino en el centro de la ciudad e indiferentes peatones.

Particularmente en el Estado Plurinacional de Bolivia, mediante datos del Instituto Nacional de Estadística, INE y Bolivia: Opinión Pública y Policía 1990 – 2007 (series históricas Comparadas), muestra, la percepción por parte de la sociedad sobre los principales problemas que aquejan al país (Tudela, 2012, p. 34).

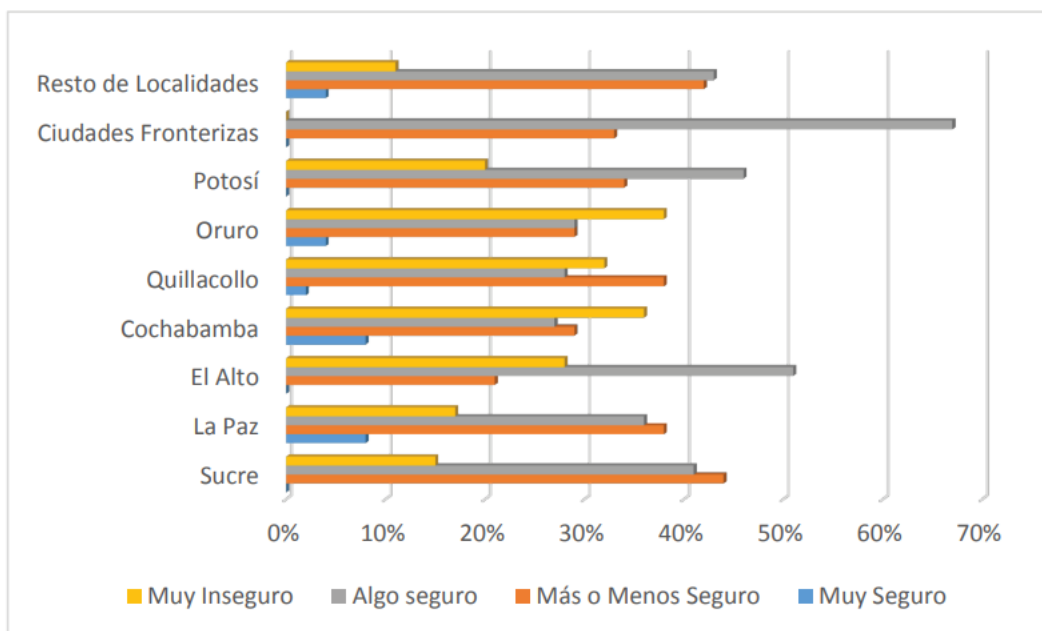
**Figura 13.**

*Problemas que posee Bolivia*



*Nota.* Percepción de la sociedad para los hechos delictivos (Tudela, 2012, p. 34).

Según (Tudela, 2012, p. 36), indica una muestra que las personas sienten o perciben una mayor inseguridad en las ciudades fronterizas, más que en las ciudades principales, esto debido a que al estar alejados tienen un número significativamente menor de policías.

**Figura 14.***Inseguridad en horarios nocturnos.*

*Nota.* Apreciación de seguridad de caminar solo por la noche por ciudades (Tudela, 2012, p. 36),

## 2.14.1 Causas

### 2.14.1.1 La Familia

El ambiente familiar desfavorable (familias disfuncionales, divorcios de los padres, desacuerdos conyugales, libertad de costumbres, ausencia del padre, deficiencia educativa, poca vigilancia a los hijos) es un factor en la criminogénesis. Las estadísticas de delincuentes que vienen de un hogar crítico son bastante numerosas tanto en el Perú y el mundo entero (Carranza, 2013, p. 16).

### 2.14.1.2 Las Condiciones Socioeconómicas

En un país como el nuestro, con problemas sociales y económicos, existe un clima permanente de intranquilidad debido al desempleo o subempleo, así como a la falta de

inclusión en el desarrollo económico, lo que nos ha llevado un alto riesgo de predisposición criminógena. La acción de estas circunstancias es compleja: la familia disfuncional, estado constante de tensión, dificultades alimenticias, inseguridad económica, costumbres licenciosas, hacinamiento de la población, promiscuidad, desesperanza, inseguridad. Todo esto ha repercutido fatalmente en el alto índice de criminalidad y la aparición de nuevas y graves modalidades delictivas. Finalmente, debemos admitir que dados los problemas socioeconómicos de nuestro país, la alternativa es una: optar por un nuevo modelo de sociedad y de hombre realmente superior, o convertirnos en retrógrados (Carranza, 2013, p. 16).

#### **2.14.1.3 Influencias Psicosociales**

Dentro de estos consideramos a la influencia psicológica del medio social en que se desarrolla el individuo, vale decir, la influencia de los grupos de edad. El barrio, el trabajo, los medios de comunicación social y otros agentes socializadores que pueden ser inadecuados o mal percibidos por el individuo. El estudio de los jóvenes delincuentes permite evidenciar la peligrosa influencia de ciertas relaciones sociales. Aunque no todos los jóvenes que viven en los barrios con alta incidencia delictiva, son delincuentes; es de señalar que los jóvenes, sin vigilancia de los padres o superficial, llegan a adquirir un esquema de comportamiento proclive al delito. Los medios de comunicación social ejercen alta influencia en los jóvenes, sobre todo en la época de la preadolescencia (pubertad). El delito y el crimen aparecen en el cine, la TV e incluso en la literatura (novelas) como un subyugante, atractivo, apasionante y audaz estilo de vida, donde la aventura y placeres a montón bien valen la pena vivirlos así se tenga que entregar la vida o la libertad a cambio. Esta es la concepción de muchos delincuentes “profesionales”, la cual es muy difícil de cambiar, por ello es mejor optar por la prevención (Carranza, 2013, p. 17).

#### **2.14.1.4 Influencia Cultural**

Estudiosos como Sellin, Kaiser, Shoham, Grumfeld (1997), y otros plantean la tesis del conflicto cultural, en la cual señalan que en una sociedad se producen, en algunas etapas críticas, conflictos diversos entre grupos de personas y este choque cultural puede ser un factor determinante en el incremento de la criminalidad. Los conflictos entre los grupos de una sociedad se pueden deber a sistemas negativos de prejuicios étnicos, económicos, sociales, etc. En el proceso migratorio, en el cual las personas llevan consigo una forma de vida (corrupción-delincuencia) diferente a la de los otros grupos, se contamina con estas conductas a los sujetos predispuestos. Shoham (1962), dice que el índice de criminalidad será tanto más elevado cuanto más divergentes sean las normas jurídicas de los valores y costumbres de un determinado grupo social. Del conflicto cultural, los grupos sociales crean su propia subcultura por oposición al grupo mayoritario. Albert Cohén (1955), es el primero en desarrollar la teoría sobre la subcultura delincuente; se cree que el fundamento de esta teoría parte de los postulados de Sellén. Cloward y Ohlin (1969), amplían los estudios.

#### **2.14.1.5 Corrupción y el Delito en las Instituciones**

Este es un grave problema que debe ser erradicado por la connotación e implicancias sociales que trae consigo. Existen en los últimos tiempos loables esfuerzos para afrontar radicalmente el problema. Los especialistas en análisis experimental de la conducta explican la psicopatización colectiva como producto de una serie de estímulos reforzados por el medio social, como “el dinero da poder y para conseguirlo no interesan los medios” o viceversa “el poder debe brindarme bienestar económico a como dé lugar”. Los sistemas de valores, el primero es el económico y el último es el ético, actitudes y patrones de comportamiento que se transmiten en ciertos medios, no son los mejores. Por ejemplo, pensar que es picardía, “criollada” o inteligencia, el logro de ciertas metas aprovechándose de los demás o del cargo que se ostenta, sin importar el daño que se ocasiona. Se dirá que siempre va a existir

delincuencia y criminalidad en todas partes del mundo, pero acaso esta no se puede reducir o, por lo menos, detener o amenguar su avance escandaloso, el delito en las instituciones sucede por las razones mencionadas y muchas otras más, y todos tenemos responsabilidad en el problema. El reforzamiento de las conductas delictivas en las instituciones, lo efectúa muchas veces la propia comunidad, pero también es verdad que en muchas de ellas existen verdaderos delincuentes, simulando una imagen de personas inofensivas, pero que pueden causar mucho sufrimiento y dolor a gente inocente, pues en todas partes existen individuos con alguna psicopatología que no es detectada por sus jefes o por los especialistas. Existen muchas personas que padecen en alguna etapa de su vida de alteraciones de tipo neurótico y más aún rasgos psicopáticos. Los rasgos psicopáticos se encuentran en todos, aunque refrenados, o sublimados, pero latentes y los factores determinantes y desencadenantes pueden hacerlos aparecer. Un alto porcentaje de delincuentes pertenece a la clase baja y a la clase media-baja. Debido a la nueva legislación de lucha contra la corrupción, hoy existen sujetos de clase socioeconómica alta en los penales, pero aún no abundan los de esa clase en las cárceles, pues son los profesionales del delito de cuello y corbata, o muy bien pasan desapercibidos en un medio aún corrupto donde el poder del dinero los hace intocables (Carranza, 2013, p. 18).

## **2.14.2 Efectos**

### **2.14.2.1 Daños Físicos y Sicológicos a las Víctimas**

**Lesiones físicas:** los delitos pueden causar lesiones físicas de diversa gravedad, desde leves hasta graves e incluso la muerte.

**Daños psicológicos:** los delitos pueden causar traumas psicológicos, como ansiedad, depresión, estrés postraumático, etc.



**Sentimiento de inseguridad:** las víctimas de delitos pueden sentir miedo, inseguridad y desconfianza hacia los demás.

#### **2.14.2.2 Pérdidas Económicas**

**Robo de bienes:** los delitos pueden ocasionar la pérdida de bienes materiales, como dinero, joyas, vehículos.

**Costos de reparación:** los daños causados por los delitos pueden ocasionar gastos en reparaciones, reposición de bienes.

**Costos médicos:** las lesiones físicas causadas por los delitos pueden implicar gastos en atención médica, hospitalización.

#### **2.14.2.3 Deterioro del Tejido Social**

**Desconfianza e inseguridad:** los delitos pueden generar un clima de desconfianza e inseguridad en la sociedad.

**Desintegración familiar:** los delitos pueden afectar negativamente a las relaciones familiares y generar desintegración.

**Deterioro de la calidad de vida:** los delitos pueden deteriorar la calidad de vida en las comunidades, ya que las personas pueden tener miedo de salir a la calle o participar en actividades sociales.

#### **2.14.2.4 Impacto en el Sistema Judicial**

**Aumento de la carga de trabajo:** los delitos generan un aumento de la carga de trabajo para el sistema judicial, que debe investigar, procesar y juzgar a los responsables.

**Costos del sistema judicial:** los delitos implican un costo para el Estado en términos de recursos humanos, materiales y financieros.

**Congestión del sistema judicial:** el aumento de la delincuencia puede congestionar el sistema judicial, lo que puede afectar a la eficiencia del mismo.

#### **2.14.2.5 Estigmatización de las Víctimas y los grupos Sociales**

**Victimización secundaria:** las víctimas de delitos pueden ser estigmatizadas por la sociedad, lo que puede aumentar su sufrimiento.

**Estigmatización de grupos sociales:** los delitos pueden contribuir a la estigmatización de ciertos grupos sociales, como los jóvenes, los inmigrantes o las personas con problemas de salud mental.

### **2.15 ÍNDICES DELICTIVOS EN EL MUNICIPIO DE EL ALTO**

El Alto es una urbe "sui generis" debido a ciertas características históricas, demográficas, culturales, económicas, sociales y políticas. En efecto, es la ciudad más joven, la más pobre, la más abigarrada, la más postergada, la más peligrosa y la más violenta. En el terreno de las estadísticas, la ciudad de El Alto es considerada como una de las urbes más violentas y peligrosas del país. A nivel nacional, en términos de violencia social urbana ocupa el segundo lugar después de la ciudad de Santa Cruz y, en relación con la violencia intrafamiliar, ocupa el primer puesto. Para el 2006, el Informe sobre Desarrollo Humano en Bolivia vinculado al tema de la "Policía Nacional y Seguridad Ciudadana" señala que ocho de cada diez habitantes alteños se sienten inseguros o muy inseguros al recorrer los espacios públicos de su ciudad. En este mismo año, un diagnóstico realizado por la Alcaldía alteña establece que cuatro de cada diez alteños o, en su defecto, algunos de sus familiares fueron víctimas de algún tipo de delito.

Para el año 2011, la FELCC informa que en esta ciudad se producen, en promedio por día, entre 3 y 5 asaltos, 2 y 3 atracos, 2 y 3 casos de víctimas "acogotadas"; asimismo, cada día en promedio, se reporta entre una y cuatro personas muertas por causas violentas. Estos

datos estadísticos revelan un panorama preocupante en relación con la inseguridad ciudadana en dos dimensiones, por un lado, mayores niveles de inseguridad fáctica que se traducen en un incremento de la violencia, los homicidios y la delincuencia y, por otro lado, mayores niveles de inseguridad simbólica que van acompañadas de un aumento en las sensaciones de miedo, temor e inseguridad. En función de esta bidimensionalidad, no resulta casual que casi la mayoría de los "alteños", tengan una historia que contar vinculada a la violencia e inseguridad fáctica y simbólica en términos de miedo, temor, recelo, amenazas, golpizas, peleas, robos, atracos, asaltos, heridas, desapariciones e inclusive muertes.

Esta apretada revisión de los datos estadísticos provenientes de organismos policiales e instituciones municipales, pone en evidencia el hecho de que una buena parte de la población alteña no se siente segura y que la inseguridad ciudadana se ha constituido en un mal social muy difundido, latente, sordo y frecuente que se está convirtiendo en pan amargo de todos los días (Valdivia, 2011).

### **2.15.1 Posibles factores delictivos**

**Asaltos y atracos:** se define legalmente al robo como la apropiación de bienes y objetos a través del uso de la violencia, la intimidación de personas o la imposición de la fuerza sobre las cosas. En relación con este tipo de delito, se debe señalar que existe una bidimensionalidad y una diferencia conceptual entre los términos "asalto" y "atraco". Por una parte, los asaltos hacen referencia a los robos perpetrados en los domicilios particulares y, por otra, los atracos incluyen a los robos que se efectúan con violencia fáctica y simbólica en diferentes espacios públicos y/o en los vehículos. Los alteños encuentran en los asaltos y los atracos, amenazas peligrosas, severas y persistentes para su seguridad y tranquilidad, a diario se reportan robos a domicilios, departamentos, viviendas, negocios, tiendas, comercios, oficinas públicas y a personas particulares. Según funcionarios policiales de la FELCC-E1 Alto, en la gestión 2009 se han registrado 1.117 delitos contra la propiedad privada y 461 contra las

personas, para la gestión 2001 estas cifras se han incrementado en 1.416 y 714 casos respectivamente, tratándose en la mayoría de los casos de robo patrimonial y de robo a particulares (Valdivia, 2011).

**Pandillas juveniles:** es evidente que no todos los jóvenes alteños son "disociales" o "potenciales delincuentes", pero también es innegable la presencia de cientos de ellos involucrados en la drogadicción, el alcohol, la prostitución y la delincuencia, a través de la conformación y el accionar de las denominadas pandillas juveniles. En la mayoría de estos jóvenes pandilleros se advierte un deterioro en la calidad de las relaciones familiares y una profunda pérdida de valores morales, que provoca una deshumanización y una degradación social que afecta negativamente su convivencia individual, familiar y social. Su frustración es tal que se han visto obligados a conformar o refugiarse en "tribus o bandas urbanas", a través de las cuales desarrollan sobre todo actividades disociales o conductas delictivas (Valdivia, 2011).

**Bares y cantinas:** es ya un lugar común decir que en El Alto el consumo "continuo" y "exagerado" del "alcohol", que se ha convertido en un elemento indispensable para celebrar acontecimientos sociales tanto privados como públicos, sean estos bautizos, matrimonios, entierros, fines de luto, festividades religiosas, eventos deportivos, celebraciones sociales e incluso actividades educativas. Solo con fines ilustrativos, cabe señalar que no existe un fin de semana donde algún vecino o familiar no se halle compartiendo bebidas alcohólicas y donde alguna *plaza*, parque, calle o avenida se convierta en un espacio público de consumo indiscriminado de bebidas alcohólicas debido a la celebración de un "preste", una "festividad religiosa" o un "aniversario de la zona" o de algún sindicato o gremio (Valdivia, 2011).

**Desapariciones, violaciones y asesinatos:** otro de los factores centrales que determina un clima de inseguridad ciudadana está relacionado con la desaparición de

personas. Según funcionarios policiales, entre los meses de enero y marzo de 2010, en las ciudades de La Paz y El Alto se registraron alrededor de 300 personas desaparecidas; se trata en particular de mujeres, adolescentes, niños y niñas que proceden de sectores sociales económicamente bastante deprimidos. Las causas de estas desapariciones son diversas, pero vinculadas principalmente a problemas familiares, disputas conyugales y situaciones violentas en el caso de las violaciones casi la mayoría de las mujeres alteñas, independientemente de su edad y su condición social, se siente amenazada por este tipo de agresión sexual (Valdivia, 2011).

**Violencia intrafamiliar:** si bien la mayoría de los alteños no se siente seguro en los distintos espacios públicos, las manifestaciones de violencia e inseguridad se hallan también presentes al interior de los hogares a través de la denominada violencia intrafamiliar. Según estadísticas nacionales, El Alto presenta los mayores índices de violencia intrafamiliar, recientes investigaciones han constatado que en 8 de cada 10 hogares alteños prevalece esta forma de maltrato en términos de agresiones físicas, psicológicas y sexuales, en particular en contra de niños, niñas, adolescentes y esposas o concubinas. Asimismo, se sabe que en Defensorías de la Niñez y Adolescencia se atienden diariamente entre 15 y 25 casos de maltrato infantil y en Servicios Legales Integrales Municipales entre 20 y 30 casos diarios de violencia en contra de las mujeres (Valdivia, 2011).

## 2.15.2 Datos sobre los Índices Delictivos

**Figura 15.**

*Denuncias de delitos*

**BOLIVIA: CANTIDAD DE DENUNCIAS DE DELITOS DE ALTA CONNOTACIÓN SOCIAL<sup>(1)</sup>,  
POR AÑO, SEGÚN MUNICIPIOS CAPITALES DE DEPARTAMENTO Y EL ALTO, 2017 – 2022**  
(En número)

| Municipios capitales de departamento y El Alto | 2017          | 2018          | 2019          | 2020          | 2021          | 2022          |
|------------------------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| <b>TOTAL</b>                                   | <b>50.640</b> | <b>48.612</b> | <b>51.774</b> | <b>39.784</b> | <b>45.741</b> | <b>51.599</b> |
| Sucre                                          | 3.919         | 3.911         | 3.233         | 2.267         | 3.579         | 4.604         |
| La Paz                                         | 6.436         | 5.745         | 6.832         | 5.611         | 6.167         | 7.224         |
| <u>El Alto</u>                                 | 5.799         | 4.848         | 4.761         | 3.890         | 4.260         | 6.039         |
| Cochabamba                                     | 4.827         | 4.157         | 4.796         | 4.263         | 5.256         | 6.048         |
| Oruro                                          | 1.765         | 1.340         | 1.366         | 763           | 1.034         | 1.804         |
| Potosí                                         | 2.979         | 3.051         | 3.165         | 2.194         | 2.537         | 2.890         |
| Tarija                                         | 2.917         | 3.355         | 3.609         | 2.831         | 3.158         | 3.276         |
| Santa Cruz de la Sierra                        | 17.774        | 18.253        | 19.795        | 15.064        | 16.793        | 16.471        |
| Trinidad                                       | 2.426         | 2.454         | 2.511         | 1.813         | 1.796         | 1.786         |
| Cobija                                         | 1.798         | 1.498         | 1.706         | 1.088         | 1.161         | 1.457         |

*Nota.* Denuncias de delitos (Choque Soto y otros, 2023, p. 33).

**Figura 16.***Denuncias de Delitos*

**BOLIVIA: TASA DE DENUNCIAS DE DELITOS DE ALTA CONNOTACIÓN SOCIAL<sup>(1)</sup>,  
POR AÑO, SEGÚN MUNICIPIOS CAPITALES DE DEPARTAMENTO Y EL ALTO, 2017 – 2022  
(Por 100 mil habitantes)**

| Municipios capitales de departamento y El Alto | 2017          | 2018          | 2019          | 2020          | 2021          | 2022          |
|------------------------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| <b>TOTAL</b>                                   | <b>875,37</b> | <b>824,56</b> | <b>862,05</b> | <b>650,47</b> | <b>734,64</b> | <b>814,36</b> |
| Sucre                                          | 1.191,01      | 1.167,13      | 947,38        | 652,25        | 1.011,03      | 1.276,96      |
| La Paz                                         | 713,64        | 629,27        | 739,40        | 600,12        | 651,98        | 755,07        |
| El Alto                                        | 573,71        | 470,61        | 453,57        | 363,78        | 391,14        | 544,52        |
| Cochabamba                                     | 617,75        | 522,00        | 591,12        | 515,90        | 624,77        | 706,38        |
| Oruro                                          | 549,41        | 409,18        | 409,41        | 224,57        | 299,02        | 512,79        |
| Potosí                                         | 1.208,45      | 1.216,54      | 1.240,20      | 844,70        | 959,52        | 1.073,63      |
| Tarija                                         | 1.185,65      | 1.334,66      | 1.405,80      | 1.080,28      | 1.181,10      | 1.201,36      |
| Santa Cruz de la Sierra                        | 1.033,27      | 1.038,48      | 1.102,92      | 822,52        | 899,14        | 865,35        |
| Trinidad                                       | 1.627,22      | 1.610,23      | 1.611,95      | 1.138,73      | 1.103,80      | 1.074,21      |
| Cobija                                         | 2.280,22      | 1.844,78      | 2.042,67      | 1.267,93      | 1.318,36      | 1.613,92      |

*Nota.* Denuncias de delitos (Choque Soto y otros, 2023, p. 35).

## 2.16 MÉTRICAS DE CALIDAD

### 2.16.1 ISO 25010

Esta ISO se trata de la evolución de la ISO 9126. (Moreno, 2018, p. 20) indica que esta norma revisa el marco propuesto, modificando y añadiendo nuevas características que deben ser analizadas para determinar la calidad del software que, al igual que su predecesora, se dividen en subcategorías. Este modelo establece la calidad en base a tres marcos diferentes, según el ámbito donde vaya a ser aplicado. El primer marco es el Modelo de Calidad de uso, orientado a los stakeholders, a los que va dirigido el producto software. El Modelo de Calidad

de producto, centrado en los propios desarrolladores del software. Y finalmente el modelo de Calidad de datos, que está recogido en la ISO 25012.

Se pueden encontrar un total de ocho características diferentes, ampliando en dos las de su predecesora, donde algunas de las categorías ya definidas han sido reformuladas y adaptadas. Estas características son:

**Figura 17.**

*Estructura de la ISO 25010*



*Nota.* Diseño de la estructura de la ISO 25010 (Moreno, 2018, p. 20).

**Adecuación funcional:** categoría, que al igual que en su predecesora, hace referencia a las funciones que implementa el software y satisface las necesidades implícitas de los usuarios, bajo unas condiciones determinadas. A su vez está dividida en diferentes subcategorías:

- **Completitud funcional:** si el software es capaz de cubrir todas las necesidades del usuario objetivo
- **Correlación funcional:** si es capaz de sacar unos resultados correctos y coherentes con los esperados.
- **Pertinencia funcional:** si proporciona un conjunto de funciones para las tareas y objetivos que sean las que el usuario necesita.



**Eficiencia de desempeño:** esta característica es medida en conjunto a atributos que evalúan el rendimiento del software en relación con la cantidad de recursos que emplea.

Dentro de esta categoría podemos encontrar:

- **Comportamiento temporal:** el tiempo de respuesta y procesamiento empleado por el software, además de la capacidad de throughput del mismo bajo condiciones determinadas
- **Utilización de recursos:** la cantidad de recursos que utiliza el software cuando está ejecutando su labor bajo unas condiciones determinadas.
- **Capacidad:** relacionado con la capacidad del sistema para funcionar bajo ciertas circunstancias extremas con relación a los requisitos.

**Compatibilidad:** hace referencia a la capacidad de varios sistemas a intercambiar información entre sí, bajo la posibilidad de compartir un mismo entorno, hardware o software.

Se divide en dos subcategorías:

- **Coexistencia:** si el software puede coexistir junto a otro de manera independiente, compartiendo recursos comunes sin perder rendimiento.
- **Interoperabilidad:** si dos o más sistemas son capaces de intercambiar datos y utilizar esa información para cumplir su labor.

**Usabilidad:** esta característica tiene en cuenta la capacidad del sistema para ser aprendido, comprendido, usado y el grado de aceptación del usuario con relación a su apariencia visual. De manera general, es evaluada por los propios usuarios del sistema. Es desglosada en varias subcategorías:

- **Capacidad para reconocer su adecuación:** si es capaz de ser entendido por un usuario y determinar si cubre sus necesidades.
- **Capacidad de aprendizaje:** relacionado con el tiempo que le lleva a un usuario aprender su funcionamiento.
- **Capacidad para ser usado:** adecuación del sistema para ser manejado y controlado por los usuarios.
- **Protección contra errores de usuario:** si el sistema es capaz de gestionar el comportamiento no esperado de los usuarios.
- **Estética de la interfaz de usuario:** determina si la interfaz es amigable para el usuario.
- **Accesibilidad:** si el sistema es capaz de adaptarse a determinadas características y discapacidades de los usuarios.

**Fiabilidad:** capacidad del software para desempeñar unas funciones específicas, bajo unas circunstancias y durante un tiempo determinado. Está subdivida en:

- **Madurez:** capacidad para satisfacer las necesidades en unas condiciones normales
- **Disponibilidad:** si el sistema es capaz de estar disponible y accesible cuando se requiere su utilización.
- **Tolerancia a fallos:** en el caso de que existan fallos de hardware o software, si el sistema es capaz de seguir funcionando correctamente.

- **Capacidad de recuperación:** si es capaz de recuperar los datos directamente afectados y restablecer un estado deseado del sistema tras un fallo o una interrupción.

**Seguridad:** está relacionado con la capacidad del sistema para proteger la información que maneja, impidiendo acceso a personas o sistemas no autorizados para que no sean capaces de leer o modificar su contenido. Se subdivide en varias categorías:

- **Confidencialidad:** si es capaz de proteger frente accesos no autorizados, ya sean intencionales o no.
- **Integridad:** si el sistema es capaz de prevenir modificaciones tanto de los datos como de los programas que almacena.
- **No repudio:** capacidad para realizar un seguimiento de todas las acciones realizadas dentro del sistema o los eventos que ocurren durante su funcionamiento.
- **Responsabilidad:** si existe un responsable de las acciones que ocurren en un sistema.
- **Autenticidad:** es capaz de demostrar la identidad de los usuarios y los recursos.

**Mantenibilidad:** hace referencia a la capacidad de un producto software para ser modificado de manera efectiva y eficiente. Esta característica se subdivide a su vez en:

- **Modularidad:** capacidad para modificar una de sus partes y que produzca un impacto mínimo en la totalidad del sistema.
- **Reusabilidad:** si el software está construido de tal manera que pueda ser utilizado en otros sistemas.

- **Analizabilidad:** facilidad con la que se puede evaluar el impacto que tiene un determinado cambio en el resto del sistema, descubrir las deficiencias o los posibles puntos de fallo.
- **Capacidad para ser modificado:** si el software posee la capacidad para ser modificado de manera eficiente y efectiva sin introducir ningún fallo.
- **Capacidad para ser probado:** si tiene la posibilidad de crear unos criterios de prueba con los que se pueda evaluar si cumple los requisitos determinados.

**Portabilidad:** si el sistema posee la capacidad de ser transferido a otro sistema con diferente hardware o software y funcionar de la manera esperada. Se divide en:

- **Adaptabilidad:** si es capaz de adaptarse de manera efectiva y eficiente a diferentes entornos con determinado hardware, software, operaciones o uso.
- **Capacidad para ser instalado:** facilidad con la que el software puede ser instalado o desinstalado de forma exitosa y completa.
- **Capacidad para ser reemplazado:** si el software es capaz de sustituir a otro que cumple el mismo propósito y está ejecutando en el mismo entorno.

## 2.17 EVALUACIÓN DE COSTOS COCOMO II

COCOMO II (Gómez y otros, 2010, p. 6), está compuesto por tres modelos que corresponden a distintos niveles de detalle y precisión. Mencionados en orden creciente son: Modelo Básico, Intermedio y Detallado. La estimación es más precisa a medida que se toman en cuenta mayor cantidad de factores que influyen en el desarrollo de un producto de software. COCOMO II permite estimar cómo se distribuye el esfuerzo y el tiempo en las distintas fases

del desarrollo de un proyecto y dentro de cada fase, en las actividades principales. Las fases consideradas por COCOMO II son:

**Diseño del Producto (PD):** se define la arquitectura del hardware, software y las estructuras de datos y control. También se desarrolla un bosquejo del manual del usuario y los planes de aceptación y testeo.

**Diseño Detallado (DD):** codificación y Testeo de Unidades (CT) En estas dos fases el diseño global de la fase anterior es implementado, creando las componentes de software, que son testeadas y evaluadas individualmente.

**Integración y Testeo (IT):** se fusionan todas las componentes de software desarrolladas con el fin de lograr que el producto de software funcione correctamente. Los requerimientos definidos son usados para controlar las aptitudes del producto liberado.

Los costos y tiempos de las fases excluidas (Requerimientos y Mantenimiento) deben ser estimados en forma separada empleando otros modelos. Se distinguen las siguientes actividades principales:

**Análisis de Requerimientos:** determinación, especificación, revisión y actualización de la funcionalidad, performance e interfase del software.

**Diseño del producto:** determinación, especificación, revisión y actualización de la arquitectura de hardware y software, diseño del programa y diseño de la base de datos.

**Programación (Diseño detallado + Unit Test):** comprende el diseño detallado, codificación, testeos unitarios e integración de las componentes individuales de software. Incluye planificación del personal de programación, adquisición de herramientas, desarrollo de la base de datos, y documentación al nivel de componente.

**Planificación del Testeo:** especificación, revisión y actualización de los planes de testeo del producto y de aceptación del mismo. Adquisición de herramientas de testeo y lotes de datos de prueba.

**Verificación y Validación:** ejecución de la validación y verificación de requerimientos, diseño, testeos del producto y testeos de aceptación. Adquisición de herramientas de V&V. Se hacen las preguntas ¿estaremos construyendo el producto correcto?, y ¿estaremos construyendo correctamente el producto?

**Actividades de oficina:** tareas referidas al gerenciamiento del proyecto. Incluye la administración de contratos y subcontratos, relaciones con los clientes.

Administración de la Configuración y Aseguramiento de la Calidad (CM/QA): La Administración de la Configuración incluye la identificación del producto, control de cambios, administración de las librerías de soporte, etc. Aseguramiento de la Calidad abarca el desarrollo y monitoreo de estándares, auditorías técnicas del producto de software y el proceso de desarrollo.

**Manuales:** desarrollo y actualización de los manuales de usuario, operador y mantenimiento.

### **2.17.1 Características**

Es una herramienta basada en las líneas de código, la cual la hace muy poderosa para la estimación de costos y no como otros que solamente miden el esfuerzo en basándose al tamaño.

Representa el más extenso modelo empírico para la estimación de software.

Existen herramientas automáticas que estiman costos basados en COCOMO como ser: Costar, COCOMO 81

### **2.17.2 Modelos de Estimación**

En el modelo COCOMO 81, uno de los factores más importantes que influye en la duración y el costo de un proyecto de software es el Modo de Desarrollo. Todo proyecto corresponde a uno de los siguientes tres modos:

**Modo Orgánico (Organic):** en esta clasificación se encuentran proyectos desarrollados en un ambiente familiar y estable. El producto a elaborar es relativamente pequeño y requiere pocas innovaciones tecnológicas en lo que refiere a algoritmos, estructuras de datos e integración de hardware.

La mayoría de las personas conectadas con el proyecto tienen gran experiencia en sistemas relacionados dentro de la organización, y un entendimiento acabado de cómo el sistema contribuirá a los objetivos de la organización. Esto significa que todo el equipo de desarrollo podrá contribuir en las etapas iniciales del proyecto sin generar confusión en las comunicaciones debido a que todos conocen qué tarea deben realizar.

Además, un proyecto clasificado dentro del modo orgánico, es relativamente flexible en el cumplimiento de los requerimientos, especificaciones de interfase y tiempos de entrega.

Muy pocos proyectos de modo orgánico han desarrollado productos con más de 50 KSLOC<sup>1</sup> de nuevo software. En los casos de productos más extensos, han sido construidos frecuentemente a partir de software existente.

Estas características permiten decir que los proyectos que se encuentran en este modo tienen una gran productividad y una pequeña diseconomía de escala. Ejemplos de software que se encuentran bajo esta clasificación son:

- Modelos de negocios.
- Modelos científicos.
- Sistemas operativos de pequeña escala.

**Modo Semiacoplado (Semidetached):** es un modelo para productos de software de tamaño y complejidad media. Las características de los proyectos se consideran intermedias a las de los modos Orgánico y Empotrado. Esto implica:

Que el equipo de desarrollo:

- Tiene un nivel intermedio de experiencia y conocimiento del sistema en desarrollo.
- Está conformado por algunas personas con vasta experiencia y otras inexpertas en el campo de aplicación.
- Está constituido por personas con amplios conocimientos solo en algunos aspectos.
- Con respecto al cumplimiento de especificaciones de interfase y funcionalidad:
- Son sistemas que presentan niveles variados de exigencia, algunas interfases rigurosas (auditadas por el gobierno) y otras interfases muy flexibles (mensajes de display al operador).

Los productos tienen un tamaño que llega a 300 KSLOC.

Ejemplos de software que se encuentran en esta clasificación son:

- Sistemas de control de producción.



- Sistemas de procesamiento de transacciones.
- Administradores de Bases de Datos.

**Modo Empotrado (Embedded):** en esta clasificación están incluidos proyectos de gran envergadura que operan en un ambiente complejo con altas restricciones de hardware, software y procedimientos operacionales, tales como los sistemas de tráfico aéreo.

Se espera que el software no solo conforme las especificaciones, sino también que sea estable frente a cambios y dificultades producidas en el ambiente. Es decir, estos proyectos no tienen opción de negociar cambios y/o arreglos provocados por modificaciones en los requerimientos y/o en las especificaciones de interfase. El proyecto dedica un gran esfuerzo para adaptarse a los cambios y arreglos, en asegurar que el software cumpla verdaderamente las especificaciones y que los cambios se efectúen correctamente. Esto implica altos costos en los procesos de Verificación y Validación y en la Administración de la Configuración, contribuyendo así a la disminución de la productividad y al aumento de las deseconomías de escala en grandes proyectos.

El equipo que interviene en el proyecto tiene un conocimiento general de los objetivos del mismo y una moderada experiencia en el tema. En general, el líder del proyecto destina en las primeras etapas a un grupo pequeño del equipo a las tareas de análisis, evitando así los problemas acarreados por la sobrecarga de comunicaciones. Una vez que se completa el diseño global del producto, la mejor estrategia es delegar a un gran equipo de programadores las tareas de diseño detallado, codificación y testeo.

Ejemplos de software que se encuentran en esta clasificación son:

- Sistemas complejos de procesamiento de transacciones.
- Sistemas operativos de gran escala.

### 2.17.3 Modelos Básicos

El Modelo Básico estima el esfuerzo y el tiempo empleado en el desarrollo de un proyecto de software usando dos variables predictivas denominadas factores de costo.

$$\text{Esfuerzo: } PM = A \times (KSLOC)^B$$

$$\text{Cronograma: } TDEV = C \times (PM)^D$$

Donde:

Donde:

- **PM** es el esfuerzo estimado. Representa los meses persona necesarios para ejecutar el proyecto.
- **KSLOC** es el tamaño del software a desarrollar en miles de líneas de código.
- **A y B** son coeficientes que varían según el Modo de Desarrollo (Orgánico, Semi acoplado, Empotrado).
- **TDEV** representa los meses de trabajo que se necesitan para ejecutar el proyecto.
- **C y D** son coeficientes que varían según el Modo de Desarrollo (Orgánico, Semi acoplado, Empotrado).

**Tabla 7.**

*Modelo Básico*

| Modo de Desarrollo | Esfuerzo                       | Cronograma                      |
|--------------------|--------------------------------|---------------------------------|
| Orgánico           | $PM=2.4 \times (KSLOC)^{1.05}$ | $TDEV = 2.5 \times (PM)^{0.38}$ |
| Semi acoplado      | $PM=3.0 \times (KSLOC)^{1.12}$ | $TDEV = 2.5 \times (PM)^{0.35}$ |
| Empotrado          | $PM=3.6 \times (KSLOC)^{1.20}$ | $TDEV = 2.5 \times (PM)^{0.32}$ |

*Nota.* Modelo Básico de Cocomo II Gómez y otros, 2010, p. 9),

### 2.17.4 Modelos Intermedios

Comparado con el modelo anterior, este provee un nivel de detalle y precisión superior, por lo cual es más apropiado para la estimación de costos en etapas de mayor especificación.

COCOMO Intermedio incorpora un conjunto de quince variables de predicción que toman en cuenta las variaciones de costos no consideradas por COCOMO Básico.

Existen diversos factores a considerar en el desarrollo de un buen modelo de estimación de costos de un proyecto de software. Para reducir el número a una cantidad relativamente manejable se utilizaron fundamentalmente dos principios:

- Eliminar aquellos factores que son significativos solamente en una fracción relativamente pequeña o en situaciones especiales.
- Eliminar los factores que están altamente correlacionados con el tamaño y comprimir aquellos factores correlacionados entre sí

Los factores seleccionados se agrupan en cuatro categorías:

- Atributos del producto de software
  - RELY Confiabilidad Requerida
  - DATA Tamaño de la Base de Datos
  - CPLX Complejidad del Producto
- Atributos del hardware
  - TIME Restricción del Tiempo de Ejecución
  - STOR Restricción del Almacenamiento Principal
  - VIRT Volatilidad de la Máquina Virtual\*
  - TURN Tiempo de Respuesta de la computadora expresado en horas

- Atributos del personal involucrado en el proyecto
  - ACAP Capacidad del Analista
  - AEXP Experiencia en Aplicaciones Similares
  - PCAP Capacidad del Programador
  - VEXP Experiencia en la máquina virtual
  - LEXP Experiencia en el Lenguaje de Programación
  
- Atributos propios del proyecto
  - MODP Prácticas Modernas de Programación.
  - TOOL Uso de Herramientas de Software.
  - SCED Cronograma de Desarrollo Requerido.

**Tabla 8.***Modelo Intermedio*

| <b>Modo de Desarrollo</b> | <b>Esfuerzo Nominal</b>            | <b>Esfuerzo Ajustado</b>    | <b>Cronograma</b>                  |
|---------------------------|------------------------------------|-----------------------------|------------------------------------|
| Orgánico                  | PMnominal=<br>$3.2x(KSLOC)^{1.05}$ | $PM=3.2xEAFx(KSLOC)^{1.05}$ | TDEV=2.5 x<br>(PM) <sup>0.38</sup> |
| Semiacoplado              | PMnominal=<br>$3.0x(KSLOC)^{1.12}$ | $PM=3.0xEAFx(KSLOC)^{1.02}$ | TDEV=2.5 x<br>(PM) <sup>0.35</sup> |
| Empotrado                 | PMnominal=<br>$2.8x(KSLOC)^{1.20}$ | $PM=3.6xEAFx(KSLOC)^{1.20}$ | TDEV=2.5 x<br>(PM) <sup>0.32</sup> |

*Nota.* Ecuaciones del Modelo Intermedio Gómez y otros, 2010, p. 13).

### **2.17.5 Modelos Detallados**

El Modelo Detallado provee los medios para generar estimaciones con mayor grado de precisión y detalle. Difiere del Modelo Intermedio en dos aspectos principales que ayudan a superar las limitaciones mencionadas.

**Jerarquía de niveles del producto:** en el Modelo Intermedio se pueden calcular valores diferentes de los factores de costo para cada componente de software. Este proceso puede resultar muy tedioso e innecesariamente repetitivo si las componentes están agrupadas en subsistemas de características generales similares. Para subsanar este problema, el Modelo Detallado aplica al producto de software una descomposición jerárquica de tres niveles. En el nivel inferior, nivel de módulo, la estimación se basa en el número de líneas de código del módulo (SLOC) y aquellos factores que tienden a variar en ese nivel: complejidad del módulo y adaptación del software existente, nivel de capacidad y experiencia del

programador, con el lenguaje y la máquina virtual sobre la que se construirá el software. El segundo nivel, nivel de subsistema, está descrito por el resto de los factores de costo que pueden variar de un subsistema a otro, pero que tienden a ser los mismos para todos los módulos dentro de un subsistema. Entre ellos se encuentran: restricciones de tiempo y espacio, capacidad del analista, herramientas, etc. El nivel superior, nivel de sistema, se usa para aplicar las ecuaciones de esfuerzo nominal y cronograma y calcular las estimaciones tanto para todo el proyecto como para cada fase.

**Multiplicadores de Esfuerzo (EM Effort Multipliers):** sensitivos a las fases. El modelo Detallado provee un conjunto de multiplicadores diferentes para cada factor de costo, según la fase del ciclo de desarrollo que se considere. De esta forma los multiplicadores se utilizan para determinar el esfuerzo requerido para completar cada fase, ver Tabla 4 y Tabla 5 que muestran los factores que afectan al proyecto al nivel de módulo y subsistema respectivamente. Las ecuaciones fundamentales de este modelo son similares a las del modelo COCOMO Intermedio, la única diferencia reside en el cálculo del Factor de Ajuste del Esfuerzo (EAF). El procedimiento incluye el cálculo de un Factor de Ajuste del Esfuerzo al nivel de módulo (EAFM) y otro al nivel de subsistema (EAFS) (Gómez y otros, 2010, p. 14).



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# **CAPÍTULO III**

## DISEÑO METODOLÓGICO



### 3 CAPÍTULO III

#### 3.1 DISEÑO METODOLÓGICO

##### 3.1.1 *Tipo de Investigación*

La metodología cuantitativa usualmente parte de cuerpos teóricos aceptados por la comunidad científica con la base en los cuales formula hipótesis sobre relaciones esperadas entre las variables que hacen parte del problema que se estudia. Su constatación se realiza mediante la recolección de información cuantitativa orientada por conceptos empíricos medibles, derivados de los conceptos teóricos con los que se construyen las hipótesis conceptuales. El análisis de la información recolectada tiene por fin determinar el grado de significación de las relaciones previstas entre las variables. El procedimiento que se sigue de la teoría, la operacionalización de las variables, la recolección, el procesamiento de los datos y la interpretación. (Álvarez, 2011, p. 11).

##### 3.1.1.1 **Investigación Correlacional**

Esta investigación fundamentalmente determina el grado en el cual las variaciones en uno o varios factores son concomitantes con la variación en otro u otros factores. La existencia y fuerza de esta covariación. Es conveniente tener en cuenta que esta covariación no significa que entre los factores existan relaciones de casualidad, estas se determinan por otros criterios que además de covariación (Álvarez, 2011, p. 101).

Características: algunas características sobresalientes de este tipo de estudio.

- a) Es indicado en situaciones complejas que importa relacionar variables, pero no las cuales no es posible el control experimental.
- b) Permite medir e interrelacionar múltiples variables simultáneamente en situaciones de observación naturales.



- c) Permite identificar asociaciones entre varias variables, pero hay que prevenir que ellas sean espurias o falsas, introduciendo los controles estadísticos ofrecidos.
- d) Es menos riesgoso que el tipo de investigación experimental porque no hay posibilidad de, manipular la variable o variables independientes ni de controlarlas rigurosamente.

#### Etapas de la investigación correlacional

1. Definir el problema.
2. Revisar la literatura.
3. Determinar el diseño correlacional:
  - a) Identificar las variables pertinentes.
  - b) Seleccionar los sujetos apropiados.
  - c) Determinar cuáles instrumentos son los apropiados para detener los datos.
  - d) Seleccionar las técnicas de correlación estadísticas apropiadas para los datos.
4. Recoger los datos.
5. Analizar los datos por medio de las correspondientes técnicas correlacionales e interpretar los resultados.

### **3.1.2 Diseño de la Investigación**

El término diseño se refiere al plan o estrategia concebida para obtener la información que se desea con el fin de responder al planteamiento del problema

En el enfoque cuantitativo se utilizan los diseños para analizar la certeza de las hipótesis formuladas en un contexto en particular o para aportar evidencias respecto a los lineamientos de la investigación (Sampieri, 2006, p. 128).

#### **3.1.2.1 Diseño no Experimental**

Es la investigación en la que se realiza sin manipular deliberadamente variables, es el estudio en los que no hacemos variar de forma intencional las variables independientes para ver el efecto que causa en otras variables. En la investigación no experimental es observar fenómenos tal como se dan en un contexto natural, para analizarlos, en un estudio no experimental no se genera ninguna situación, sino que se observan situaciones ya existentes, no provocadas intencionalmente.

En la investigación no experimental no se tiene un control directo sobre dichas variables ni se puede influir en ellas, porque ya sucedieron al igual que sus efectos, en un estudio no experimental la investigación es sistemática y empírica (Sampieri, 2006, p. 153).

#### **3.1.2.2 Tipos de Diseños no Experimentales**

La siguiente manera de clasificar dicha investigación: por su dimensión temporal o el número de momentos o puntos en el tiempo en los cuales se recolectan datos. En algunas ocasiones la investigación se centra en:

- a) Analizar cuál es el nivel o modalidad de una o diversas variables en un momento dado.

- b) Evaluar una situación, comunidad, evento, fenómeno o contexto en un punto del tiempo.
- c) Determinar o ubicar cuál es la relación entre un conjunto de variables en un momento.

El diseño apropiado (con un enfoque no experimental) es el transversal o transeccional. Ya sea que su alcance inicial o final sea exploratorio, descriptivo, correlacional o explicativo.

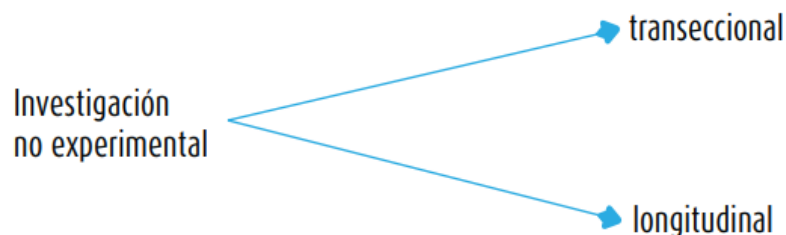
Otras veces, la investigación se enfoca en:

- a) Estudiar cómo evolucionan una o más variables o las relaciones entre ellas.
- b) Analizar los cambios al paso del tiempo de un evento, comunidad, proceso, fenómeno o contexto.

En situaciones como estas el diseño apropiado (en un enfoque no experimental) es el longitudinal. Dicho de otro modo, los diseños no experimentales se pueden clasificar en transeccionales y longitudinales. (Sampieri, 2006, p. 154)

### Figura 18.

*Diseño no experimental.*



*Nota.* Clasificación del diseño no experimental (Sampieri, 2006, p. 154).

### 3.1.3 *Variables de la Investigación*

La variable es una característica o propiedad de la realidad que puede variar entre individuos o conjuntos, cualquier aspecto o propiedad de la realidad que sea susceptible de asumir valores, esto es, de variar de una unidad de observación a otra, de un tiempo a otro en una misma unidad de observación. (Álvarez, 2011, p. 91).

- **Independiente:** las supuestas causas, es la característica o propiedad que se supone el antecedente o causa del fenómeno estudiado.
  - **Variable Independiente:** Modelo Predictivo.
  
- **Dependiente:** el efecto supuesto, los cambios esperados o producidos por la variable independiente, el resultado atribuible a la existencia o manipulación de la variable independiente.
  - **Variable Dependiente:** Índices delictivos.

### 3.1.4 *Ambiente de la Investigación*

El ambiente puede ser tan variado (un hospital, una o varias empresas, el área de producción de una fábrica, una obra civil, una zona selvática si estudiamos el comportamiento de una especie animal, una comunidad indígena, una universidad, una plaza pública, un consultorio, una casa donde sesiona un grupo). Y el contexto implica una definición geográfica, pero es inicial, puesto que puede variar, ampliarse o reducirse. Imaginemos que queremos estudiar los valores de ciertos estudiantes universitarios mediante la observación es una interpretación detallada de casos, seres vivos, personas, objetos, lugares específicos y eventos del contexto (Sampieri, 2006, p. 365).

#### **3.1.4.1 Muestreo**

El análisis de la investigación, se recolectó una muestra probabilística. Donde se limita el universo de estudio a los distritos que conforman al Municipio de El Alto.

La información sobre los distritos fue proporcionada por el Gobierno Autónomo Municipal de El Alto.

También se cuenta con los datos recolectados de los hechos delictivos, datos proporcionados por el Comando General de la Policía Boliviana de las gestiones 2017 hasta la gestión 2023.

### **3.1.5 Descripción de la Metodología a usar**

#### **3.1.5.1 Método Científico**

Para que se dé el conocimiento científico en forma razonada y válida, un método general de investigación deberá cumplir unos requisitos específicos de la ciencia.

El profesor Muñoz (1998), basado en Mario Bunge, presenta un inventario de las principales características de la ciencia (Bernal Torres, 2010, p. 67).

**El conocimiento científico es fáctico (verdadero):** la ciencia intenta descubrir los hechos como son, independientemente del valor comercial o emocional que se les otorgue.

**El conocimiento trasciende los hechos:** la investigación científica observa los hechos, descarta los que no le son útiles, produce acontecimientos nuevos y los explica después de observarlos y describirlos, y de realizar cierta experimentación.

**La ciencia es analítica:** la ciencia aborda problemas específicos y trata de descomponer sus elementos, con la finalidad de entenderlos de manera integral y en sus relaciones con el medio que los rodea.

**La investigación científica es especializada:** a pesar de utilizar muchas y muy variadas técnicas de observación y experimentación, métodos, procedimientos, análisis y alcances, etcétera, la investigación científica se enmarca en una disciplina en particular.

**El conocimiento científico es claro y preciso:** la ciencia es mucho más que un método organizado, pues constituye una alternativa de conocimiento que se apoya en métodos y técnicas comprobados para darle claridad a la investigación y precisar sus resultados.

**El conocimiento científico es comunicable:** la comunicación de resultados y técnicas utilizadas para lograr conocimiento científico perfecciona la ciencia, y multiplica las posibilidades de confirmación, refutación y expansión.

**El conocimiento científico es verificable:** para que el conocimiento sea admitido como ciencia, tendrá que someterse a la comprobación y a la crítica de la comunidad científica.

**La investigación científica es metódica:** cualquier trabajo de investigación científica se fundamenta en un método, unas técnicas y unos procedimientos que han resultado eficaces en el pasado.

**El conocimiento científico es sistemático:** la ciencia es un sistema de ideas interconectadas que buscan la verdad. El fundamento de la ciencia es un conjunto ordenado de principios, hipótesis y resultados, que se conjugan con un método lógico y coherente que les da racionalidad y validez.

**El conocimiento científico es general:** la ciencia ubica los hechos singulares en pautas generales y promueve que de enunciados particulares derivan esquemas más amplios. En este punto es importante recordar que el debate de la filosofía de la ciencia se refiere a que el conocimiento científico es hipotético-deductivo y no inductivo, es decir, que va de lo general a lo particular y no inversamente.

**El conocimiento científico es legal:** el conocimiento científico busca leyes y se apoya en pautas generales. Estas leyes deben servir como marco de referencia y no como una norma rígida.

**La ciencia es explicativa:** la ciencia no solo se conforma con realizar la descripción detallada de un fenómeno o una situación, sino que busca entender el porqué de los hechos.

**El conocimiento científico es predictivo:** la ciencia supone los fenómenos del pasado para proyectarlos al futuro. A partir de resultados de investigaciones se predicen nuevos hechos y consecuencias.

**La ciencia es abierta:** el conocimiento científico, a pesar de fundamentarse en leyes, considera que el conocimiento actual es susceptible de corregirse y remplazarse.

**La ciencia es útil:** la ciencia busca la verdad y la objetividad de los resultados, pero en particular solucionar problemas. Evidentemente, estas quince características responden a un tipo específico de modelo o idea de ciencia, como es la ciencia fáctica o el modelo general de ciencia positivista, uno de los modelos de investigación científica que existen.

## 3.2 HERRAMIENTAS

### 3.2.1 *Python*

presenta una serie de ventajas que lo hacen muy atractivo, tanto para su uso profesional como para el aprendizaje de la programación. Entre las más interesantes desde el punto de vista didáctico tenemos.

1. Python es un lenguaje muy expresivo, es decir, los programas Python son muy compactos: un programa Python suele ser bastante más corto que su equivalente en lenguajes como C. (Python llega a ser considerado por muchos un lenguaje de programación de muy alto nivel.)

2. Python es muy legible. La sintaxis de Python es muy elegante y permite la escritura de programas cuya lectura resulta más fácil que si utilizáramos otros lenguajes de programación.
3. Python ofrece un entorno interactivo que facilita la realización de pruebas y ayuda a despejar dudas acerca de ciertas características del lenguaje.
4. El entorno de ejecución de Python detecta muchos de los errores de programación que escapan al control de los compiladores y proporciona información muy rica para detectarlos y corregirlos. Python puede usarse como lenguaje imperativo procedimental o como lenguaje orientado a objetos.
5. Posee un rico juego de estructuras de datos que se pueden manipular de modo sencillo.

Estas características hacen que sea relativamente fácil traducir métodos de cálculo a programas Python. Una ventaja fundamental es la gratuidad de su intérprete (Marzal & Gracia, 2009, p. 16).

### **3.2.2 Postgres**

PostgreSQL es un avanzado sistema de bases de datos relacionales basado en Open Source. Esto quiere decir que el código fuente del programa está disponible a cualquier persona libre de cargos directos, permitiendo a cualquiera colaborar con el desarrollo del proyecto o modificar el sistema para ajustarlo a sus necesidades. PostgreSQL está bajo licencia BSD. Un sistema de base de datos relacionales es un sistema que permite la manipulación de acuerdo con las reglas del álgebra relacional. Los datos se almacenan en tablas de columnas y renglones. Con el uso de llaves, esas tablas se pueden relacionar unas con otras.



PostgreSQL se caracteriza por ser un sistema estable, de alto rendimiento, gran flexibilidad, ya que funciona la mayoría de los sistemas Unix, además tiene características que permiten extender fácilmente el sistema. PostgreSQL puede ser integrada al ambiente Windows permitiendo de esta manera a los desarrolladores, generar nuevas aplicaciones o mantener las ya existentes. Permite desarrollar o migrar aplicaciones desde Access, Visual Basic, Foxpro, Visual Foxpro, C/C++ Visual C/C++, Delphi, para que utilicen a PostgreSQL como servidor de BD; Por lo expuesto PostgreSQL se convierte en una gran alternativa al momento de decidirse por un sistema de bases de datos (Denzer, 2002, p. 2).

### **3.2.3 *Visual Studio Code***

Visual Studio Code es un editor de código fuente ligero, pero potente que se ejecuta en su escritorio y está disponible para Windows, macOS y Linux. Incluye soporte incorporado para JavaScript, TypeScript y Node.js y tiene un rico ecosistema de extensiones para otros lenguajes (como C ++, C #, Java, Python, PHP, Go) y tiempos de ejecución (como .NET y Unity).

Visual Studio Code es un editor de código fuente. Es compatible con varios lenguajes de programación y un conjunto de características que pueden o no estar disponibles para un idioma dado. Muchas de las características de Visual Studio Code no están expuestas a través de los menús o la interfaz de usuario. Más bien, se accede a través de la paleta de comandos o a través de archivos. Json.

También incluye soporte para la depuración, control integrado de Git, resaltado de sintaxis, finalización inteligente de código, fragmentos y refactorización de código. También es personalizable, por lo que los usuarios pueden cambiar el tema del editor, los atajos de teclado y las preferencias. Es gratuito y de código abierto, aunque la descarga oficial está bajo software propietario, requiriendo tus datos de uso del programa legalmente. (Emanuel, 2019, p. 33).

### **3.3 HERRAMIENTAS A USAR**

#### **3.3.1 *Técnicas de Investigación e Instrumentos***

##### **3.3.1.1 Técnica de Investigación**

Método cuantitativo o método tradicional se fundamenta en la medición de las características de los fenómenos sociales, lo cual supone derivar de un marco conceptual pertinente al problema analizado, una serie de postulados que expresen relaciones entre las variables estudiadas de forma deductiva. Este método tiende a generalizar y normalizar resultados

La investigación cuantitativa parte de cuerpos teóricos aceptados por la comunidad científica, en tanto que la investigación cualitativa pretende conceptuar sobre la realidad, con base en la información obtenida de la población o las personas estudiadas (Bernal Torres, 2010, p. 60).

#### **3.3.2 *Instrumentos de investigación***

Los instrumentos de investigación son las herramientas que se utilizan para la recolección de datos, el cual se basa en la muestra seleccionada y con el correcto análisis aportaremos conocimientos significativos en nuestro tema de investigación. Los pasos para la recolección de datos son los siguientes:

- Identificar fuentes.
- Ubicar fuentes.
- Identificar cómo se recopilan los datos.
- Determinar cómo preparar los datos.

Para que la recolección de datos aporte valor debemos medir dichos datos. Medir significa asignar números, símbolos o valores a las propiedades de objetos o eventos de acuerdo con reglas.

Estos instrumentos deben contar con las siguientes características:

- **Confiabilidad:** nos referimos a confiabilidad cuando adquirimos diferentes electrodomésticos en el mismo almacén y estos funcionan sin molestia alguna, además cumplen con su periodo de vida útil. Por ende, si vamos a comprar al mismo almacén esto se interpreta que confiamos en la calidad de los productos que venden. Por lo tanto, la confiabilidad se genera cuando se repite un instrumento con el mismo sujeto y produce el mismo resultado o similar. Por esta razón se determina que la confiabilidad es una medida de estabilidad.
  
- **Validez:** Por lo general cuando se ingresa a un curso o seminario de cualquier área, matemáticas, por ejemplo, nos realizan un pequeño test en donde se pretende conocer el grado de conocimiento que tiene cada individuo. Por lo tanto, para que dicho test tenga validez tiene que estar acorde a lo que se desea evaluar. No sería válido hacer preguntas sobre hobbies, religión, grado de estudios, etc. Podemos concluir que la validez indica el grado en que un instrumento mide la variable que se busca medir.

Hay tres tipos de validez:

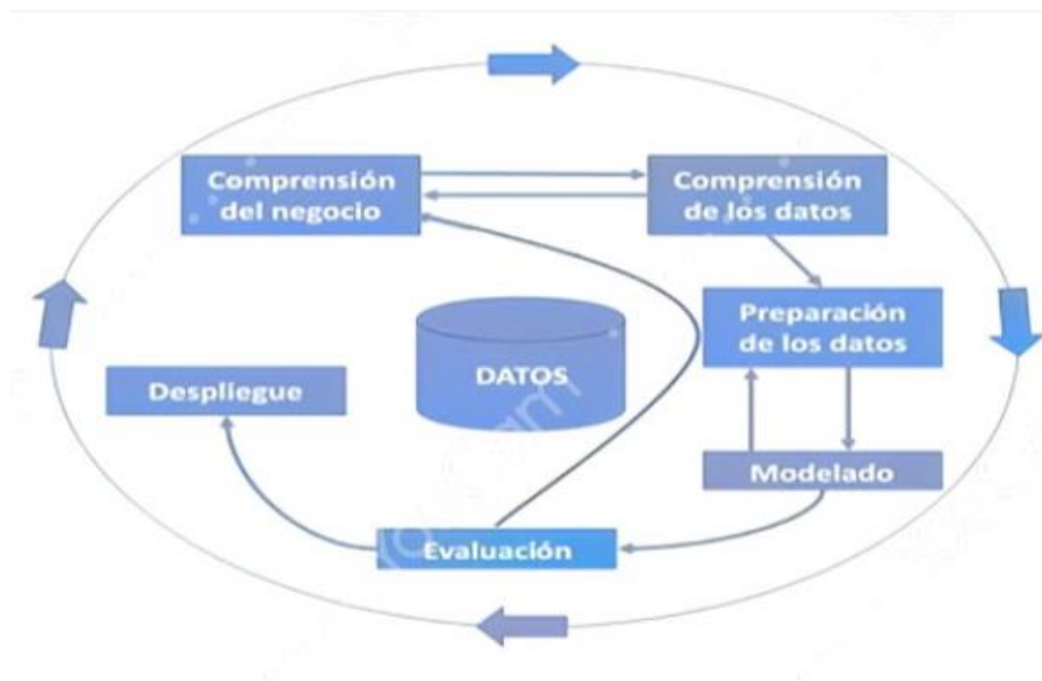
- Contenido: dominio específico de contenido de lo que se mide.
  
  - Criterio: su resultado al ser comparado con otro que mide lo mismo está correlacionado.
  
  - Constructo: el instrumento representa y mide un concepto teórico.
- 
- **Objetividad:** hace referencia al grado en que el instrumento es o no permeable a la influencia de sesgos y tendencias del investigador. Por ejemplo: Si reunimos un

grupo de niños de 1 a 12 años y les proporcionamos los siguientes juguetes: muñecas, animalitos e instrumentos musicales. Se comprobará que no todos los niños se sentirán atraídos a jugar con dichos juguetes, por lo general los niños de 9 a 12 años se interesan por los deportes, videojuegos, experimentos, lectura, etc. Por lo tanto, para que haya objetividad en el instrumento debe estar estandarizado, es decir, mismas condiciones e instrucciones. “Es importante recalcar que tanto la confiabilidad, validez y objetividad no deben tratarse de forma separada. Si el instrumento no posee una de estas características, este no será útil para el estudio. (Sampieri, 2006).

### 3.4 APLICACIÓN DE CRISP-DM

**Figura 19.**

*Metodología CRISP-DM*



*Nota.* Metodología CRISP-DM (Hernández Cedano, 2015, p. 24).

### **3.4.1 FASE I: *Comprensión del negocio***

En el Municipio de El Alto, originario de la provincia Pedro Domingo Murillo, situada en el departamento de La Paz, es destacada por ser la ciudad más joven y tener un alto crecimiento en población, llegando a ser una de las ciudades más pobladas de Bolivia.

El Municipio de El Alto está atravesando un incremento en los hechos delincuenciales, esta situación en consecuencia llegó a tener un gran impacto a tal punto de ser considerada como una de las ciudades más peligrosas, causantes de este suceso los hechos delictivos en el Municipio de El Alto

Esta investigación adopta un enfoque de Minería de Datos para generar información valiosa y predictiva a partir de los datos recopilados, permitiendo encontrar el índice de crecimiento de los hechos delictivos.

### **3.4.2 FASE II: *Comprensión de los datos***

#### **3.4.2.1 Recolección de los datos iniciales**

Los datos recopilados sirvieron como base fundamental para llevar a cabo esta investigación, estos datos hacen referencia al índice de hechos delictivos en el Municipio de El Alto, se realizó una solicitud formal de información sobre estos hechos delictivos al Comando General de la Policía Boliviana.

#### **3.4.2.2 Datos recolectados**

Los datos proporcionados del Comando General de la Policía Boliviana sobre los hechos delictivos, se obtuvieron en formato de hoja de cálculo Excel.

Figura 20.

## Datos de hechos Delictivos

| ID | ZONA              | NOM_CASO                | ENERO | FEBRERO | MARZO | ABRIL | MAYO | JUNIO | JULIO | AGOSTO | SEPTIEMBRE | OCTUBRE | NOVIEMBRE | DICIEMBRE | TOTAL_CASO | DISTRITO | AÑO  |
|----|-------------------|-------------------------|-------|---------|-------|-------|------|-------|-------|--------|------------|---------|-----------|-----------|------------|----------|------|
| 1  | 12 DE OCTUBRE     | LESIONES GRAVES Y LEVES | 10    | 3       | 19    | 0     | 16   | 13    | 13    | 6      | 6          | 7       | 2         | 1         | 96         | 1        | 2017 |
| 2  | 16 DE JULIO       | LESIONES GRAVES Y LEVES | 0     | 5       | 4     | 0     | 4    | 1     | 6     | 2      | 1          | 3       | 1         | 1         | 28         | 6        | 2017 |
| 3  | VILLA BOLIVAR D   | LESIONES GRAVES Y LEVES | 0     | 0       | 3     | 0     | 1    | 0     | 0     | 1      | 21         | 0       | 0         | 0         | 26         | 2        | 2017 |
| 4  | ALTO LIMA         | LESIONES GRAVES Y LEVES | 0     | 2       | 2     | 0     | 2    | 2     | 2     | 2      | 2          | 3       | 1         | 1         | 19         | 6        | 2017 |
| 5  | VILLA DOLORES     | LESIONES GRAVES Y LEVES | 0     | 0       | 1     | 0     | 4    | 0     | 3     | 2      | 0          | 5       | 0         | 1         | 16         | 3        | 2017 |
| 6  | SENKATA           | LESIONES GRAVES Y LEVES | 1     | 0       | 2     | 0     | 1    | 3     | 3     | 0      | 0          | 0       | 0         | 0         | 10         | 8        | 2017 |
| 7  | RIO SECO          | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 6    | 2     | 2     | 3      | 2          | 0       | 1         | 0         | 16         | 4        | 2017 |
| 8  | VILLA TUNARI      | LESIONES GRAVES Y LEVES | 1     | 1       | 0     | 0     | 1    | 0     | 0     | 0      | 3          | 3       | 0         | 1         | 10         | 4        | 2017 |
| 9  | CIUDAD SATELITE   | LESIONES GRAVES Y LEVES | 0     | 0       | 3     | 0     | 1    | 1     | 5     | 0      | 0          | 2       | 0         | 2         | 14         | 1        | 2017 |
| 10 | VILLA ADELA       | LESIONES GRAVES Y LEVES | 1     | 0       | 2     | 0     | 2    | 2     | 1     | 0      | 1          | 0       | 0         | 0         | 9          | 3        | 2017 |
| 11 | FERROPETROL       | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 2    | 2     | 0     | 0      | 1          | 1       | 0         | 0         | 6          | 6        | 2017 |
| 12 | SANTA ROSA        | LESIONES GRAVES Y LEVES | 2     | 0       | 1     | 0     | 0    | 0     | 1     | 0      | 1          | 3       | 0         | 0         | 8          | 1        | 2017 |
| 13 | VILLA INGENIO     | LESIONES GRAVES Y LEVES | 1     | 0       | 2     | 0     | 0    | 0     | 1     | 1      | 0          | 0       | 0         | 0         | 5          | 5        | 2017 |
| 14 | BALLIVIAN         | LESIONES GRAVES Y LEVES | 0     | 1       | 0     | 0     | 0    | 0     | 0     | 0      | 0          | 2       | 1         | 1         | 4          | 6        | 2017 |
| 15 | SAN MARTIN        | LESIONES GRAVES Y LEVES | 0     | 1       | 0     | 0     | 1    | 1     | 1     | 0      | 0          | 0       | 0         | 1         | 5          | 7        | 2017 |
| 16 | SAN LUIS PAMPA    | LESIONES GRAVES Y LEVES | 0     | 0       | 1     | 0     | 0    | 1     | 0     | 0      | 1          | 0       | 0         | 0         | 3          | 3        | 2017 |
| 17 | VILLA ESPERANZA   | LESIONES GRAVES Y LEVES | 0     | 1       | 1     | 0     | 0    | 1     | 1     | 0      | 1          | 0       | 1         | 0         | 6          | 6        | 2017 |
| 18 | HUAYNA POTOSI     | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 0    | 1     | 1     | 1      | 0          | 0       | 0         | 0         | 3          | 5        | 2017 |
| 19 | 21 DE SEPTIEMBRE  | LESIONES GRAVES Y LEVES | 2     | 0       | 0     | 0     | 1    | 1     | 0     | 0      | 0          | 0       | 0         | 0         | 4          | 7        | 2017 |
| 20 | KENKO             | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 1    | 0     | 2     | 0      | 0          | 0       | 0         | 0         | 3          | 2        | 2017 |
| 21 | COSMOS 79         | LESIONES GRAVES Y LEVES | 2     | 0       | 1     | 0     | 0    | 1     | 1     | 0      | 1          | 2       | 0         | 1         | 9          | 3        | 2017 |
| 22 | NUEVOS HORIZONTES | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 0    | 0     | 1     | 0      | 1          | 2       | 1         | 0         | 5          | 2        | 2017 |
| 23 | 25 DE JULIO       | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 2    | 1     | 0     | 0      | 0          | 0       | 0         | 0         | 3          | 4        | 2017 |
| 24 | 16 DE FEBRERO     | LESIONES GRAVES Y LEVES | 0     | 1       | 0     | 0     | 1    | 0     | 0     | 0      | 1          | 0       | 0         | 0         | 3          | 4        | 2017 |
| 25 | VENTILLA          | LESIONES GRAVES Y LEVES | 1     | 0       | 0     | 1     | 1    | 1     | 0     | 1      | 0          | 0       | 0         | 0         | 5          | 8        | 2017 |
| 26 | VILLA DOLORES F   | LESIONES GRAVES Y LEVES | 0     | 0       | 1     | 0     | 0    | 0     | 0     | 1      | 0          | 0       | 0         | 0         | 2          | 3        | 2017 |
| 27 | BAUTISTA SAAVEDRA | LESIONES GRAVES Y LEVES | 1     | 0       | 0     | 0     | 0    | 0     | 0     | 1      | 0          | 0       | 0         | 0         | 2          | 14       | 2017 |
| 28 | SAN ANTONIO       | LESIONES GRAVES Y LEVES | 1     | 0       | 1     | 0     | 0    | 0     | 0     | 0      | 0          | 0       | 0         | 0         | 2          | 7        | 2017 |
| 29 | CRUCE VILLA ADELA | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 1    | 0     | 1     | 0      | 0          | 0       | 1         | 0         | 3          | 3        | 2017 |
| 30 | LOS ANDES         | LESIONES GRAVES Y LEVES | 0     | 0       | 0     | 0     | 1    | 1     | 1     | 0      | 0          | 0       | 0         | 0         | 3          | 6        | 2017 |

Nota. Datos de los hechos delictivos del Municipio de El Alto.

### 3.4.2.3 Descripción de los datos

La información que se tiene en las columnas de la hoja de cálculo Excel es:

- Zona, es el lugar en el que se registró algún tipo de suceso delictivo.
- Nombre del caso, es El nombre del tipo de suceso delictivo que ocurrió en el lugar específico.
- Meses, los meses están separados por su respectivo mes con los números de casos que se registraron en el mes correspondiente.
- Total, casos es el conjunto de sucesos delictivos ocurridos en toda la gestión.
- Distrito, se registraron todos los distritos correspondientes al Municipio de El Alto
- Año, es el año en el cual se registraron los sucesos delictivos.

### 3.4.3 FASE III: Preparación de los datos

Se deben cargar los datos obtenidos a un gestor de base de datos para su respectivo uso requerido en el prototipo.

**Tabla 9.**

*Datos de la Tabla*

| <b>Campo</b>    | <b>Muestra</b>          |
|-----------------|-------------------------|
| <b>Id</b>       | 1                       |
| <b>Zona</b>     | 12 de octubre           |
| <b>Nom_caso</b> | Lesiones Graves y leves |
| <b>Caso_ene</b> | 10                      |
| <b>Caso_feb</b> | 3                       |
| <b>Caso_mar</b> | 19                      |
| <b>Caso_abr</b> | 0                       |
| <b>Caso_may</b> | 16                      |
| <b>Caso_jun</b> | 13                      |
| <b>Caso_jul</b> | 13                      |
| <b>Caso_ago</b> | 6                       |
| <b>Caso_sep</b> | 6                       |

|                   |      |
|-------------------|------|
| <b>Caso_dic</b>   | 7    |
| <b>Caso_nov</b>   | 2    |
| <b>Caso_dic</b>   | 1    |
| <b>Total_caso</b> | 96   |
| <b>Distrito</b>   | 6    |
| <b>Año</b>        | 2017 |

*Nota.* Elaboración propia.

**Tabla 10.**

*Total, datos delictivos*

| <b>Id</b> | <b>Total_caso</b> | <b>Año</b> |
|-----------|-------------------|------------|
| 1         | 1797              | 2017       |
| 2         | 1553              | 2018       |
| 3         | 1393              | 2019       |
| 4         | 1713              | 2020       |
| 5         | 1737              | 2021       |
| 6         | 2188              | 2022       |
| 7         | 2855              | 2023       |

*Nota.* Total, datos de las gestiones.



### 3.4.3.1 Importando a la Base de Datos

Se necesita usar los datos desde una base de datos para poder realizar las consultas SQL y obtener los datos correspondientes que nos ayudaran a lograr la predicción.

**Figura 21.**

*Base de Datos PostgreSQL*

| id      | zona                    | nom_caso                | caso_e  | caso_feb | caso_mar | caso_abr | caso_ma | caso_jun | caso_jul | caso_ag | caso_sep | caso_oi | caso_nov | caso_dic | total_cas | distrito | año     |
|---------|-------------------------|-------------------------|---------|----------|----------|----------|---------|----------|----------|---------|----------|---------|----------|----------|-----------|----------|---------|
| integer | character varying (200) | character varying (200) | integer | integer  | integer  | integer  | integer | integer  | integer  | integer | integer  | integer | integer  | integer  | integer   | integer  | integer |
| 1       | 12 DE OCTUBRE           | LESIONES GRAVES Y LEVES | 10      | 3        | 19       | 0        | 16      | 13       | 13       | 6       | 6        | 7       | 2        | 1        | 96        | 1        | 2017    |
| 2       | 16 DE JULIO             | LESIONES GRAVES Y LEVES | 0       | 5        | 4        | 0        | 4       | 1        | 6        | 2       | 1        | 3       | 1        | 1        | 28        | 6        | 2017    |
| 3       | VILLA BOLIVAR D         | LESIONES GRAVES Y LEVES | 0       | 0        | 3        | 0        | 1       | 0        | 0        | 1       | 21       | 0       | 0        | 0        | 26        | 2        | 2017    |
| 4       | ALTO LIMA               | LESIONES GRAVES Y LEVES | 0       | 2        | 2        | 0        | 2       | 2        | 2        | 2       | 2        | 3       | 1        | 1        | 19        | 6        | 2017    |
| 5       | VILLA DOLORES           | LESIONES GRAVES Y LEVES | 0       | 0        | 1        | 0        | 4       | 0        | 3        | 2       | 0        | 5       | 0        | 1        | 16        | 3        | 2017    |
| 6       | SENKATA                 | LESIONES GRAVES Y LEVES | 1       | 0        | 2        | 0        | 1       | 3        | 3        | 0       | 0        | 0       | 0        | 0        | 10        | 8        | 2017    |
| 7       | RIO SECO                | LESIONES GRAVES Y LEVES | 0       | 0        | 0        | 0        | 6       | 2        | 2        | 3       | 2        | 0       | 1        | 0        | 16        | 4        | 2017    |
| 8       | VILLA TUNARI            | LESIONES GRAVES Y LEVES | 1       | 1        | 0        | 0        | 1       | 0        | 0        | 0       | 3        | 3       | 0        | 1        | 10        | 4        | 2017    |
| 9       | CIUDAD SATELITE         | LESIONES GRAVES Y LEVES | 0       | 0        | 3        | 0        | 1       | 1        | 5        | 0       | 0        | 2       | 0        | 2        | 14        | 1        | 2017    |
| 10      | VILLA ADELA             | LESIONES GRAVES Y LEVES | 1       | 0        | 2        | 0        | 2       | 2        | 1        | 0       | 1        | 0       | 0        | 0        | 9         | 3        | 2017    |
| 11      | FERROPETROL             | LESIONES GRAVES Y LEVES | 0       | 0        | 0        | 0        | 2       | 2        | 0        | 0       | 1        | 1       | 0        | 0        | 6         | 6        | 2017    |
| 12      | SANTA ROSA              | LESIONES GRAVES Y LEVES | 2       | 0        | 1        | 0        | 0       | 0        | 1        | 0       | 1        | 3       | 0        | 0        | 8         | 1        | 2017    |
| 13      | VILLA INGENIO           | LESIONES GRAVES Y LEVES | 1       | 0        | 2        | 0        | 0       | 0        | 1        | 1       | 0        | 0       | 0        | 0        | 5         | 5        | 2017    |
| 14      | BALLIVIAN               | LESIONES GRAVES Y LEVES | 0       | 1        | 0        | 0        | 0       | 0        | 0        | 0       | 0        | 0       | 2        | 1        | 4         | 6        | 2017    |
| 15      | SAN MARTIN              | LESIONES GRAVES Y LEVES | 0       | 1        | 0        | 0        | 1       | 1        | 1        | 0       | 0        | 0       | 0        | 1        | 5         | 7        | 2017    |
| 16      | SAN LUIS PAMPA          | LESIONES GRAVES Y LEVES | 0       | 0        | 1        | 0        | 0       | 1        | 0        | 0       | 1        | 0       | 0        | 0        | 3         | 3        | 2017    |

*Nota.* Datos ingresados en el gestor de base de Datos PostgreSQL.

### 3.4.4 FASE IV: Modelado

En la Minería de Datos mediante la información cargados a la base de datos es con la aplicación de algoritmo. Los datos cumplen con los criterios estadísticos de validez y confiabilidad.

### 3.4.5 FASE V: Evaluación

El algoritmo implementado en el modelo de predicción, el modelo Arima Luego de las varias correcciones en el prototipo se obtuvo resultados en la ejecución del modelo con el algoritmo mencionado.

**Figura 22.***Resultado de Arima*

```

=====
SARIMAX Results
=====
Dep. Variable:      SUM  No. Observations:      7
Model:             ARIMA(1, 0, 1)  Log Likelihood      -50.450
Date:             Mon, 17 Jun 2024  AIC              108.900
Time:             11:37:38  BIC              108.683
Sample:           0  HQIC              106.226
                  - 7
Covariance Type:   opg
=====
              coef  std err      z  P>|z|  [0.025  0.975]
-----
const      2005.9070  534.193   3.755  0.000  958.909  3052.905
ar.L1       0.4286   0.422   1.017  0.309  -0.398   1.255
ma.L1       1.0000   1.300   0.769  0.442  -1.548   3.548
sigma2      7.023e+04   0.000  4.55e+08  0.000  7.02e+04  7.02e+04
=====
Ljung-Box (L1) (Q):      0.19  Jarque-Bera (JB):      0.87
Prob(Q):                 0.67  Prob(JB):              0.65
Heteroskedasticity (H):  4.79  Skew:                  0.56
Prob(H) (two-sided):    0.35  Kurtosis:              1.68
=====

```

*Nota:* datos resultantes del entrenamiento Arima.

### 3.5 APLICANDO LA METODOLOGIA SCRUM

#### 3.5.1 Planificación de la iteración (*Sprint Planning*)

**Objetivo:** Diseñar un modelo predictivo de índices delictivos basado en Minería de Datos, aplicado al Municipio de El Alto, utilizando datos recolectados de las diferentes instituciones relacionadas con la temática de Seguridad, con el fin de anticipar el incremento de los índices delictivos.

**Duración del sprint:** 2 semanas

**Tabla 11.**

*Sprint Planning.*

| Equipo                     | Tareas                  |                                        |
|----------------------------|-------------------------|----------------------------------------|
| Científico de Datos.       | Revisión de requisitos. | Recopilación de Datos.                 |
| Desarrollador de Software. | Selección del modelo.   | Entrenamiento y evaluación del modelo. |
| Analista de Datos.         | Exploración de Datos.   | Preparación de Datos.                  |

*Nota.* Planificación de la iteración, elaboración propia.

### 3.5.2 Ejecución de la iteración (Sprint)

| Equipo                     | Para hacer                                       | Haciendo                                          | Hecho                                                |
|----------------------------|--------------------------------------------------|---------------------------------------------------|------------------------------------------------------|
| Científico de Datos.       | Una revisión detallada de los requisitos.        | Análisis de datos históricos, datos relevantes.   | Selección de datos.                                  |
| Desarrollador de Software. | Selección del algoritmo adecuado.                | Entrenar al modelo, mediante los datos preparados | Modelo seleccionado mediante el ajuste de los datos. |
| Analista de Datos.         | Identificar patrones, relaciones en los índices. | Limpieza de los datos.                            | Incorporar los datos transformados al modelo.        |

*Nota.* Ejecución de la Iteración, Elaboración Propia.

### 3.5.3 Reunión diaria de sincronización del equipo (Daily meeting)

| Equipo                     | Tareas                    | Obstáculos                                          | Objetivos                          |
|----------------------------|---------------------------|-----------------------------------------------------|------------------------------------|
| Científico de Datos.       | Revisión de requisitos.   | Datos atípicos.                                     | Datos históricos seleccionados.    |
| Desarrollador de software. | Selección del modelo.     | Identificación del modelo adecuado al tipo de dato. | Modelo preparado para las pruebas. |
| Analista de Datos.         | Exploración de los Datos. | Datos duplicados.                                   | Datos preparados para el modelo.   |

*Nota.* Sincronización del equipo y seguimiento, Elaboración propia.

### 3.5.4 *Demostración de requisitos completados (Sprint Review)*

- **Recolección de datos:**
  - **Datos Históricos:** datos de los hechos delictivos del Municipio de El Alto.
  - **Fuentes de recopilación de datos:** comando General de la Policía Boliviana.
  - **Formato adecuado:** información obtenida en formato hoja de cálculo Excel.
- **Procesamiento de Datos:**
  - **Limpieza y normalización de datos:** se realizó la limpieza correspondiente.
  - **eliminación de valores atípico:** no se detectó valores atípicos considerables.
- **Análisis Exploratorio de Datos:**
  - **Generación de gráficos:** se generaron los gráficos con los datos procesados.
- **Selección de Algoritmos:**
  - Algoritmos: se seleccionó los algoritmos adecuados para una mejor predicción con los datos procesados.
- **Entrenamiento del modelo:** se entreno al modelo con el algoritmo adecuado y con el procesamiento de los datos.
- **Resultados del Modelo:** el modelo seleccionado dio resultados de los datos históricos que se seleccionaron para su proyección, y como resultado fue capaz de predecir datos futuros.

## 3.6 METRICAS DE CALIDA

Para asegurar un producto de alta calidad, la norma ISO/IEC 25010 de calidad de software ofrece una clasificación detallada en un marco estructurado.

### 3.6.1 *Funcionalidad*

Son las necesidades que satisfacen, a continuación, una muestra en la siguiente tabla.

**Tabla 12.***Puntuación de la Funcionalidad*

| Aspectos    | Porcentaje |
|-------------|------------|
| Usabilidad  | 80%        |
| Exactitud   | 80%        |
| Conformidad | 80%        |
| Adecuación  | 80%        |
| Promedio    | 80%        |

**3.6.2 Usabilidad**

Se evaluará el esfuerzo necesario que se invirtió para usar el sistema, en la siguiente tabla se mencionara los aspectos a tomar en cuenta.

**Tabla 13.***Cálculo de la Usabilidad*

| N ° | Preguntas                          | SI | No | Puntuación |
|-----|------------------------------------|----|----|------------|
| 1   | ¿Se adaptó a usar el sistema?      | 9  | 1  | 90%        |
| 2   | ¿La interfaz le pareció agradable? | 8  | 2  | 80%        |

|           |                                                                |    |   |      |
|-----------|----------------------------------------------------------------|----|---|------|
| <b>3</b>  | ¿Le pareció comprensible la interfaz?                          | 9  | 1 | 90%  |
| <b>4</b>  | ¿El Sistema responde con rapidez a sus peticiones?             | 7  | 3 | 70%  |
| <b>5</b>  | ¿El Sistema responde a sus peticiones?                         | 9  | 1 | 90%  |
| <b>6</b>  | ¿El Sistema le facilita su trabajo?                            | 9  | 1 | 90%  |
| <b>7</b>  | ¿Es fácil el desplazamiento entre las opciones de la interfaz? | 9  | 1 | 90%  |
| <b>8</b>  | ¿Las operaciones que se realizan no son complejas?             | 10 | 0 | 100% |
| <b>9</b>  | ¿El Sistema le proporciona los resultados esperados?           | 9  | 1 | 90%  |
| <b>10</b> | ¿El Sistema no presenta errores recuentes?                     | 8  | 2 | 80%  |

Resultados de la Usabilidad es del: 84%

### 3.6.3 Confiabilidad

La confiabilidad del sistema se define mediante la probabilidad de operación libre de fallos en un entorno determinado y durante un tiempo específico. Para determinar la

confiabilidad de un software especificamos desde el instante que empieza a funcionar, es decir, que a partir de ese momento se realiza las observaciones pertinentes.

***N***=Número de fallas (indica el número de fallas que puedan llegar a ocurrir)

***F(t)***= Tiempo total de operaciones (Tiempo en el cual el software está en ejecución), en el tiempo total se tomará en cuenta 15 días de ejecución.

**( 1) Porcentaje de fallas**

$$\lambda = \frac{N}{F(t)}$$

$$\lambda = \frac{2}{15} * 100$$

$$\lambda = 13.3\%$$

La tasa de fallas representa la cantidad promedio de fallas que ocurren por unidad de tiempo. Un valor de  $\lambda$  alto indica una confiabilidad baja, mientras que un valor bajo indica una confiabilidad alta.

El 13.3% es el porcentaje de fallas que puede llegar a tener el software en el transcurso de 15 días.

Siguiendo con la fórmula para obtener el tiempo promedio entre fallas.



**( 2 ) Confiabilidad del sistema**

$$MTBF = \frac{15}{2}$$

$$MTBF = \frac{15}{2} * 100$$

$$MTBF = 75\%$$

MTBF alto indica una confiabilidad alta, ya que las fallas llegan a ocurrir con menos frecuencia, Entonces, la confiabilidad del sistema es del 75%.

**3.6.4 Eficiencia.**

Se obtendrá la eficiencia del sistema con las características esenciales que el sistema desempeña para poderlas ponderar.

**Tabla 14.***Evaluación de la Eficiencia*

| <b>Desempeño del sistema</b> | <b>Ponderación</b> |
|------------------------------|--------------------|
| Disponibilidad               | 5                  |
| Fluidez                      | 4                  |
| Rapidez de inicio            | 4                  |
| Rapidez de procesos          | 4                  |
| Proceso de información       | 3                  |

*Nota.* Evaluación de la eficiencia.

En base a los datos obtenidos del desempeño del sistema, se puede llegar con exactitud a comprender la idea de la eficiencia, siguiendo esta fórmula se obtendrá el valor a la eficiencia.

**( 3 ) Valor de la eficiencia**

$$Eficiencia = \frac{\sum xi}{n} * 100/n$$

$$Eficiencia = \frac{20}{5} * 100/5$$

$$Eficiencia = 80\%$$

**3.6.5 Mantenibilidad**

Es la facilidad con la que un programa de software se puede modificar para corregir errores, mejorar su funcionalidad o adaptarlo a nuevos requisitos. Es una característica esencial del software de alta calidad, ya que permite a los desarrolladores realizar cambios de manera eficiente y sin introducir nuevos errores.

$$Mantenibilidad = (Mt - (Fc + Fa + Fd))/Mt$$

Donde:

- t = número de módulos en la versión actual.
- Fc = número de módulos en la versión actual que han cambiado.
- Fa = número de módulos en la versión actual añadido.
- Fd = número de módulos en la versión anterior que se ha borrado

Entonces:

$$Mt = 1 ; Fc = 0 ; Fa = 0 ; Fd = 0$$

**( 4 ) Valor de la mantenibilidad**

$$\text{Mantenibilidad} = (1 - (0 + 0 + 0))/1$$

$$\text{Mantenibilidad} = 1 * 100$$

$$\text{Mantenibilidad} = 100\%$$

**3.6.6 Portabilidad**

Es la facilidad con la que un software se pueda adaptar para ejecutarse en diferentes plataformas de hardware, sistemas operativos o entornos de programación.

$$\text{Portabilidad} = 1 - (\text{ndpm}/\text{ndim})$$

Donde:

**( 5 ) Valor de la portabilidad**

$$\text{ndim} = \text{dias para convertir el modelo a portable}$$

$$\text{dias, ndim} = \text{dias para la implementacion del modelo}$$

$$\text{portabilidad} = 1 - (1/5)$$

$$\text{portabilidad} = 0.8 * 100$$

$$\text{portabilidad} = 80\%$$

**3.7 EVALUACION DE COSTOS**

Se estimará el costo de producción del Software desarrollado mediante COCOMO II.

**3.7.1 Adecuación Funcional**

Para esta parte se agrupará los atributos que calificaran si el producto de Software maneja las funciones para satisfacer las necesidades para las que fue diseñado. Para este cálculo funcional se determinó las características de dominios de información de la siguiente manera.

- **Número de entradas de usuario:** es cuando el usuario proporciona distintos datos en el sistema.
- **Número de salida de usuario:** son datos que ofrecen información al usuario.
- **Número de peticiones:** se refieren a las consultas desde la base de datos que muestra información en el sistema.
- **Número de archivos:** se refieren a la base de datos o flujos de información requerida por el sistema.
- **Número de interfaces externos:** representa las interfaces externas con las que se conecta nuestro sistema.

**Tabla 15.**

*Cálculo de Adecuación Funcional*

| N ° | PARAMETROS DE MEDIDA          | CANTIDAD |
|-----|-------------------------------|----------|
| 1   | Número de entradas de usuario | 0        |
| 2   | Número de salida de usuario   | 26       |
| 3   | Número de peticiones          | 26       |
| 4   | Número de archivos            | 1        |
| 5   | Número de interfaces externos | 0        |

*Nota.* Cuadro de resultados obtenidos del cálculo de la adecuación funcional.

**Tabla 16.***Cálculo de cuenta total*

| <b>Parámetros de medición</b> | <b>Cuenta total</b> | <b>Factor de ponderación</b> | <b>Valor obtenido</b> |
|-------------------------------|---------------------|------------------------------|-----------------------|
| Número de entradas de usuario | 0                   | 4                            | 0                     |
| Número de salida de usuario   | 7                   | 5                            | 35                    |
| Número de peticiones          | 8                   | 5                            | 40                    |
| Número de archivos            | 1                   | 4                            | 4                     |
| Número de interfaces externos | 0                   | 4                            | 0                     |
| <b>TOTAL</b>                  |                     |                              | <b>79</b>             |

*Nota.* Cálculo general de la cuenta total.

**Tabla 17.***Cálculo factor de ajuste de complejidad*

| N ° | Factores                                                                              | 0 | 1 | 2 | 3 | 4 | 5 | Fi |
|-----|---------------------------------------------------------------------------------------|---|---|---|---|---|---|----|
| 1   | ¿Requiere el sistema copias de seguridad y de recuperación fiables?                   |   |   |   |   | X |   | 4  |
| 2   | ¿Se requiere comunicación de datos?                                                   |   |   |   |   |   | X | 5  |
| 3   | ¿Existen funciones de procesos distribuidos?                                          |   | X |   |   |   |   | 1  |
| 4   | ¿Es crítico el rendimiento?                                                           |   |   |   |   | X |   | 4  |
| 5   | ¿Será ejecutado el sistema en un entorno operativo existente y fuertemente utilizado? |   |   |   |   | X |   | 4  |
| 6   | ¿Requiere el sistema entrada de datos interactivo?                                    |   | X |   |   |   |   | 1  |
| 7   | ¿Se utilizaron los archivos maestros de forma iterativa?                              |   | X |   |   |   |   | 1  |
| 8   | ¿Tiene facilidad operativa?                                                           |   |   | X |   |   |   | 2  |
| 9   | ¿Son complejas las entradas, salidas y/o peticiones?                                  |   | X |   |   |   |   | 1  |
| 10  | ¿Es complejo el procesamiento interno?                                                |   |   | X |   |   |   | 2  |

---

|                                     |                                                                                                          |   |           |
|-------------------------------------|----------------------------------------------------------------------------------------------------------|---|-----------|
| 11                                  | ¿Se ha diseñado el código para ser reutilizable?                                                         | X | 3         |
| 12                                  | ¿Están incluidas en el diseño la conversación y la instalación?                                          | X | 2         |
| 13                                  | ¿Se ha diseñado el sistema para soportar diferentes instalaciones en diferentes organizaciones?          | X | 2         |
| 14                                  | ¿Se ha diseñado la aplicación para facilitar los cambios y para ser fácilmente utilizada por el usuario? | X | 4         |
| <b>Factor ajusté de complejidad</b> |                                                                                                          |   | <b>36</b> |

---

*Nota.* Cálculo general del factor de ajuste de complejidad.

Para realizar el cálculo de punto de función se emplea la siguiente fórmula:

$$PF = CuentaTotal * (x + \min(Y) * \sum f_i)$$

Donde:

**PF:** medida de la ecuación funcional

**Cuenta Total:** es la sumatoria de los números de entrada, números de salida, número de peticiones, número de archivos y de número de interfaces externas.

**X:** Confiabilidad del proyecto, varía entre el 1% o 100%

**Min(Y):** error mínimo aceptable de complejidad

$\sum F_i$ : es el valor de ajuste de complejidad donde  $1 \leq i \leq 14$ .

Reemplazando los valores obtenidos anteriormente se tendrá el siguiente resultado:

**( 6 ) Punto de función**

$$PF = Cuenta\ Total * [0.75(0.013 * \sum F_i)]$$

$$PF = 79 * (0.65 + 0.013 * 36)$$

$$PF = 79 * 1.118 = 88.32$$

Como resultado se obtuvo la funcionalidad de la aplicación con el valor de 88% de funcionalidad, el cual significa que el sistema responde de manera óptima a las funcionalidades requeridas.

**3.7.2 Aplicación de COCOMO II**

El desarrollo del sistema implica una cuidadosa consideración de varias plataformas de soporte, el lenguaje de programación que se empleará y el sistema de gestión de bases de datos que respaldará la aplicación. Para estimar las líneas de código requeridas, utilizamos el valor del punto de función ajustado y el factor de línea de código asociado con el lenguaje de programación elegido.



**Tabla 18.***Factor LCD/PF de lenguaje de Programación*

| Lenguaje     | Nivel | Factor LCD/PF |
|--------------|-------|---------------|
| C            | 2.5   | 128           |
| ANSI/BASIC   | 5     | 64            |
| Java         | 6     | 53            |
| PL/I         | 4     | 80            |
| Visual Basic | 7     | 46            |
| ASP          | 9     | 36            |
| PHP          | 11    | 29            |
| Visual C++   | 9.5   | 34            |

*Nota.* Cuadro de factor LCD/PF.

Cálculo de costo del Modelo mediante las siguientes fórmulas.

$$E = a(KLDC)^b ; \text{Personas} = \text{mes}$$

$$Tdev = c(E)^d ; \text{Meses}$$

$$P = E/dev ; \text{Personas}$$

Donde:

**E:** esfuerzo requerido por el proyecto expresado en persona-mes.

**D:** tiempo requerido por el proyecto expresado en meses.

**P:** número de personas requeridas para el proyecto.

**A, B, C y D:** constantes con valores definidos según cada sub modelo.

**KLDC:** cantidad de líneas de código distribuidas en miles.

A la vez cada modelo se subdivide en tres modos:

**Modo orgánico:** es un pequeño grupo de programadores experimentados desarrollando proyectos de software en un entorno familiar.

**Modo semilibre:** corresponde a un esquema intermedio entre el modo orgánico y el rígido, el grupo de desarrollo puede incluir una mezcla de personas experimentadas y no experimentadas.

**Modo rígido:** el proyecto tiene fuertes restricciones, que pueden estar relacionadas con la funcionalidad y/o pueden ser técnicas.

**Tabla 19.**

*Constantes A, B, C, D COCOMO*

| <b>Modo</b>      | <b>A</b> | <b>B</b> | <b>C</b> | <b>D</b> |
|------------------|----------|----------|----------|----------|
| <b>Orgánico</b>  | 2.4      | 1.05     | 2.5      | 0.38     |
| <b>Semilibre</b> | 3.0      | 1.12     | 2.5      | 0.35     |
| <b>Rígido</b>    | 3.6      | 2.20     | 2.5      | 0.32     |

*Nota.* Constantes de los Modos.

### **3.7.2.1 Costos del sistema**

Para la estimación de costos del sistema se realizará el desarrollo tomando en cuenta las KLDC (Kilo líneas de Código) en la cual detallaremos:

**( 7 ) Kilo líneas de código**

$$LDC = PFA * Factor LDC / PF$$

$$LDC = 88.32 * 35$$

$$LDC = 3091.2$$

$$KLDC = 3091.2/1000$$

$$KLDC = 3.09$$

Esfuerzo Requerido para el proyecto persona – mes

**( 8 ) Esfuerzo requerido**

$$E = a(KLDC)^b ; Person\text{-}mes$$

$$E = 2.4 * (3.09)^{1.05} = 7 \text{ persona} - \text{mes}$$

Para el cálculo del tiempo requerido para el proyecto en meses.

**( 9 ) Tiempo requerido**

$$Tdev = c(E)^d ; Meses$$

$$Tdev = 2.5 * (7)^{0.38} = 5 \text{ meses}$$

Para el cálculo del número de programadores para desarrollar el proyecto.

**( 10 ) Cálculo de programadores**

$$P = E/Tdev ; Person\text{-}as$$

$$P = 8/5 = 1.6 = 1 \text{ personas}$$

Cálculo de estimación de salario, el salario mínimo nacional es de Bs. 2500, esta cifra se tomará en cuenta para la estimación de salario.

**( 11 ) Estimación de salario**

*Costo del software = N° de programadores \* salario de un programador*

*Costo del Software por Persona = 1 \* 2500 = 2500 Bs.*

*Costo Total del Software Desarrollado = 2500 \* 5 = 12500 Bs.*

**Tabla 20.**

*Costo total del Modelo*

| Descripción                                                | Estimación       |
|------------------------------------------------------------|------------------|
| Personas necesarias por mes para llevar a cabo el proyecto | 7 personas – mes |
| Tiempo de desarrollo del proyecto                          | 5 meses          |
| Personas necesarias para realizar el proyecto              | 1 persona        |
| <b>Costo total del proyecto</b>                            | <b>Bs.12500</b>  |

*Nota.* Costo total del modelo.

El costo del sistema desarrollado con un cálculo de tiempo requerido de 5 meses con un costo total de software de Bs. 12500, el costo total del software equivalente en dólares es de \$ 1795.



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# **CAPÍTULO IV**

## PRUEBAS Y RESULTADOS



## 4 CAPÍTULO IV

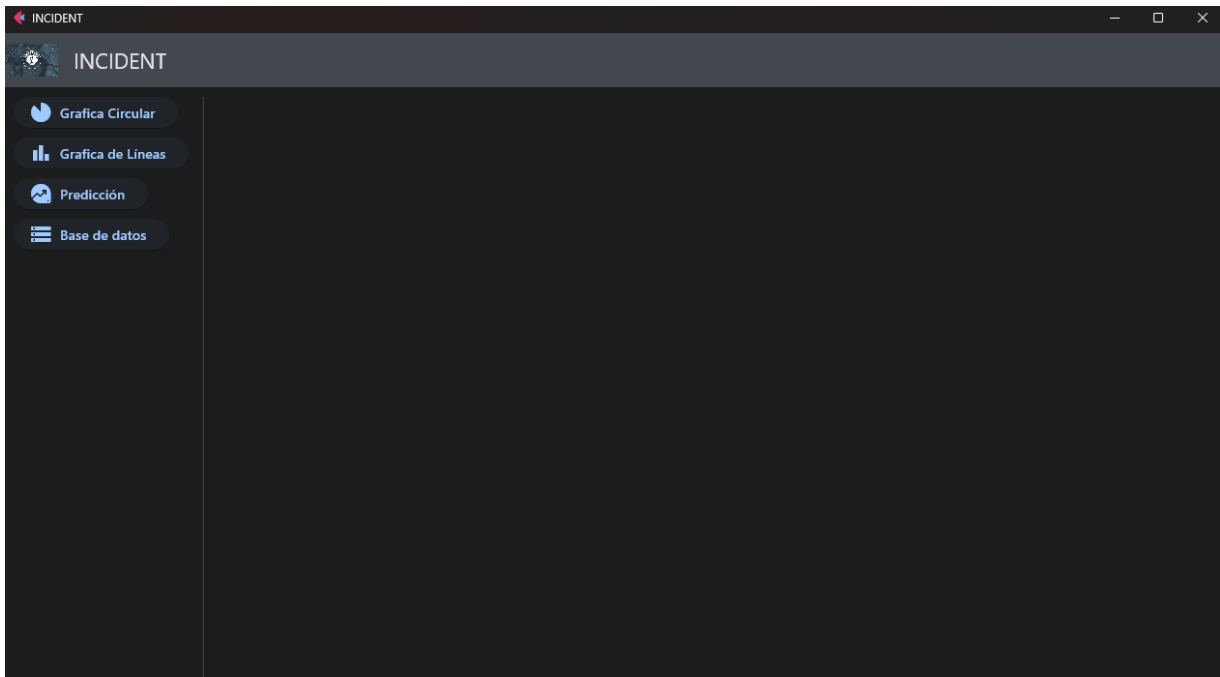
### 4.1 PRUEBAS Y RESULTADOS

Tiene como principal objetivo dar a conocer las pruebas de Hipótesis del presente trabajo de investigación y su interpretación, que es fundamental para el trabajo de investigación.

### 4.2 PRESENTACIÓN DEL MODELO

#### Figura 23.

*Interfaz principal*



*Nota.* Interfaz principal del prototipo diseñado para la ejecución de la predicción.

### 4.3 DESARROLLO DEL MODELO

Se importaron las librerías necesarias para el diseño del prototipo que ejecutara el análisis de datos del modelo, como pandas, statsmodels, matplotlib, flet, psycopg2.

Se cargó la serie para analizar y se preparó el modelo, esto incluyó la limpieza de datos y modificaciones para adecuar el modelo a la predicción si fuese necesario.

Se seleccionó los valores apropiados para la predicción más óptima.

Se entrenó al modelo ARIMA utilizando la serie temporal preparada.

Se debe realizar el pronóstico futuro para la serie temporal.

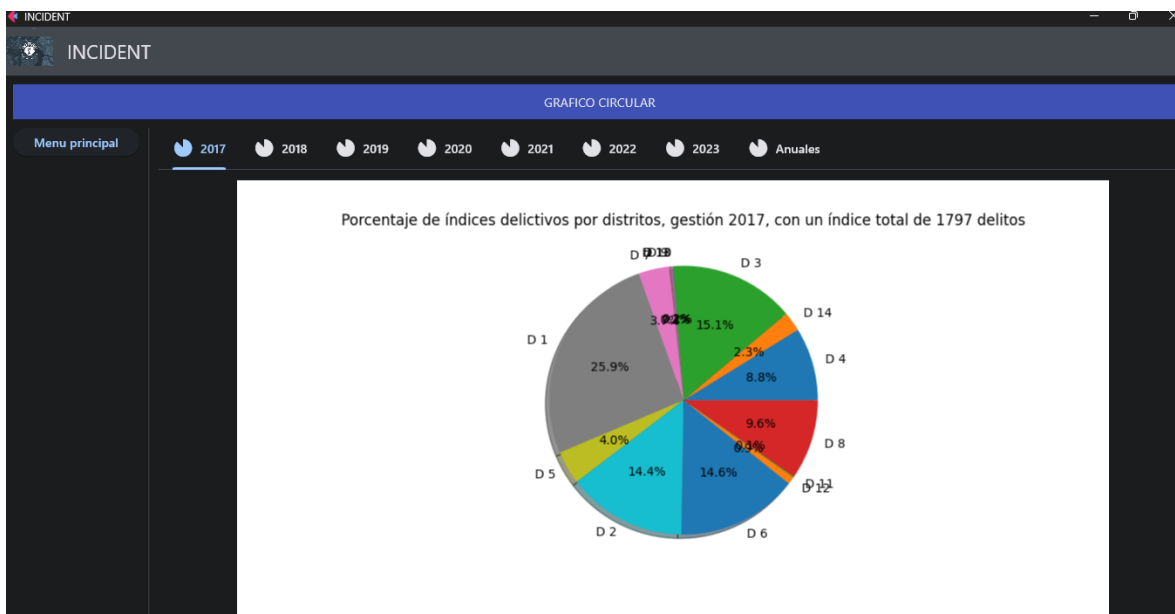
Se visualizó los resultados del modelo ARIMA, incluyendo la serie original y los pronósticos.

El desarrollo del modelo predictivo en ARIMA utilizando Python me permitió analizar una serie temporal, y realizar pronósticos futuros con precisión.

#### 4.4 DEMOSTRACIÓN DEL PROTOTIPO

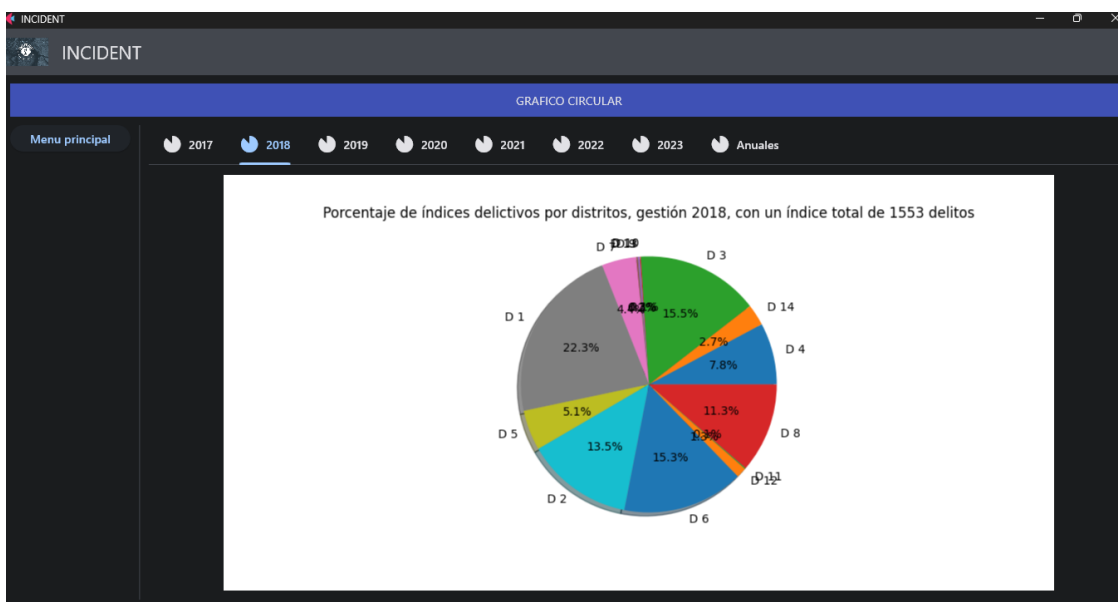
##### Figura 24.

##### *Porcentaje de índices delictivos Gestión 2017*



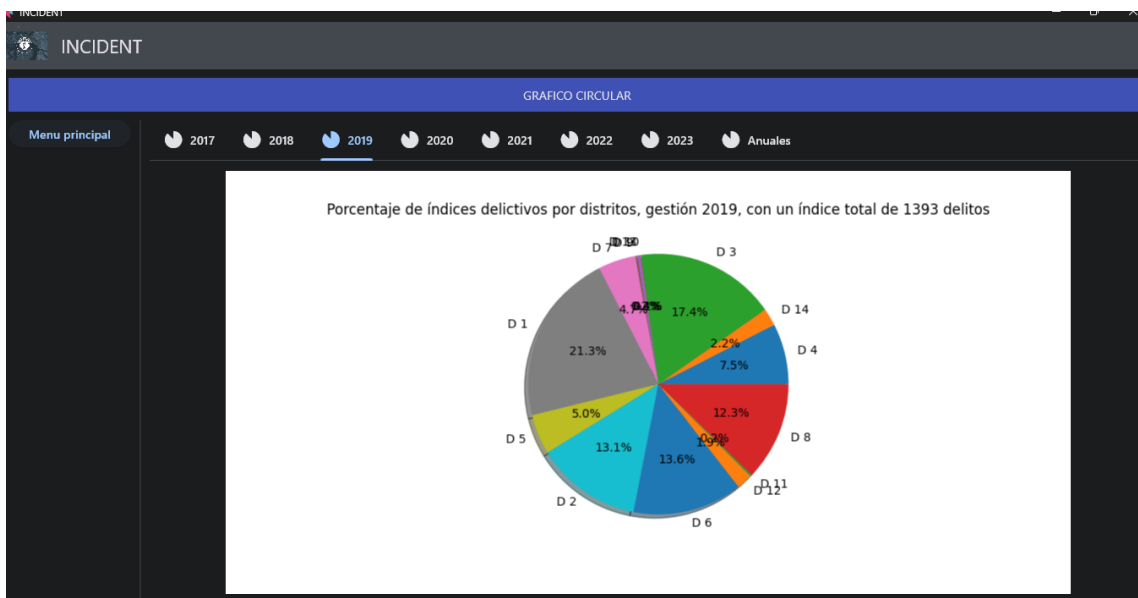
**Figura 25.**

*Porcentaje de índices delictivos Gestión 2018*



**Figura 26.**

*Porcentaje de índices delictivos Gestión 2019*





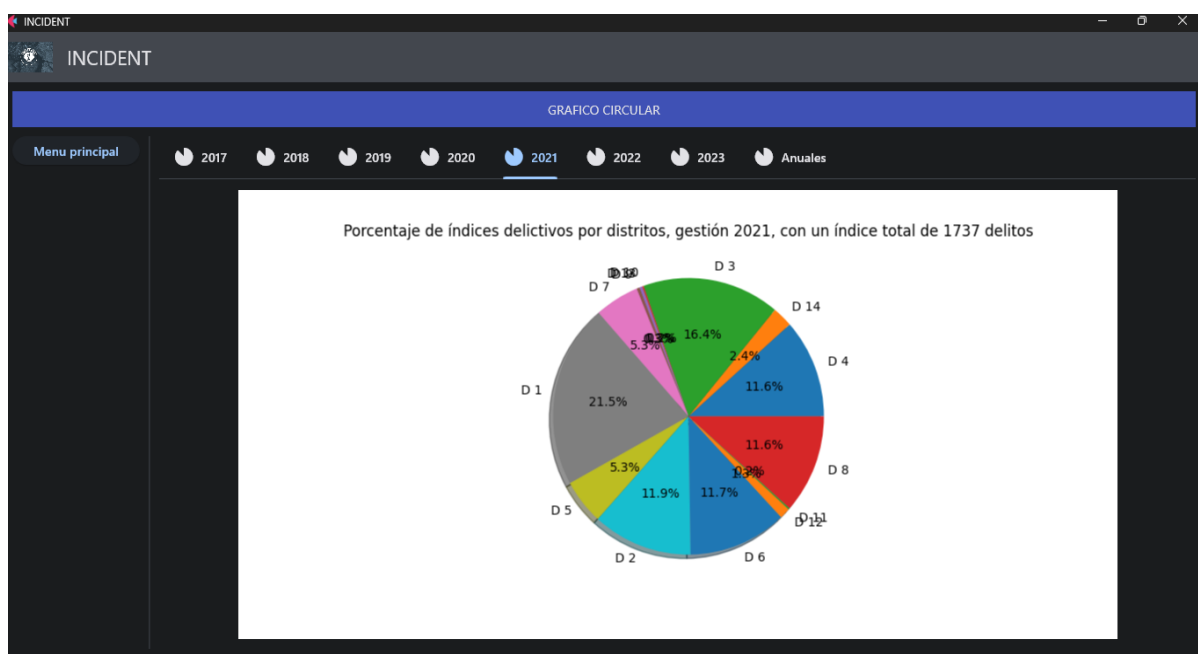
**Figura 27.**

*Porcentaje de índices delictivos Gestión 2020*



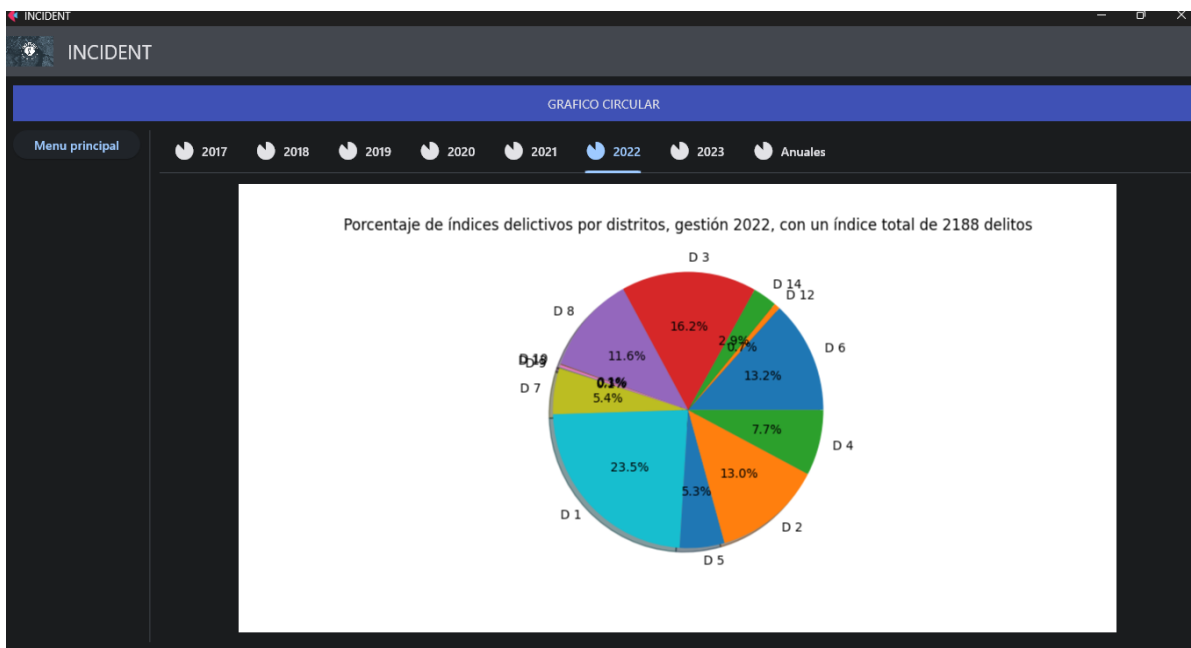
**Figura 28.**

*Porcentaje de índices delictivos Gestión 2021*



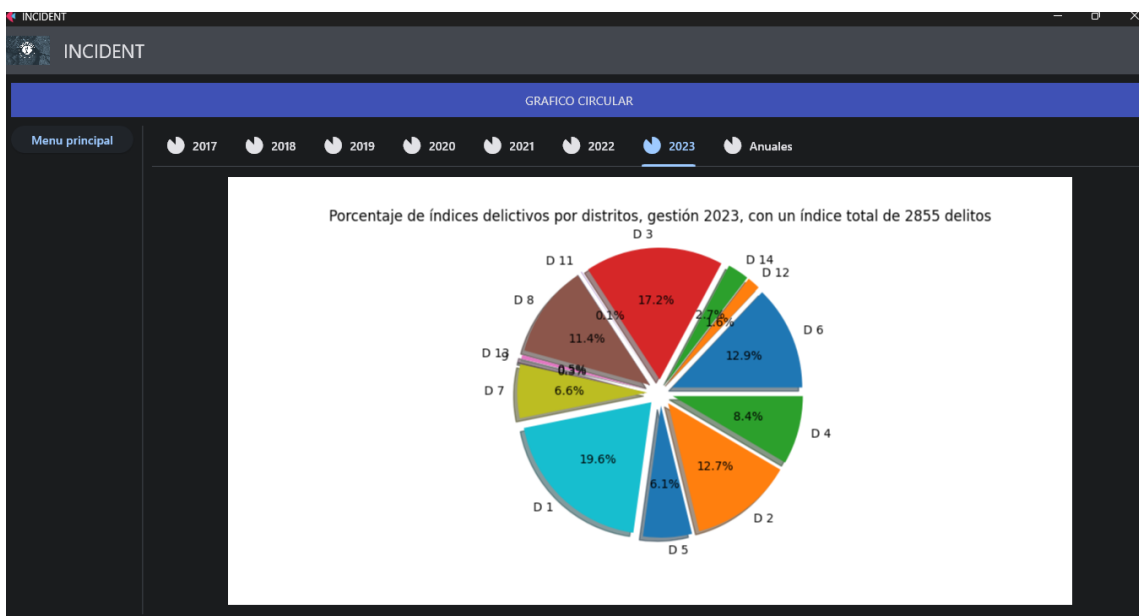
**Figura 29.**

*Porcentaje de índices delictivos Gestión 2022*



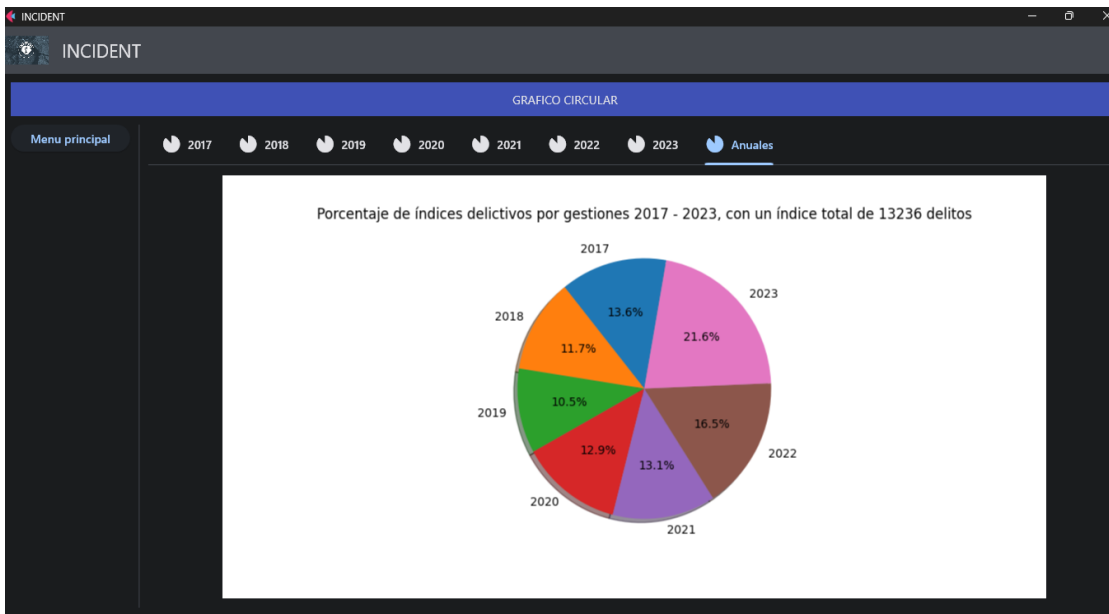
**Figura 30.**

*Porcentaje de índices delictivos Gestión 2023*



**Figura 31.**

*Porcentaje de índices delictivos Gestiones 2017 al 2023*



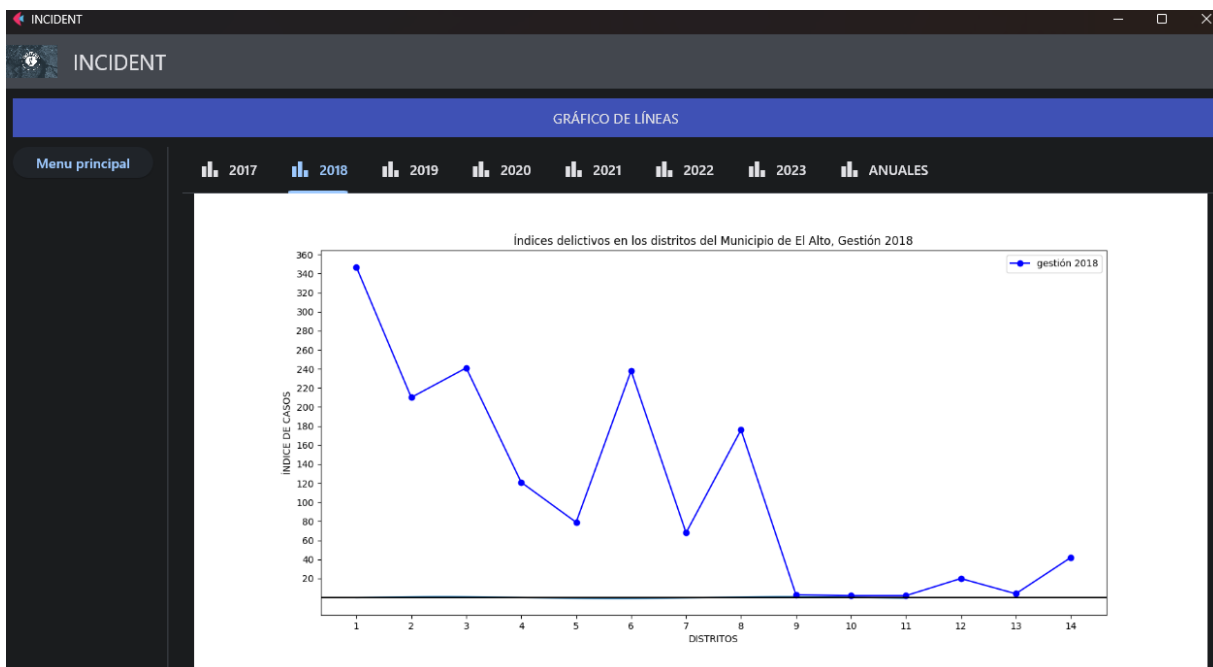
**Figura 32.**

*Índices delictivos altos y bajos por distritos gestión 2017*



**Figura 33.**

*Índices delictivos altos y bajos por distritos gestión 2018*



**Figura 34.**

*Índices delictivos altos y bajos por distritos gestión 2019*

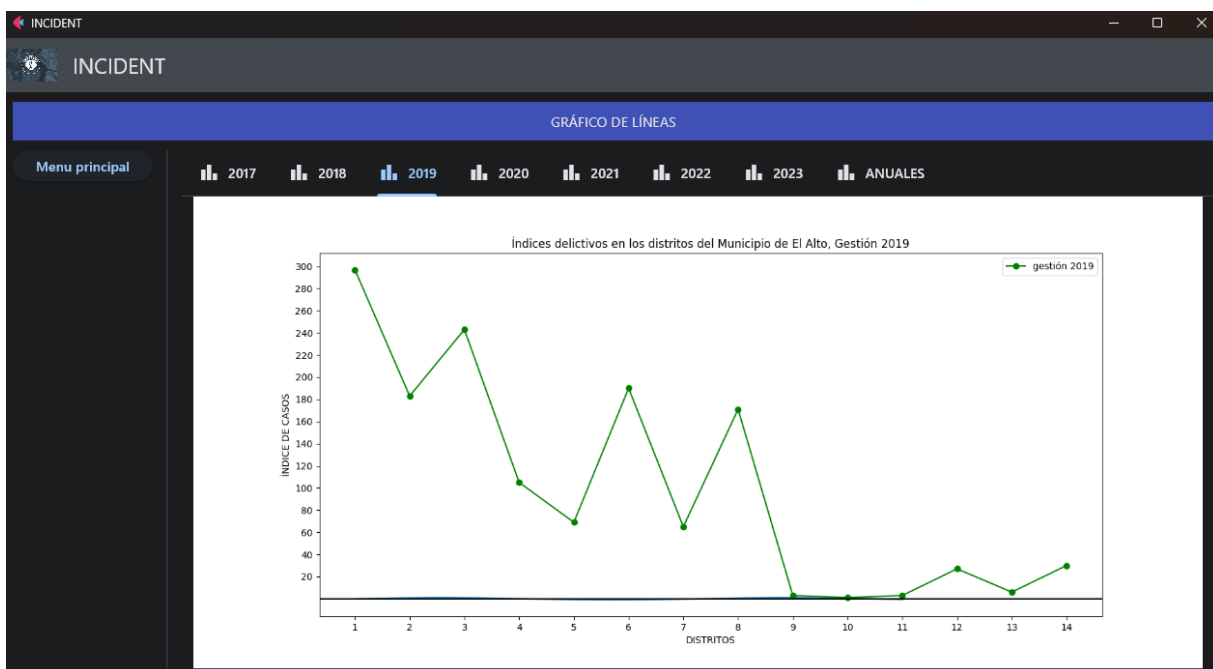


Figura 35.

*Índices delictivos altos y bajos por distritos gestión 2020*

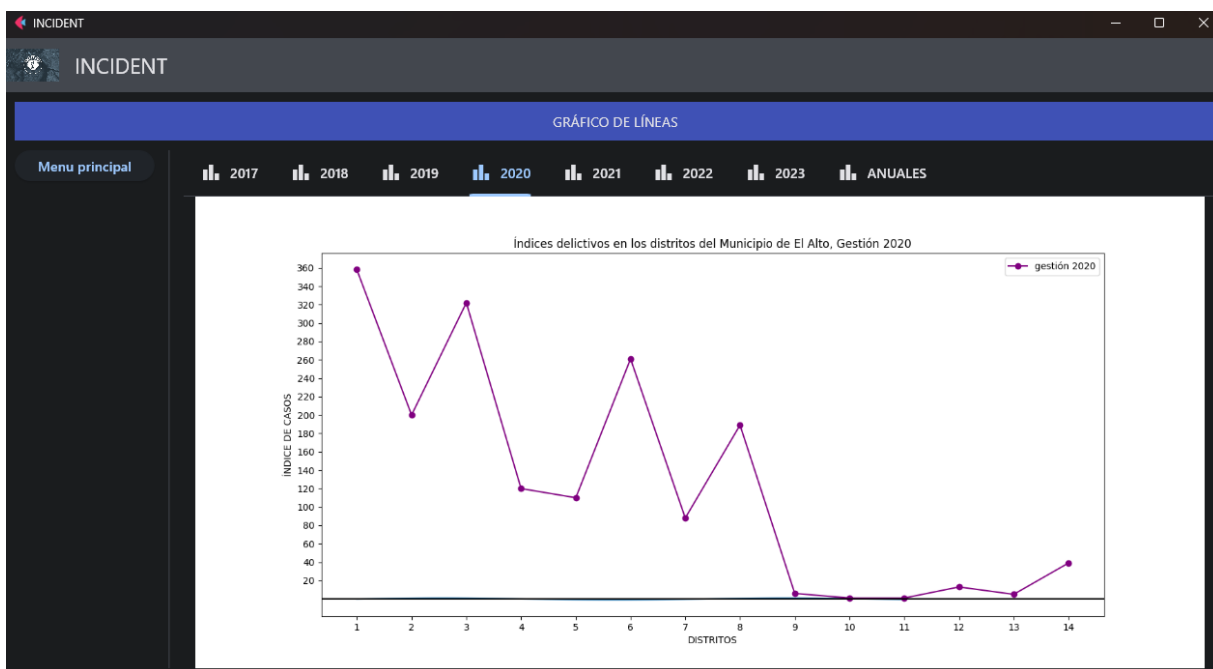


Figura 36.

*Índices delictivos altos y bajos por distritos gestión 2021*

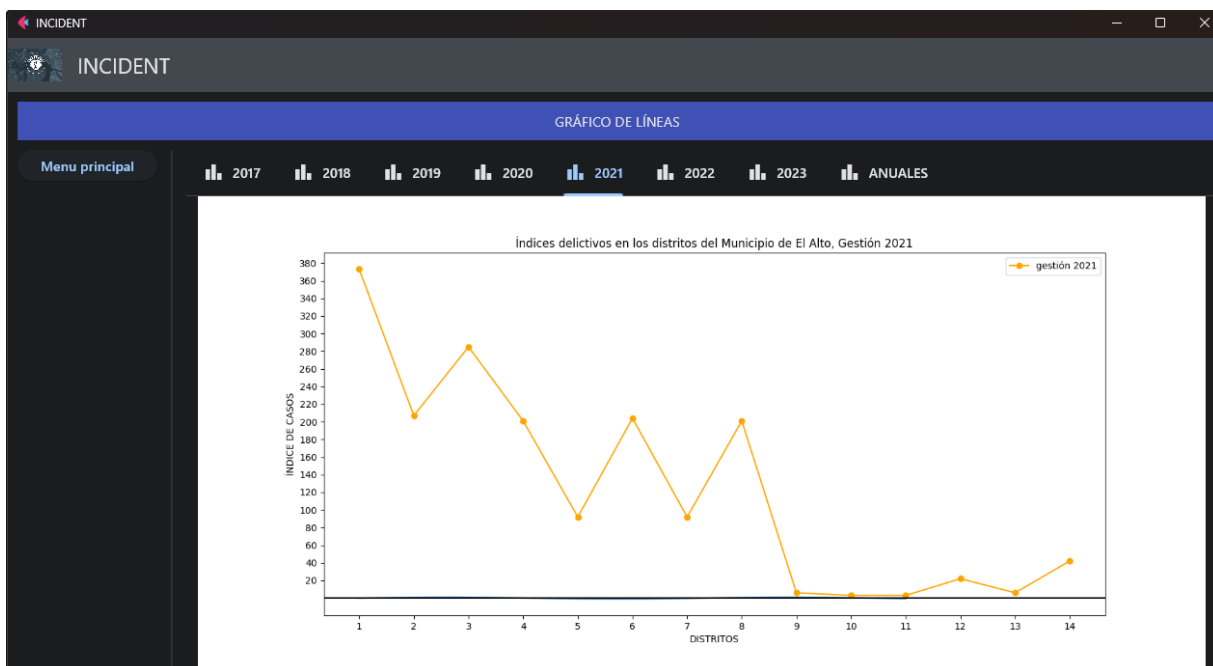


Figura 37.

*Índices delictivos altos y bajos por distritos gestión 2022*



Figura 38.

*Índices delictivos altos y bajos por distritos gestión 2023*



Figura 39.

Índices delictivos altos y bajos por distritos gestiones 2017 al 2023



Figura 40.

Información del Modelo Arima

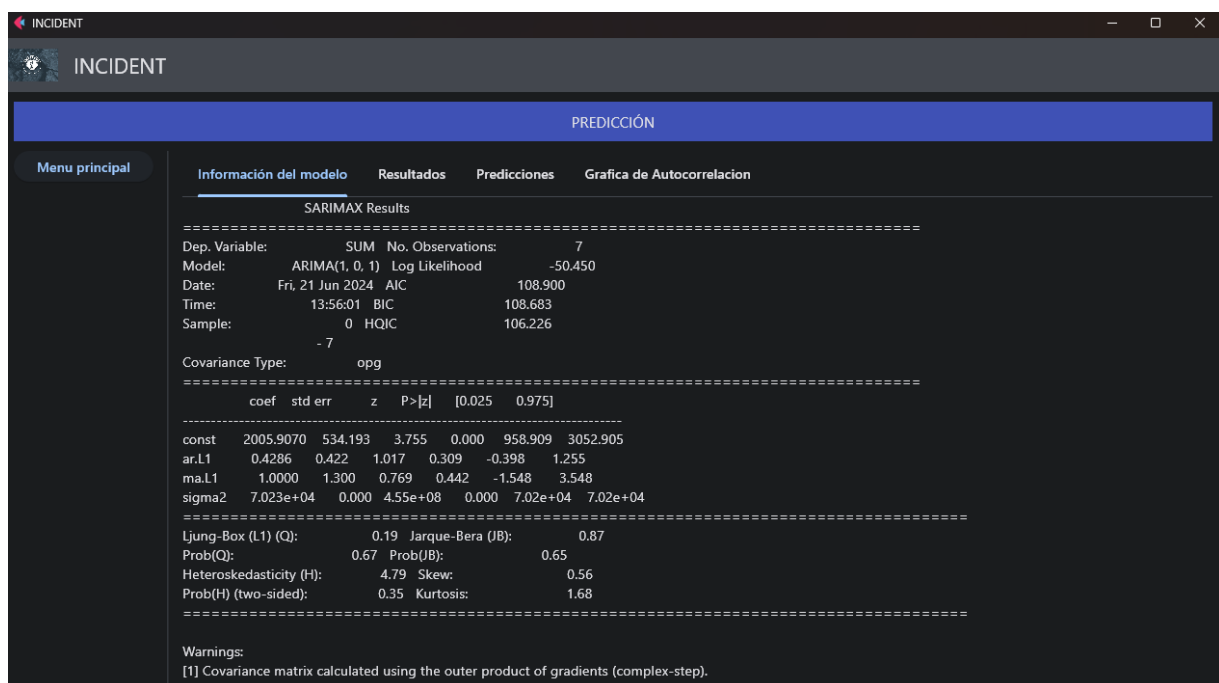


Figura 41.

Resultados obtenidos

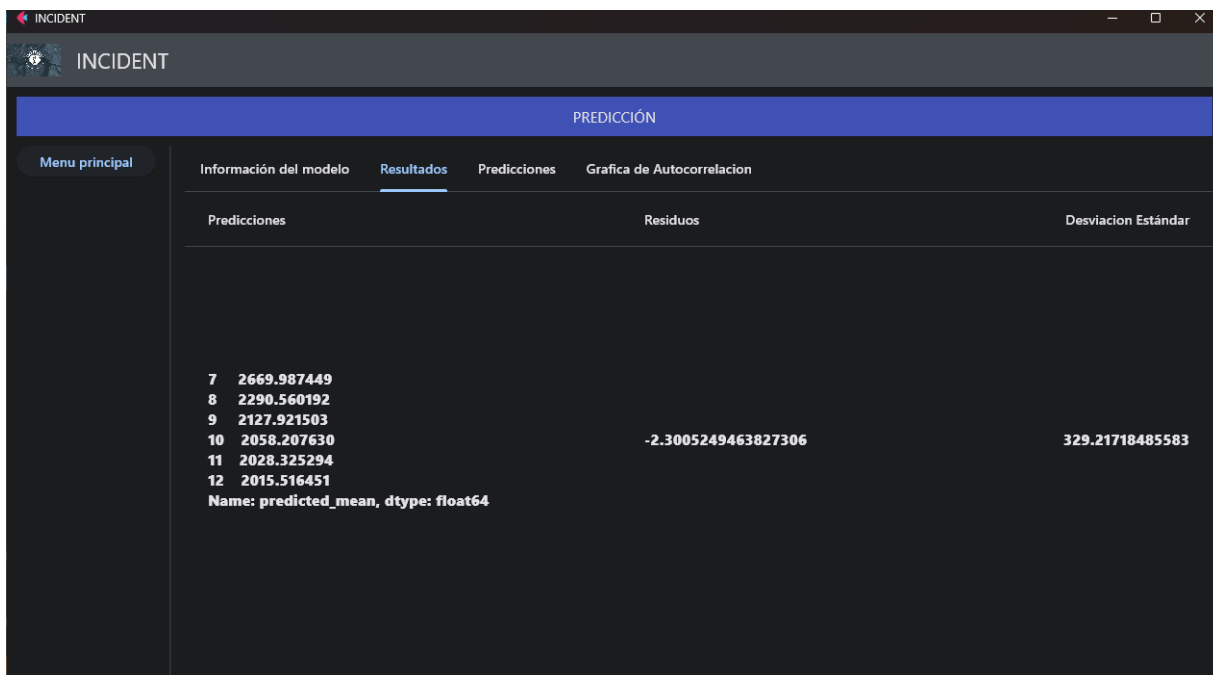


Figura 42.

Predicción de 6 años futuros

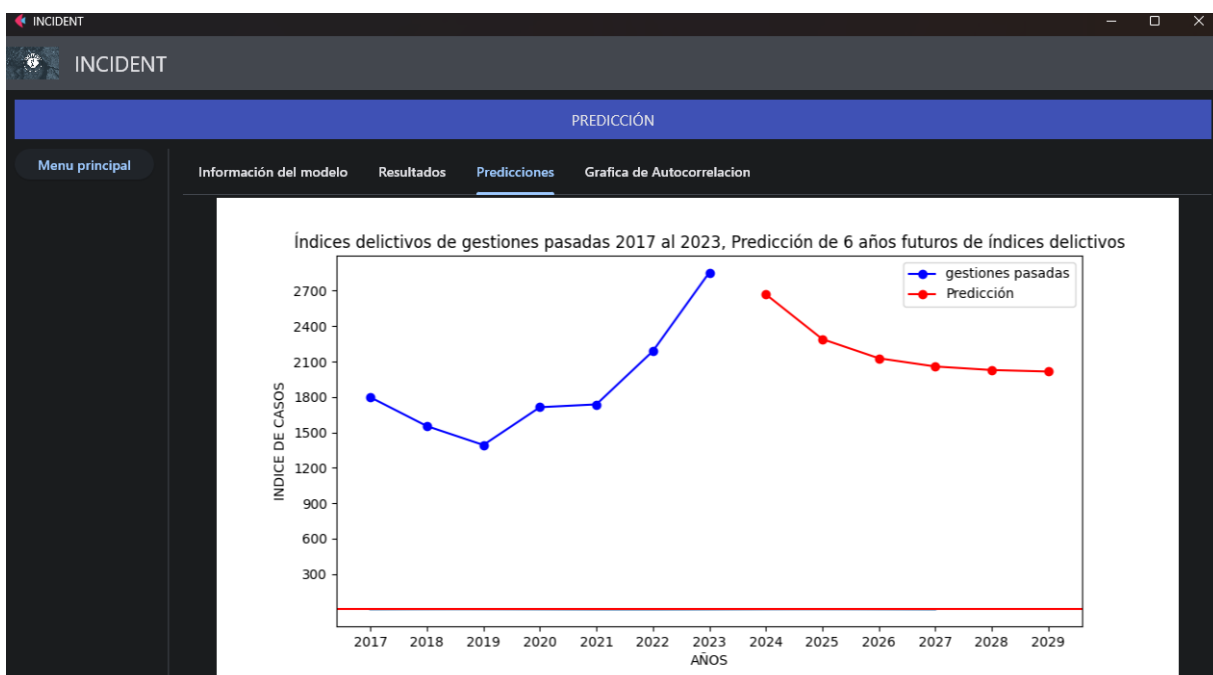
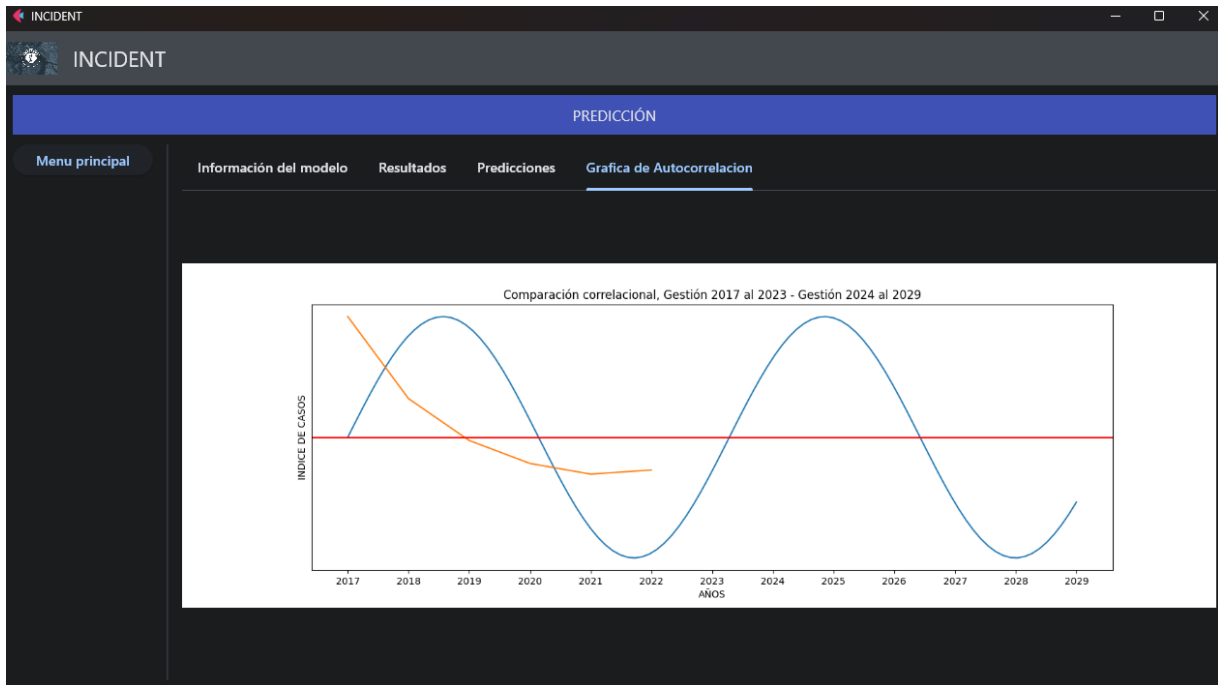




Tabla 21.

*Comparación Correlacional*

#### 4.5 PRUEBAS DE HIPÓTESIS

Se obtendrá un modelo predictivo basado en la Minería de Datos, teniendo un prototipo con la capacidad de realizar predicciones a 3 años de los índices delictivos del Municipio de El Alto con una eficiencia del 85% para un mejor conocimiento y contribución a la toma de decisiones por parte de las autoridades pertinentes.

- **H<sub>1</sub>**

Se obtendrá un modelo predictivo basado en la Minería de datos, teniendo un prototipo con la capacidad de realizar predicciones a 1 años de los índices delictivos del Municipio de El Alto con una eficiencia del 55% para un mejor conocimiento y contribución a la toma de decisiones por parte de las autoridades pertinentes.

- **H<sub>0</sub>**

No se obtendrá un modelo predictivo basado en la Minería de Datos, teniendo un prototipo con la capacidad de realizar predicciones, no logrará la predicción de los índices delictivos del Municipio de El Alto.

**Tabla 22.**

*Índices Pasados, índices Futuros*

| <b>GESTIONES PASADAS</b> | <b>ÍNDICES PASADOS</b> |
|--------------------------|------------------------|
| 2017                     | 1797                   |
| 2018                     | 1553                   |
| 2019                     | 1393                   |
| 2020                     | 1713                   |
| 2021                     | 1737                   |
| 2022                     | 2188                   |
| 2023                     | 2855                   |
| <b>GESTIONES FUTURAS</b> | <b>ÍNDICES FUTUROS</b> |
| 2024                     | 2669                   |
| 2025                     | 2290                   |
| 2026                     | 2127                   |
| 2027                     | 2058                   |
| 2028                     | 2028                   |
| 2029                     | 2015                   |

*Nota.* Índices pasados, Índices futuros.

**Tabla 23.***Media, Desviación estándar*

| <b>Gestiones</b> | <b>Media</b> | <b>Desviación<br/>estándar</b> | <b>Varianza</b> | <b>Promedio</b> |
|------------------|--------------|--------------------------------|-----------------|-----------------|
| 2017 – 2023      | 1890         | 490.57                         | 240665          | 1890            |
| 2024 - 2029      | 2197         | 240.79                         | 63466           | 2197            |

*Nota.* Cálculo de la media y la desviación estándar.

$$H = P \geq 85\% ; H1 = P_1 \leq 55\% ; n_1 = 7 ; n_2 = 6$$

Varianza común:

**( 12 ) Varianza común**

$$s_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$s_c^2 = \frac{(7 - 1) * 240665 + (6 - 1) * 63466}{7 + 6 - 2}$$

$$s_c^2 = 160120$$

Estadístico de prueba:

**( 13 ) Estadístico de prueba**

$$t = \frac{X_1 - X_2}{\sqrt{\frac{S_c^2}{n_1} + \frac{S_c^2}{n_2}}}$$

$$t = \frac{1890 - 2197}{\sqrt{\frac{160120}{7} + \frac{160120}{6}}}$$

$$t = -1,379$$

Valor crítico:

**( 14 ) Valor crítico**

$$gl = (n_1 + n_2 - 2)$$

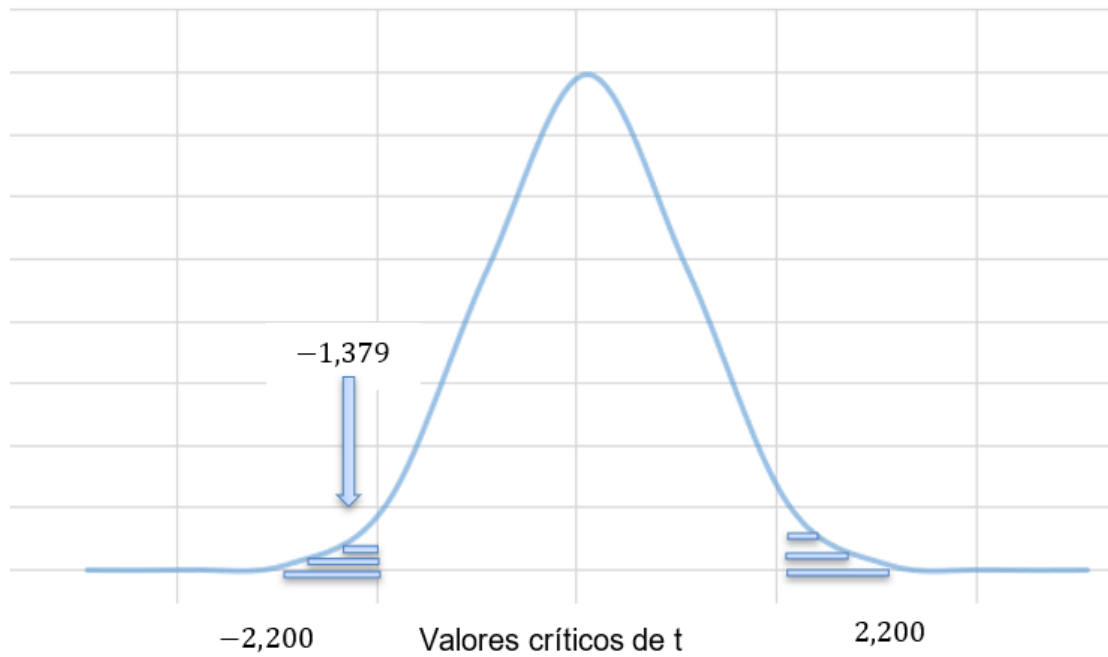
$$gl = (7 + 6 - 2)$$

$$gl = 11$$

$$\alpha = 0.05$$

$$t_{\left(1-\frac{\alpha}{2}\right) * (n_1+n_2-2)} = 2,200$$

$$p - valor = 0,1952$$



Se obtiene como resultado el cálculo  $t = -1,379$  por lo que cae en la zona de crítica, por lo que se rechaza la Hipótesis Nula  $H_0$ , y se acepta la Hipótesis Alternativa  $H_1$ , con un intervalo de confianza de 85%



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# **CAPÍTULO V**

## CONCLUSIONES Y RECOMENDACIONES



## 5 CAPÍTULO V

### 5.1 CONCLUSIONES Y RECOMENDACIONES

#### 5.1.1 CONCLUSIONES

Se concluyo el trabajo de investigación y haber realizado el desarrollo del modelo, se ha logrado el objetivo principal planteado anteriormente.

Diseñar un modelo predictivo de índices delictivos basado en Minería de Datos, aplicado al Municipio de El Alto, utilizando datos recolectados de las diferentes instituciones relacionadas con la temática de Seguridad, con el fin de anticipar el incremento de los índices delictivos.

En el Diseño metodológico se demuestra el análisis de los datos, la aplicación de la minería de datos y el desarrollo del modelo de proyección INCIDENT llegando a la totalidad del objetivo.

Los objetivos específicos se cumplieron de la siguiente manera:

- Analizar y relevar los requerimientos necesarios sobre el índice delictivo en el Municipio de El Alto.
- Diseñar un modelo de Minería de Datos para el análisis del índice delictivo.
- Identificar los requerimientos de un modelo predictivo, para los datos recolectados.
- Desarrollar el algoritmo del modelo predictivo para el índice delictivo.
- Analizar los índices delictivos mediante los datos obtenidos utilizando el modelo de Minería de Datos.
- Establecer la relación entre los datos y el prototipo diseñado.

- Verificar e interpretar los resultados obtenidos por el prototipo diseñado con el modelo de Minería de Datos.

¿Podrá un modelo predictivo basado en Minería de Datos ayudar a predecir los futuros índices delictivos en el Municipio de El Alto?

Al diseñar un Modelo predictivo donde se visualiza los posibles datos futuros sobre los índices delictivos en el Municipio de El Alto, las autoridades competentes podrán tener conocimiento sobre estos sucesos y se espera que puedan tomar decisiones previniendo el posible crecimiento en los índices delictivos, que estos provocan que el Municipio de El Alto sea potencialmente muy peligrosa.

### **5.1.2 RECOMENDACIONES**

Como recomendaciones propongo que.

Registrar más datos, de gestiones pasadas e incorporarlas a la base de datos para obtener una mejor predicción.

Recopilar información adicional sobre los índices delictivos en diferentes instituciones, y complementar a la base de datos existente en el modelo.

Diseñar una segunda versión con más registros sobre los índices delictivos del Municipio de El Alto y aplicarla a otros Municipios.

Estudiar diferentes algoritmos y aplicar otros modelos de predicción basados en Minería de Datos para mejorar la predicción e interpretación de resultados.





**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# BIBLIOGRAFIA



## BIBLIOGRAFÍA

- Alejandro Aquino, X. G. (22 de 06 de 2015). *Analítica Predictiva de Big Data en Sistemas de Base de Datos Relacionales [Tesis de grado, Universidad Mayor de San Andres]*.  
<https://repositorio.umsa.bo/handle/123456789/7748>
- Almora Castro, N. C. (22 de 06 de 2018). *Analisis y usos del Big Data Aplicado en la Universidad Nacional San Luis Gonzaga de Ica [Tesis de grado, Universidad Nacional "San Luis Gonzaga]*. <http://repositorio.unica.edu.pe/handle/20.500.13028/3095>
- Álvarez, C. A. (2011). *Metodología de la Investigación Cuantitativa y Cualitativa*.
- Apaza Ajtona, M. (22 de 06 de 2022). *Modelo de Predicción Sobre El Índice Decrecimiento Del Cáncer De Mama En Las Mujeres De Edades Entre 20 A 40 Años De La Ciudad De La Paz, Basado En Minería De Datos [Tesis de grado, Universida Publica de El Alto]*. <http://repositorio.upea.bo/jspui/handle/123456789/116>
- Aular, M., Yelitza , J., & Talavera Pereira, R. (22 de 06 de 2007). Minería de Datos como soporte a la toma de decisiones empresariales. *Scielo*, 23(52), 107-108.  
[https://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S1012-15872007000100008](https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1012-15872007000100008)
- Bel, W. (2020). *Algoritmos y Estructuras de Datos en Python*. UADER.
- Beltrán, B. (2018). *Minería de datos*.
- Bernal Torres, C. A. (2010). *Metodología de la Investigación* (tercera edición ed.). (O. F. Palma, Ed.) Pearson Educación de Colombia Ltd.
- Berry, M., & Linoff, G. (1997). *Data Mining Techniques For Marketing, Sales*. Editorial Wiley.
- Cadena, I., Rendon, M., Aguilar, A., Salinas Cruz, J., Cruz Morales, E., & Sangerman, J. (09 de 2017). Métodos cuantitativos, métodos cualitativos o su combinación en la

investigación: un acercamiento en las ciencias sociales. *Revista Mexicana de Ciencias Agrícolas*, 8(7), 16. <https://www.redalyc.org/pdf/2631/263153520009.pdf>

Carranza, H. R. (2013). La Conducta delictiva, factores causales. *Big Bang Faustiniiano*, 5.

Carrión, F. (2002). *Seguridad Ciudadana* (1 ed.).

<https://www.casede.org/BibliotecaCasede/sfseguridadciudadana.pdf>

Choque Soto, C. C., Vargas Camberos, M. R., & Aguirre Alvarez, W. (2023). *Dossier de estadísticas em seguridad ciudadana 2017-2022*. Editorial del Estado Plurinacion de Bolivia. [https://www.mingobierno.gob.bo/sites/default/files/2023-12/Dossier\\_Seguridad\\_Ciudadana.pdf](https://www.mingobierno.gob.bo/sites/default/files/2023-12/Dossier_Seguridad_Ciudadana.pdf)

ConceptoDefinition. (22 de 06 de 2022). <https://conceptodefinicion.de/windows-10/>

Crespo, J. R. (2021). *El test de la T de Student, ¿sólo en poblaciones normales*.

Denzer, P. (2002). *PostgreSQL*.

<http://profesores.elo.utfsm.cl/~agv/elo330/2s02/projects/denzer/informe.pdf>

Derecho. (22 de 06 de 2022). <http://universojus.com/definicion/delictivo>

Emanuel, L. G. (2019). *Implementación del editor de código fuente Visual, [Tesina, Universidad Politécnica de Sinaloa Programa Académico de Ingeniería en Informática]*.

<http://repositorio.upsin.edu.mx/formatos/632016030156LopezGuzmanSergioEmanuel6114.pdf>

eni. (22 de 06 de 2022). <https://www.ediciones->

[eni.com/open/mediabook.aspx?idR=ae7ae21eaafa369ac7341b1b92b0d7d8](https://www.ediciones-eni.com/open/mediabook.aspx?idR=ae7ae21eaafa369ac7341b1b92b0d7d8)

Fassler, K. W. (2020). *Metodos Cuantitativos de Analisis*.

<https://archivos.juridicas.unam.mx/www/bjv/libros/13/6180/7.pdf>

- Fernández Villaroel, R. F. (22 de 06 de 2016). *Big Data En El Comportamiento De Datos Climatológicos Y Estrategias Internacionales De Reducción De Desastres Para La Gestion De Riesgo Ambiental [Tesis de grado, Universidad Mayor de San Andres]*.  
<http://repositorio.umsa.bo/xmlui/handle/123456789/10689>
- Fernández, S. d. (2022). *Estadística Descriptiva [Universidad Autonoma de Madrid]*.  
<https://www.fuenterrebollo.com/Economicas2013/indices-teoria.pdf>
- Gómez, A., Migani, M., Migani, S., & Otazú, A. (22 de 06 de 2010). *Cocomo II un modelo de estimacion de proyectos de software*.  
<https://blogadmi1.files.wordpress.com/2010/11/cocomoIIfull.pdf>
- González, A., & González, S. (2000). *Introduccion a R*. <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- Gutierrez Alvarez, R. A. (2019). *Implementacion de un Modelo Predictivo Basado en DataMining para la Mejora de la Gestion de ventas en la Distribuidora Jimenez e Iriarte S.A [Tesis de grado, Universidad Nacional Tecnológica de Lima Sur]*.  
<https://repositorio.untels.edu.pe/jspui/handle/123456789/534>
- Gutiérrez cruz, D., Albarrán Fernández, Y. A., & Rodríguez Páez, C. L. (2019). *MINERÍA DE DATOS I*.
- Hernández Avalos, P. A. (2015). *Minería de datos*.
- Hernández Cedano, J. Á. (2015). *Modelo de minería de datos para identificacion de patrones que influyen en el aprovechamiento academico*.
- Hidalgo Mamani, J. M. (22 de 06 de 2020). *Modelo De Predicción Basado En Minería De Datos Sobre Indices De Deserción De Alumnos [Tesis de grado, Universidad Publica de El Alto]*. <http://repositorio.upea.bo/jspui/handle/123456789/213>

ISO 25000. (22 de 06 de 2022). *La familia de normas ISO/IEC 25000*.

<https://iso25000.com/index.php/normas-iso-25000>

Linares Berrocal, C. (22 de 06 de 2019). *Implementación de un sistema de big data aplicado a la migración de datos bajo la distribución cloudera con apache hadoop, en el banco interbank [Tesis de grado, Universidad Tecnológica del Perú]*.

<https://hdl.handle.net/20.500.12867/1944>

Maida, E. G., & Pacienza, J. (2015). *Metodologías de desarrollo de software. [Tesis de Licenciatura en Sistemas y Cuptación. Facultad de Química E ingeniería]*.

Martínez de Pisón Ascacibar, F. J. (2003). *Optimización mediante técnicas de minería de datos del ciclo de recocido de una línea de galvanizado [Tesis doctoral, Universidad de la Rioja]*. <https://dialnet.unirioja.es/servlet/tesis?codigo=81>

Marzal, A., & Gracia, I. (2009). *Introducción a la programación con Python*.

<https://repositori.uji.es/xmlui/handle/10234/24305>

Microsoft. (23 de 06 de 2022). *Microsoft*. <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>

Moine, J. M. (2013). *Metodologías para el descubrimiento de conocimientos un estudio comparativo [Tesis de maestría, Universidad Nacional de la Plata]*.

<https://sedici.unlp.edu.ar/handle/10915/29582>

Moreno, M. O. (2018). *Sistema de evaluación de la calidad de los componentes web centrados en los usuarios finales [Tesis de maestría, Univesidad Politecnica de Madrid]*. [https://oa.upm.es/51462/1/TFM\\_MIGUEL\\_ORTEGA\\_MORENO.pdf](https://oa.upm.es/51462/1/TFM_MIGUEL_ORTEGA_MORENO.pdf)

Pressman, R. S. (22 de 06 de 2010). *Ingeniería de Software Un enfoque práctico (7 ed.)*.

McGRAW-HILL INTERAMERICANA EDITORES, S.A. DE C.V.

<https://www.javier8a.com/itc/bd1/ld->

[Ingenieria.de.software.enfoque.practico.7ed.Pressman.PDF](#)

profile. (22 de 06 de 2022). <https://profile.es/blog/mejores-frameworks-python/>

Ramírez, Q., Hernandez, O., & Ferri, R. (2004). *Introducción a la minería de datos*. Pearson Educación. S.A.

Riquelme, J., Ruiz, A., & Gilbert, K. (2006). Minería de Datos Conceptos y Tendencias. *Inteligencia Artificial*, 10(14), 29.

Rodríguez Suárez, Y., & Anolandy Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana*.

Rojas, A. E. (17 de 10 de 2022). *Gestión del conocimiento en la educación universitaria*.

Dialnet: <https://www.google.com/search?q=Dialnet->

[GestionDelConocimientoEnLaEducacionUniversitaria-5654272.pdf&oq=Dialnet-](#)

[GestionDelConocimientoEnLaEducacionUniversitaria-](#)

[5654272.pdf&aqs=chrome..69i57j69i60.4830j0j1&sourceid=chrome&ie=UTF-8](#)

Rubin de Celis, R., Sanjinés Tudela, G. N., & Aliaga Londermann, J. (2012). Delincuencia en Bolivia desde una perspectiva espacial. *Revista Latinoamericana de Desarrollo Económico*. [http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S2074-47062012000200006](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2074-47062012000200006)

Sagrario, R. G. (2011). *Minería de Datos en Encuestas de Profesores [Tesis de Grado, Universidad Nacional Autónoma de México]*.

<http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/227/tesis.pdf?sequence=1>

Sampieri, R. H. (2006). *Metodología de La investigacion*. sexta edicion.

- Shapiro , P., & WJ, F. (1991). *El proceso de descubrimiento de conocimiento en base de datos*. Prensa AAAI / MIT.
- Sommerville, I. (2011). *Ingeniería de Software* (9 ed.). Pearson Educación de Mexic, S.A.  
[https://gc.scalahed.com/recursos/files/r161r/w25469w/ingdelsoftwarelibro9\\_compress ed.pdf](https://gc.scalahed.com/recursos/files/r161r/w25469w/ingdelsoftwarelibro9_compress ed.pdf)
- Suarez Yuniet, R., & Diaz Amador, A. (07 de 2009). Herramientas de Minería de Datos.  
*Revista Cubana de Ciencias Informaticas*, 3(3-4), 9.  
<https://www.redalyc.org/pdf/3783/378343637009.pdf>
- Thompson, I. (17 de 10 de 2022). *Definición de Información*.  
<https://www.itmerida.mx/panel/posgrado/archivos/mga/PDF.pdf>
- Timón, C. E. (2017). *Analisis Predictivo Tecnicas y Modelos y Aplicaciones del mismo - Herramientas Open Source que Permiten su uso*.  
<https://openaccess.uoc.edu/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf>
- Tudela, G. N. (2012). Análisis de las actividades delictivas en Bolivia desde la perspectiva económica. *Dialnet*, 23. file:///C:/Users/adan/Downloads/Dialnet-AnalisisDeLasActividadesDelictivasEnBoliviaDesdeLa-5114829.pdf
- Valdivia, A. E. (2011). Violencia e Inseguridad en la Ciudad de El Alto. *Scielo Bolivia*.  
[http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S0040-29152011000100005](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S0040-29152011000100005)
- Vizcanio, P. A. (2008). *Aplicación de Técnicas de Inducción de Árboles de Decision a Problemas de Clasificación Mediante el Uso de Weka*.

Ximénes, C., & Revuelta, J. (2022). *Análisis de Datos en Lenguaje R* (EDICIONES DE LA UNIVERSIDAD AUTÓNOMA DE MADRID ed.).

<https://libros.uam.es/uam/catalog/download/1163/2054/2338?inline=1>





**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

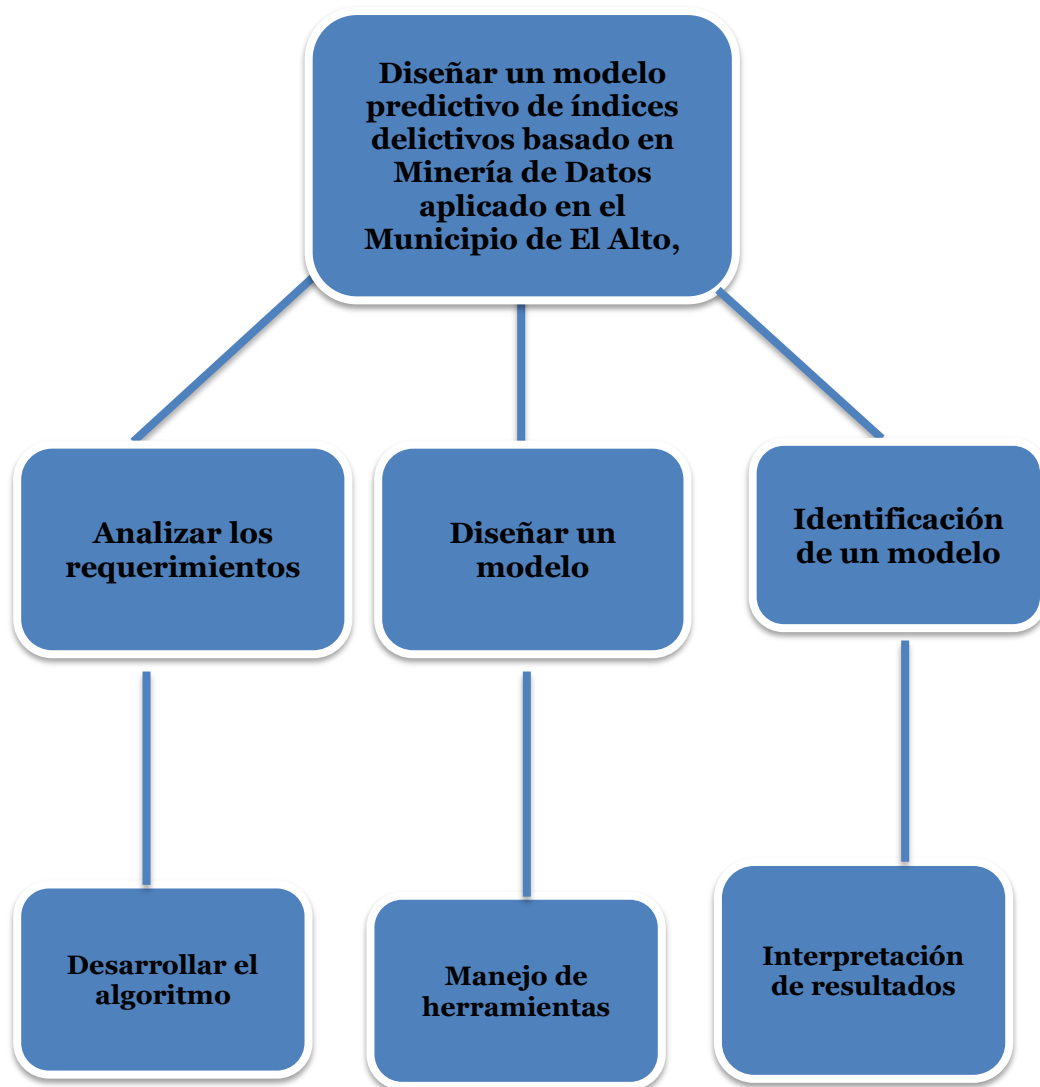
# ANEXOS



**ANEXO A.  
ARBOL DE PROBLEMAS**



**ANEXO B**  
**ARBOL DE OBJETIVOS**



El Alto, 20 de junio de 2024

Señor:  
Lic. Ing William Roque Roque  
**DIRECTOR DE CARRERA**  
**INGENIERÍA DE SISTEMAS**  
Presente. –

**REF. AVAL DE CONFORMIDAD**

Distinguido director de carrera:  
Mediante la presente tengo a bien comunicarle mi conformidad del Trabajo de Grado:

**TITULO: MODELO PREDICTIVO DE ÍNDICES DELICTIVOS BASADO EN MINERÍA DE DATOS APLICADO EN EL MUNICIPIO DE EL ALTO**

**MODALIDAD: TESIS DE GRADO**

Univ. Adan Eddy Paredes Barrionuevo

Registro Universitario: 200017211

Cedula de Identidad: 8362144 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la Carrera de Ingeniería de sistemas de la Universidad Pública de El Alto.

Atentamente,

.....  
M. Sc. Lic. Ing Fanny Helen Perez Mamani  
TUTOR METODOLÓGICO  
TALLER DE GRADO II

El Alto, 20 de junio de 2024

Señor:  
M. Sc. Lic. Ing Fanny Helen Perez Mamani  
**TUTOR METODOLÓGICO**  
**TALLER DE GRADO II**  
Presente. –

**REF. AVAL DE CONFORMIDAD**

Distinguido tutor metodológico:  
Mediante la presente tengo a bien comunicarle mi conformidad del Trabajo de Grado:

**TITULO: MODELO PREDICTIVO DE ÍNDICES DELICTIVOS BASADO EN MINERÍA DE DATOS APLICADO EN EL MUNICIPIO DE EL ALTO**

**MODALIDAD: TESIS DE GRADO**

Univ. Adan Eddy Paredes Barrionuevo

Registro Universitario: 200017211

Cedula de Identidad: 8362144 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la Carrera de Ingeniería de sistemas de la Universidad Pública de El Alto.

Atentamente,

.....  
Lic. Ing. Carmiña Argani Gutiérrez  
TUTOR REVISOR

El Alto, 20 de junio de 2024

Señor:  
M. Sc. Lic. Ing Fanny Helen Perez Mamani  
**TUTOR METODOLÓGICO**  
**TALLER DE GRADO II**  
Presente. –

**REF. AVAL DE CONFORMIDAD**

Distinguido tutor metodológico:  
Mediante la presente tengo a bien comunicarle mi conformidad del Trabajo de Grado:

**TITULO: MODELO PREDICTIVO DE ÍNDICES DELICTIVOS BASADO EN MINERÍA DE DATOS APLICADO EN EL MUNICIPIO DE EL ALTO**

**MODALIDAD: TESIS DE GRADO**

Univ. Adan Eddy Paredes Barrionuevo

Registro Universitario: 200017211

Cedula de Identidad: 8362144 LP

Para su defensa pública y evaluación correspondiente a la materia de Taller de Grado II, de acuerdo al reglamento vigente de la Carrera de Ingeniería de sistemas de la Universidad Pública de El Alto.

Atentamente,

.....  
Lic. Ing. Sergio Ramiro Rojas Saire  
TUTOR ESPECIALISTA



**INGENIERÍA  
DE SISTEMAS**  
UNIVERSIDAD PÚBLICA DE EL ALTO

# MANUAL DE USUARIO

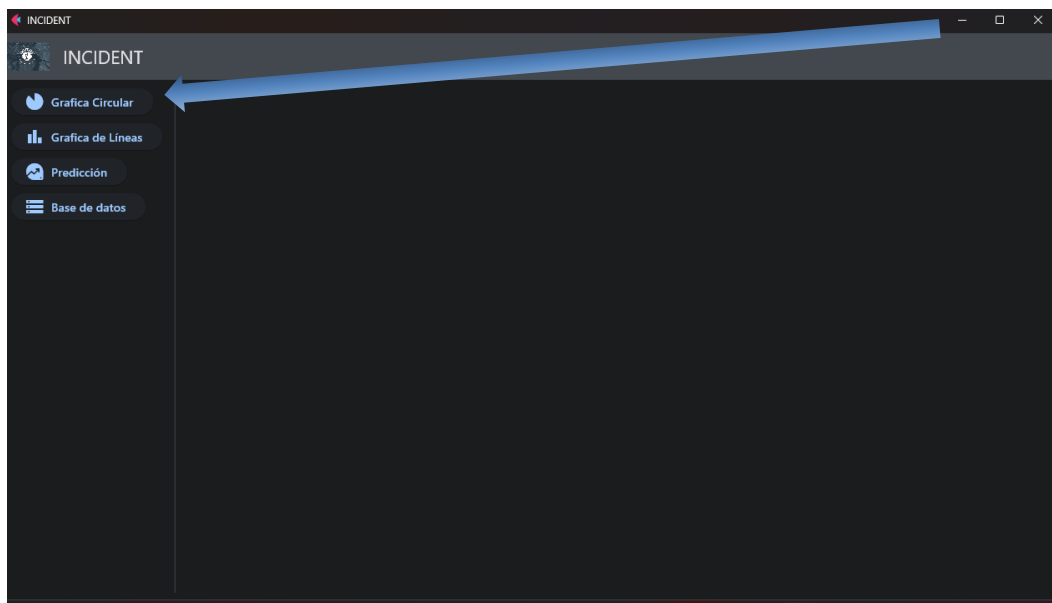


## MANUAL DE USUARIO



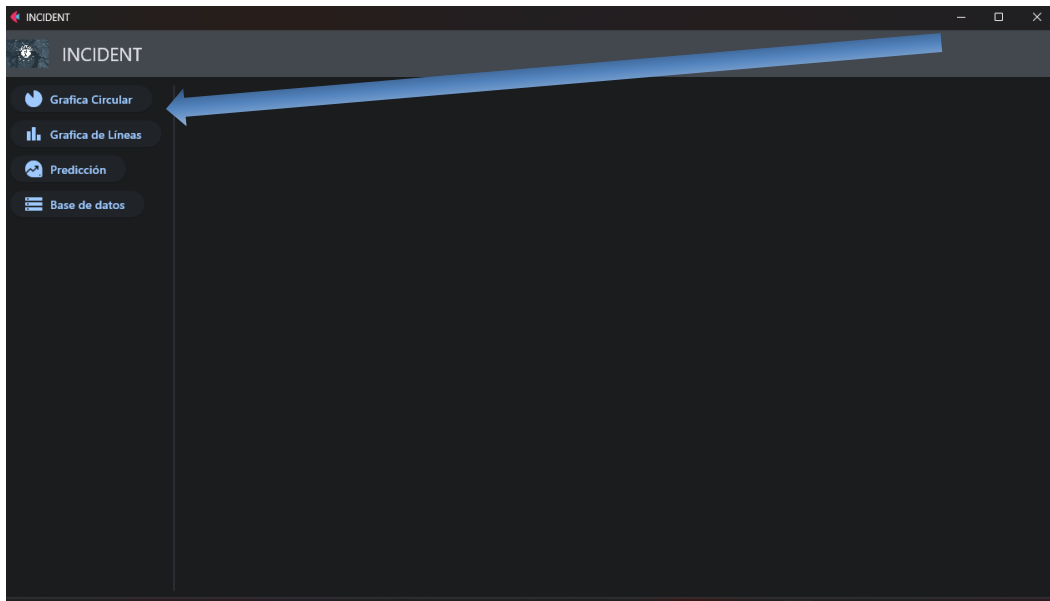
El modelo predictivo Basado en Minería de datos, fue diseñada con la necesidad de predecir datos futuros de los índices delictivos.

Interfaz gráfica: Grafica circular

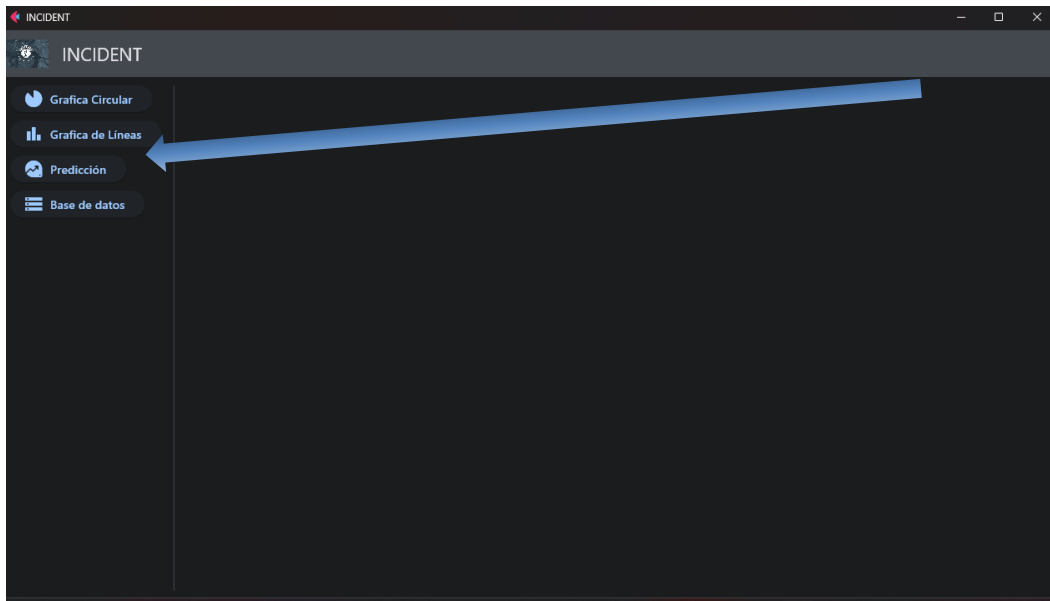




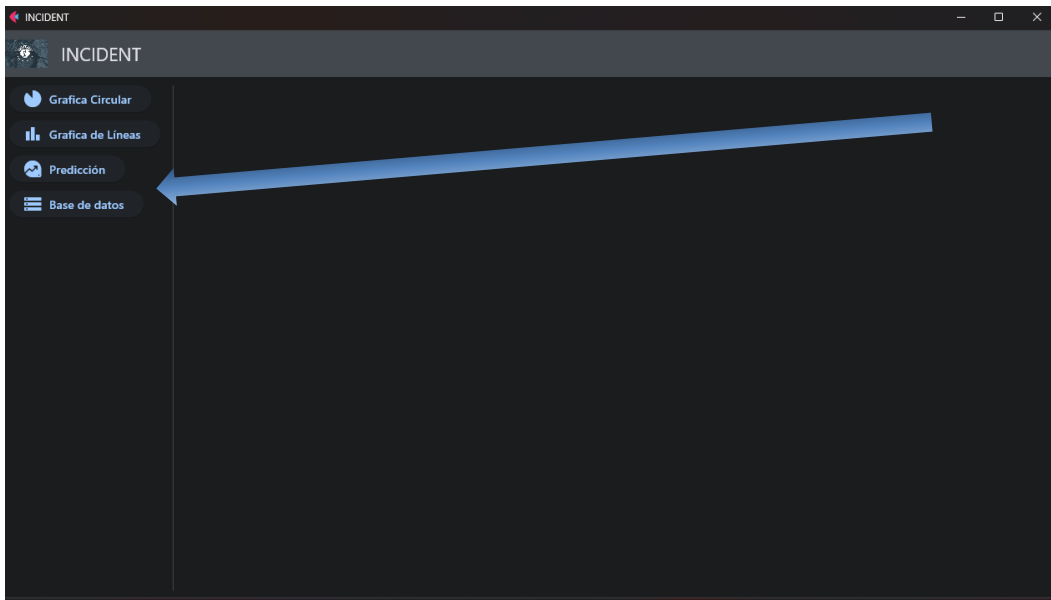
## Interfaz gráfica: Grafica de Líneas



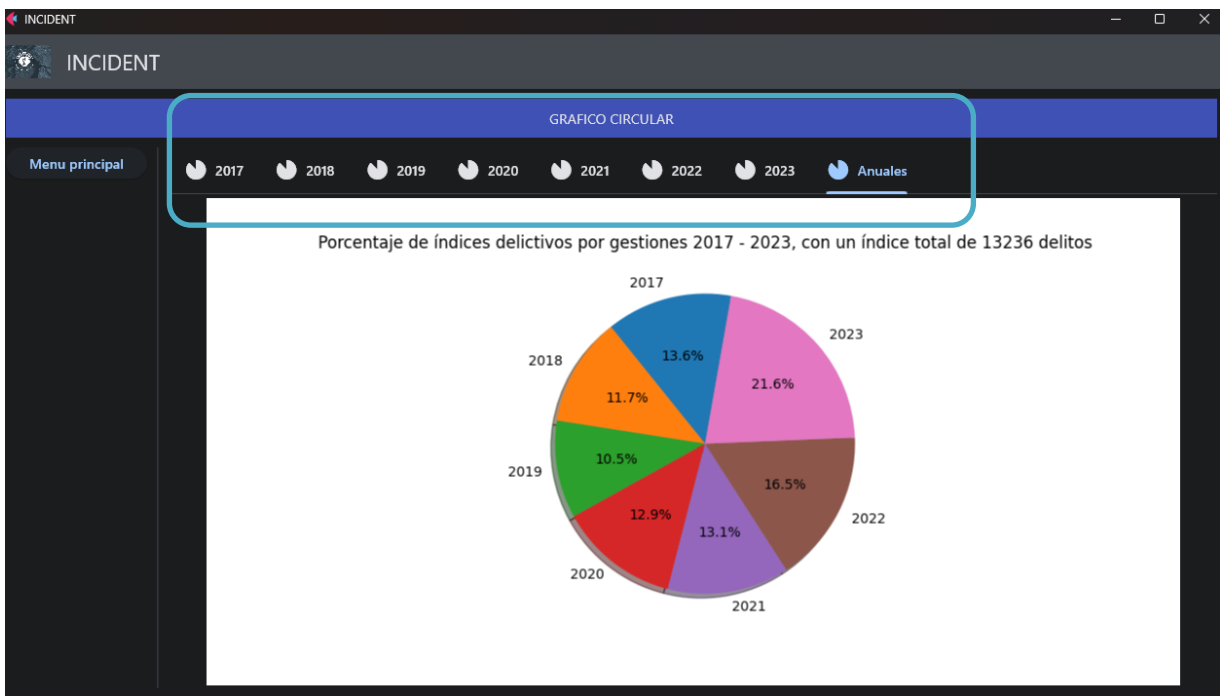
## Interfaz de predicción: Predicción



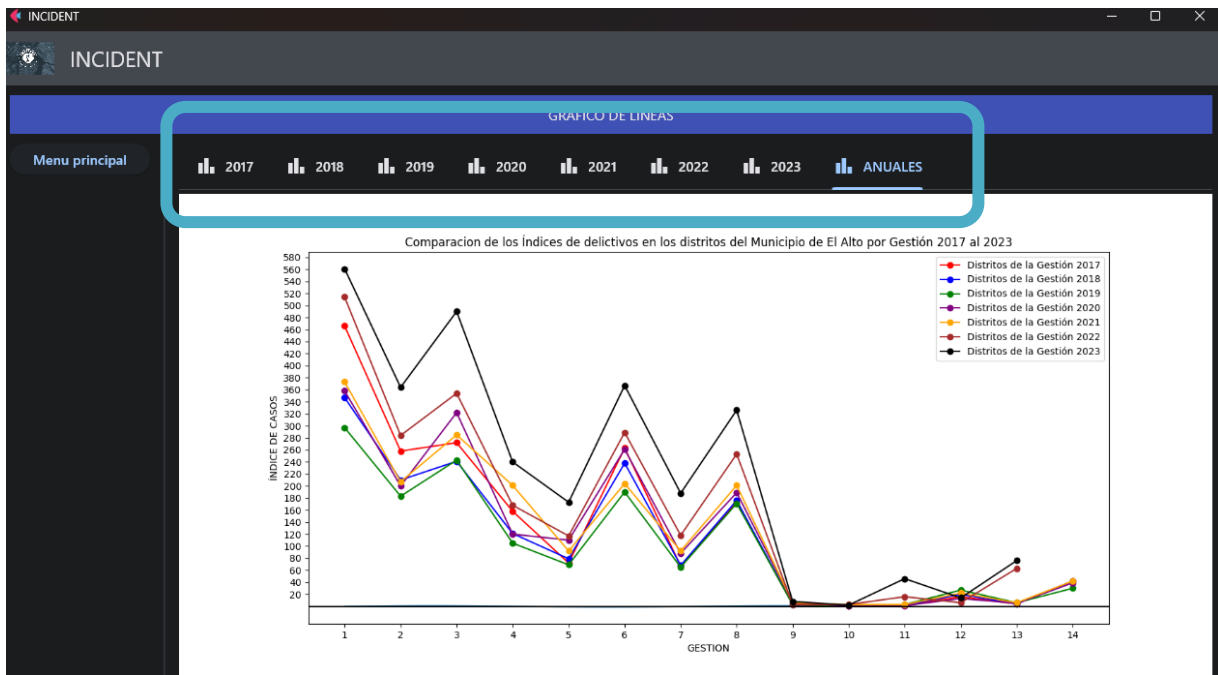
## Interfaz de Base de Datos: Base de datos



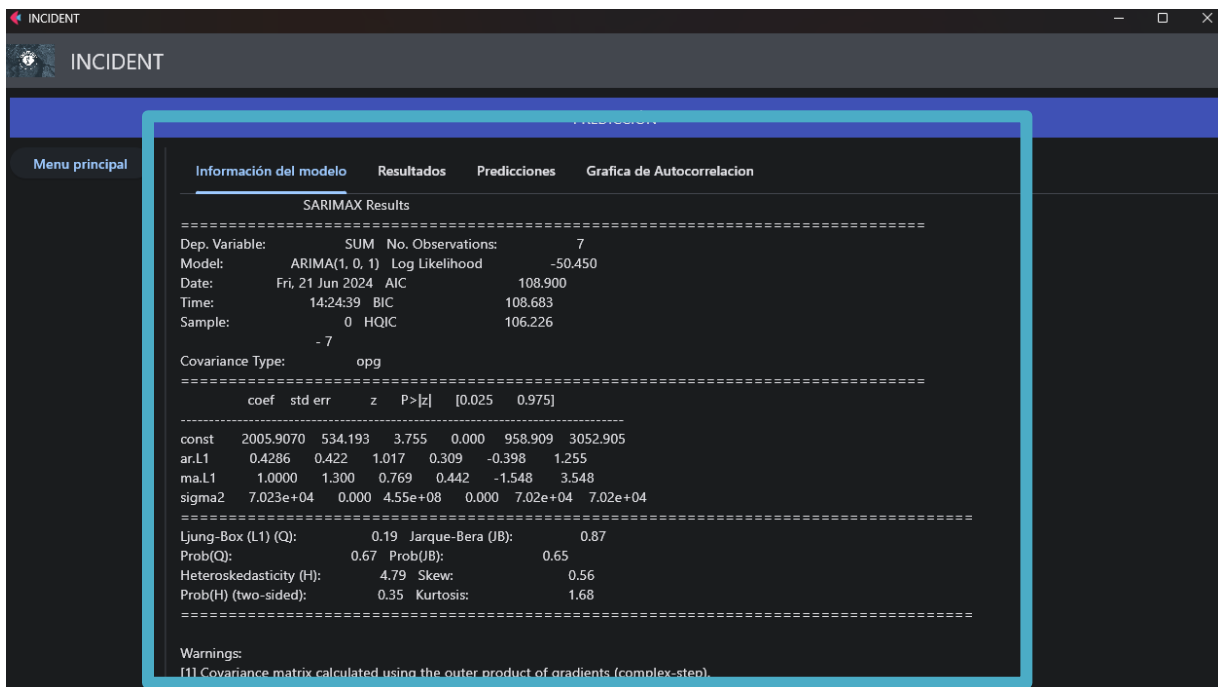
## Graficas: Grafica Circular



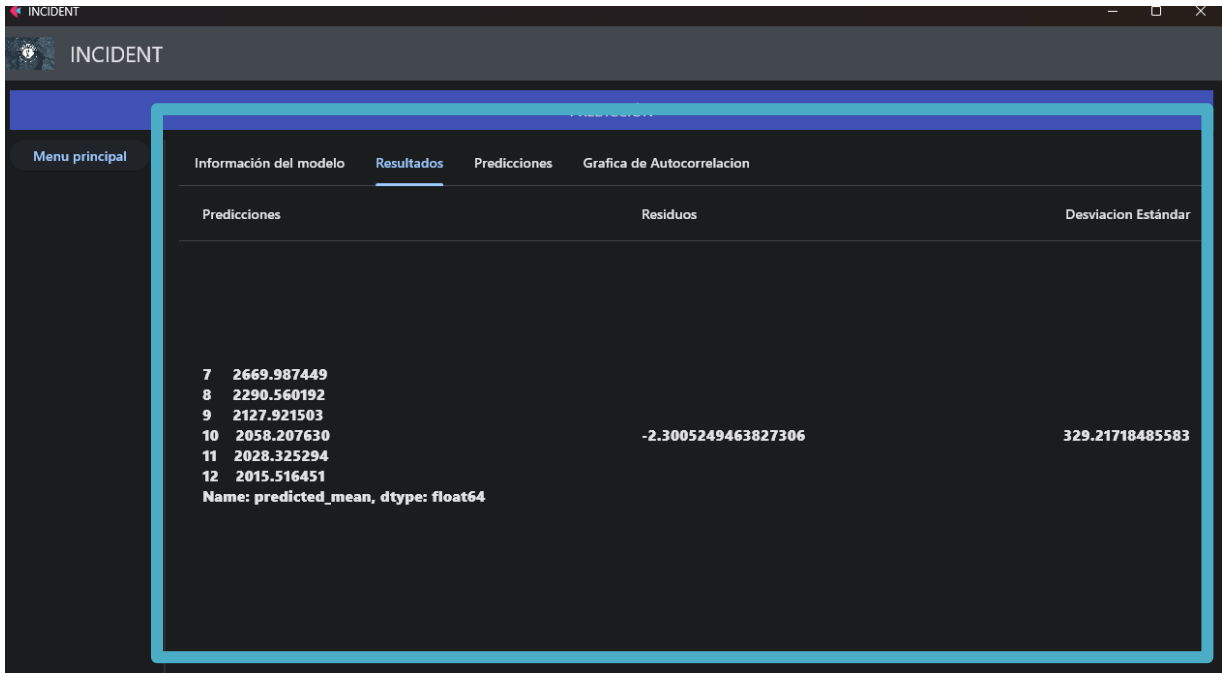
## Graficas: Grafica de Línea:



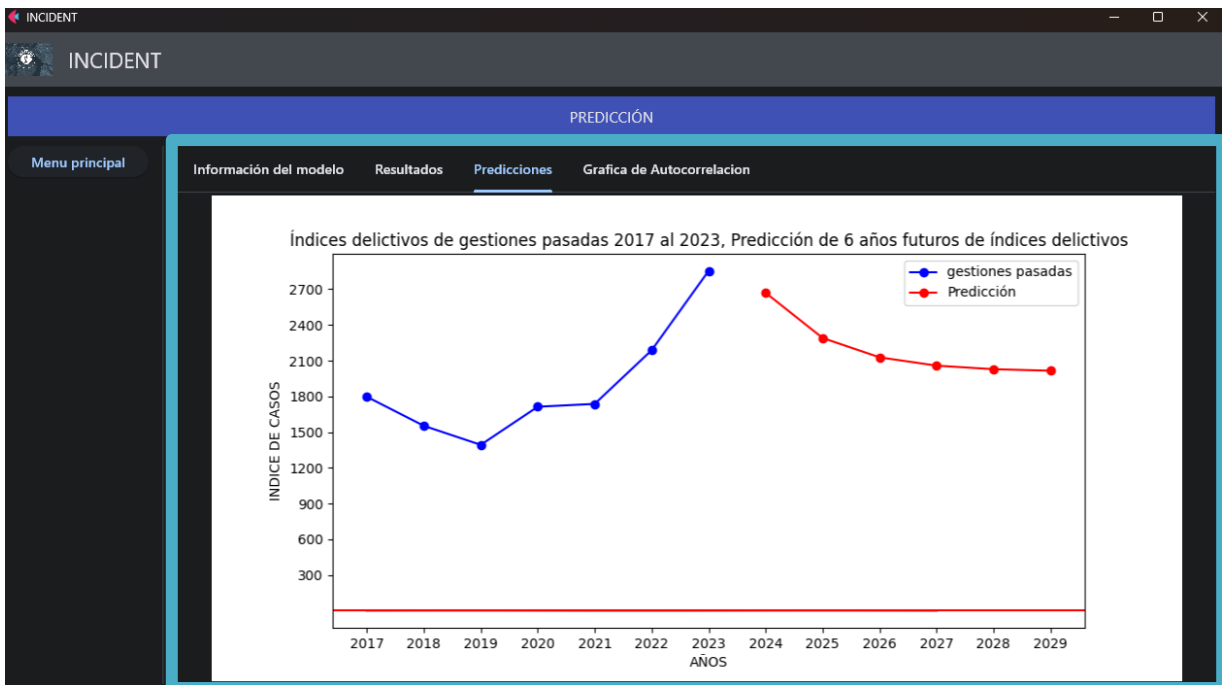
## Interfaz de Predicción:



## Resultados del modelo:



## Predicciones del modelo:



## Interfaz de la Base de Datos:

The screenshot shows a web application window titled 'INCIDENT'. On the left, there is a sidebar with a 'Menu principal' button. The main content area displays a table with tabs for 'GESTIÓN 2017', 'GESTIÓN 2018', 'GESTIÓN 2019', 'GESTIÓN 2020', 'GESTIÓN 2021', 'GESTIÓN 2022', and 'GESTIÓN 2023'. The 'GESTIÓN 2017' tab is selected, showing a table with 9 rows of data. The table columns are: N°, ZONA, NOMBRE DE CASO, TOTAL CASO, DISTRITO, and AÑO. The data is as follows:

| N° | ZONA            | NOMBRE DE CASO          | TOTAL CASO | DISTRITO | AÑO  |
|----|-----------------|-------------------------|------------|----------|------|
| 1  | 12 DE OCTUBRE   | LESIONES GRAVES Y LEVES | 96         | 1        | 2017 |
| 2  | 16 DE JULIO     | LESIONES GRAVES Y LEVES | 28         | 6        | 2017 |
| 3  | VILLA BOLIVAR D | LESIONES GRAVES Y LEVES | 26         | 2        | 2017 |
| 4  | ALTO LIMA       | LESIONES GRAVES Y LEVES | 19         | 6        | 2017 |
| 5  | VILLA DOLORES   | LESIONES GRAVES Y LEVES | 16         | 3        | 2017 |
| 6  | SENKATA         | LESIONES GRAVES Y LEVES | 10         | 8        | 2017 |
| 7  | RIO SECO        | LESIONES GRAVES Y LEVES | 16         | 4        | 2017 |
| 8  | VILLA TUNARI    | LESIONES GRAVES Y LEVES | 10         | 4        | 2017 |
| 9  | CIUDAD SATELITE | LESIONES GRAVES Y LEVES | 14         | 1        | 2017 |