UCV HACER

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

Revista de Investigación y Cultura - Universidad César Vallejo

UCV HACER Rev. Inv. Cult. Volumen 11, Número 1, Enero - Marzo 2022
Chiclayo - Lambayeque - Perú

Aplicación de minería de datos en datos abiertos de Ecuador: Delitos

Application of data mining in open data from Ecuador: Crimes

COLINA VARGAS, Alejandra Mercedes[1]; ESPINOZA MINA, Marcos Antonio[2]

Universidad Ecotec

RESUMEN

Ecuador en los últimos años ha registrado un significativo incremento de diversos delitos, principalmente homicidios y robos. El gobierno y la ciudadanía deben obtener, de forma permanente y oportuna, datos e información significativa de los delitos consumados; que favorezcan a la toma de decisiones, en la definición de políticas y estrategias ajustadas al entorno local, para la disminución de los niveles de la delincuencia, que afecta a la sociedad y a su desarrollo. Este articulo propone hacer un reconocimiento de la realidad de los datos abiertos en el Ecuador sobre delincuencia, y del proceso de minería de datos, utilizando Pentaho y Orange. Se siguió el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD), para desarrollar el proceso de análisis de datos criminales y la correspondiente identificación de patrones relacionados con los delitos. Este estudio permitió identificar la existencia de un conjunto de documentos que dan sustento legal a la aplicación de datos abiertos en la Administración pública; sin embargo, se identificaron limitadas fuentes oficiales de datos abiertos relacionadas con delitos. Se extrajo y se tomó datos del Ministerio de Gobierno, validando, a través de herramientas de minería de datos, la potencial utilidad para la exploración y detección de patrones delictivos y su consecuente beneficio en el poder de decisión de organismos competentes.

Palabras clave: Análisis de datos, software de ordenador, delincuencia, reconocimiento de patrones.

ABSTRACT

In recent years, Ecuador has seen a significant increase in various crimes, mainly homicides and robberies. The government and citizens must obtain, in a permanent and timely manner, significant data and information on the crimes committed, which will help in decision-making, in the definition of policies and strategies adjusted to the local environment, in order to reduce the levels of crime, which affects society and its development. This article proposes to make a recognition of the reality of open data in Ecuador on crime, and the process of data mining, using Pentaho and Orange. The process of Knowledge Discovery in Databases (KDD) was followed to develop the process of analysis of criminal data and the corresponding identification of patterns related to crime. This study identified the existence of a set of documents that provide legal support for the application of open data in public administration; however, limited official sources of crime-related open data were identified. Data was extracted and taken from the Ministry of Government, validating, through data mining tools, the potential usefulness for the exploration and detection of crime patterns and their consequent benefit in the decision making power of competent bodies.

Keywords: Data analysis, computer software, delinquency, pattern recognition.

Recibido: 22 de febrero de 2022
Aceptado: 10 de marzo de 2022
Publicado: 11 de marzo de 2022

[1]Doctora en Educación, Magister Scientarium en Gerencia de las Tecnologías de Información y Comunicación, Docente Universidad ECOTEC, e-mail: acolina@ecotec.edu.ec, https://orcid.org/0000-0003-1514-8852
[2]Magister en Sistemas de Información, Magister en Negocios Internacionales y Gestión en Comercio Exterior, Docente Universidad ECOTEC, e-mail: mespinoza@ecotec.edu.ec, https://orcid.org/0000-0003-1530-7243

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

INTRODUCCIÓN

América Latina enfrenta grandes desafíos en el tema de seguridad; ha sido descrita como una región insegura y violenta. La evidencia de los registros delictivos, las encuestas de víctimas, y las estadísticas de salud, sugieren que los temores públicos, sobre la seguridad, están fundamentados y existe una incidencia cada vez mayor de este grave problema social (Villalta et al., 2016).

En el Ecuador se tiene la percepción de que la capacidad de las autoridades, llamadas a mantener la seguridad ciudadana, es muy limitada. La falta de una estrategia efectiva sobre delitos está minando la confianza de la ciudadanía en el gobierno y la policía nacional. La violencia, los delitos y los crímenes se incrementan de forma más acelerada, que a su vez generan costos a la sociedad, para combatirlos y disminuir el daño que causa (Feijoó et al., 2018).

Los países latinoamericanos, en su gran mayoría, han optado durante las últimas décadas por enfrentar al fenómeno criminal, a través de la sanción penal y la cárcel. En consecuencia, han adoptado políticas de mano dura, en la que prevalece el castigo por sobre la prevención y la rehabilitación. Pero ese camino no ha dado resultados, no se ha disuadido la delincuencia, por lo contrario, se percibe su incremento. Esas políticas públicas deben ser modificadas de forma inteligente (Coimbra & Briones, 2019).

Con el apoyo del Banco Interamericano de Desarrollo (BID), el Ministerio del Interior del Ecuador, se estableció el "Plan Nacional de Seguridad Ciudadana y Convivencia Social Pacífica 2019-2030", en donde se muestra cada uno de los nueve objetivos estratégicos a ser alcanzados por este plan; por su parte, el sexto, señala la necesidad de fortalecer los sistemas de información, inteligencia e investigación que permitan producir conocimiento a todo nivel sobre todos los riesgos, amenazas y oportunidades, que afecten a la seguridad ciudadana y pública.

Los investigadores buscan nuevos enfoques para agrupar distintos tipos de delitos dentro de las categorías de delitos existentes, para ayudar a los analistas de delitos a evaluar grandes volúmenes de datos, con el objetivo de ampliar la comprensión de los problemas delictivos y respaldar el diseño de intervenciones de prevención (Birks et al., 2020).

En marzo de 2015, la Comisión de Estadística de las Naciones Unidas aprobó la "Clasificación Internacional de Delitos con Fines Estadísticos", como norma estadística, con el fin de mejorar la coherencia y comparabilidad global de las estadísticas sobre el delito, además de mejorar la capacidad de análisis de información. Esta clasificación puede aplicarse a todas las formas de datos sobre el delito, cualquiera que sea la etapa del proceso de justicia penal en que se recopilan, así como a los datos recopilados en las encuestas de victimas (UNODC, 2015). En Ecuador, en abril de 2018, el Instituto Nacional de Estadísticas y Censos (INEC) presentó una versión provisional de "La Clasificación Nacional de Delitos con Fines Estadísticos" (CNDE).

Muchas veces la estadística descriptiva clásica, que es presentada por las instituciones del gobierno a la ciudadanía, no refleja el problema real; para intentar hacerlo, es necesario mejorar el tratamiento de la información, que obligue a evolucionar a las instituciones gubernamentales, en la forma que tratan el análisis de información criminal (Britos et al., 2008). La minería de datos es una opción, integra a las bases de datos, con las concepciones de la estadística, el aprendizaje automático y los visores de datos. Esta unión de disciplinas y herramientas se ha dado por el incremento del volumen de datos de los sistemas informáticos, y a la necesidad de disponer de información significativa para tomar decisiones (Valenga et al., 2008).

Por otro lado, la Constitución de la República del Ecuador en su artículo 18 señala que todas las personas, en forma individual o colectiva, tienen derecho a buscar, recibir, intercambiar, producir y difundir información veraz, verificada, oportuna, contextualizada, plural, sin censura previa acerca de los hechos, acontecimientos y procesos de interés general, y con responsabilidad ulterior. Además, a acceder libremente a la información generada en entidades públicas, o en las privadas que manejen fondos del Estado o realicen funciones públicas.

Los datos abiertos, se convierten en un elemento integral del modelo de gobierno digital, el cual muchos países intentan aplicar; en el que se

promueve una participación cada vez más activa de la ciudadanía en el proceso de resolución de diversos problemas de la Administración pública. Los datos brindan no solo un soporte analítico a las decisiones tomadas, sino que también, apoyan el funcionamiento del mecanismo de retroalimentación, que permite detectar problemas y corregir errores en tiempo real (Kosorukov, 2017).

El Ecuador desde hace algunos años ha celebrado representativos acuerdos internacionales y locales, referentes a la importancia de la disponibilidad de datos abiertos; la Tabla 1, presenta los más destacables y parte de sus definiciones.

Tabla 1
Identificación de variables objetivos del estudio

Acuerdos	Definiciones
Carta Iberoamericana de Gobierno Electrónico (CLAD):	Las administraciones públicas serán responsables de la integridad, veracidad y calidad de los datos, servicios e informaciones en sus sitios electrónicos y portales. (IX Conferencia Iberoamericana de Ministros de Administración Pública y Reforma del Estado, 2007).
Carta Iberoamericana Gobierno Abierto:	Los gobiernos deberían diseñar, implementar y desarrollar portales de datos abiertos y elaborar normativas y/o pautas metodológicas para su adecuada categorización, uso y reutilización por parte de la ciudadanía y otros actores del ecosistema del gobierno abierto (XVII Conferencia Iberoamericana de Ministras y Ministros de Administración Pública y Reforma del Estado, 2016).
Compromiso de Lima en la VIII Cumbre de las Américas:	Se pacta promover y/o fortalecer la implementación de políticas y planes nacionales y, cuando corresponda, sub-nacionales, en materia de: gobierno abierto, gobierno digital, datos abiertos, …, considerando para ello la participación de la sociedad civil y otros actores sociales. (VIII Cumbre de las Américas, 2018).
Acuerdo Ministerial 011-2020:	Se emite la "Política de datos abiertos, de aplicación en la administración pública central", (Ministerio de Telecomunicaciones y de la Sociedad de la Información, 2020).
Acuerdo Ministerial 035-2020:	Se emite "Guía de datos abiertos de aplicación en la Administración pública central ecuatoriana" (Ministerio de Telecomunicaciones y de la Sociedad de la Información, 2020).

Fuente. Elaboración propia

En este tenor, la implementación de datos abiertos por parte de las Administraciones públicas, es una práctica que ha traído innumerables beneficios para la sociedad; sin embargo, en muchos países, incluido el Ecuador, se presentan barreras que dificultan la reutilización, lo que desalienta el desarrollo; y de forma adversa permite a los administradores públicos argumentar que por el bajo uso, se reducen la cantidad de publicaciones o incluso toman decisiones de dejar en el olvido estas plataformas (Alves et al., 2018).

En el Ecuador, el Ministerio de Gobierno tiene la misión de garantizar la seguridad ciudadana, y a pesar, de ser uno de los organismos llamados a cumplir con los acuerdos locales e internacionales relacionados a datos abiertos, es lamentable que no forme parte del grupo de las pocas instituciones que han publicado datos en el Portal Datos Abiertos Ecuador. El portal web propio de este organismo, tiene algunos enlaces a información estadística relacionados con temas de seguridad, pero en su mayoría, los datos son solo de consulta y con resultados reducidos, que no permiten reutilizarse.

Las labores de prevención, detección y esclarecimiento del delito por parte de los organismos competentes para la mejora de la seguridad ciudadana, requiere concentrar los recursos en actividades contra la delincuencia basadas en información efectiva. La toma de decisiones y las consecuentes estrategias se deben crear con base al análisis de los registros criminales históricos, que delinean los incrementos de los tipos de delitos y los lugares en donde se realizan; deben implementarse intervenciones de prevención específicas, basadas la evidencia estadística generada por estos datos.

Es necesario el despliegue de agentes de policía en las calles, particularmente en los llamados "puntos críticos". Utilizando información histórica, se puede estimar los períodos de tiempo en que ocurren los delitos con mayor regularidad, y predecir los puntos críticos por áreas, a nivel provincial, cantonal, parroquial o sector. Si las intervenciones policiales fueran enfocadas, gracias a la información pertinente, consistente y oportuna, pudieran reducirse los delitos en el país significativamente.

A pesar de esta problemática social, son pocos los análisis realizados sobre la situación de la delincuencia en el país, y así presentar tentativas soluciones; los interesados comentan, que la dificultad radica en los reducidos datos públicos disponibles para hacer estudios. Es así que, frente a estos problemas, el presente trabajo de investigación plantea una revisión sobre la realidad en el Ecuador de los datos abiertos, los resultados de la búsqueda y toma de datos abiertos vinculados con los delitos, y se demuestra, a través de dos importantes herramientas informáticas la aplicación de minería de datos, presentando finalmente algunos patrones significativos encontrados, relacionados con incidentes delictivos.

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

METODOLOGÍA

El proceso de minería de datos, consiste en la extracción de conocimiento a partir de un gran volumen de datos. Este comprende un conjunto de actividades que van desde la selección de las fuentes de datos, limpieza y preparación de los datos, aplicación de algoritmos o modelos y por último la interpretación de los resultados obtenidos.

Existen varios métodos de apoyo a la minería de datos, los cuales se pueden agrupar según el objetivo del análisis (Lausch, et al., 2014). Para la demostración de aplicación de minería de datos a los datos abiertos relacionados a delitos en el Ecuador, se siguió la secuencia metodológica propuesta en el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD, siglas en inglés de Knowledge Discovery in Databases); conocido como el proceso de descubrimiento de conocimientos útiles, no trivial, a partir de datos, para la identificación de patrones válidos, novedosos y entendibles (Fayyad et al., 1996).

El término KDD fue acuñado en 1989, resaltando que el conocimiento es el producto final de un descubrimiento basado en datos, a partir del cual se ha popularizado en los campos de la inteligencia artificial y el aprendizaje automático (Piatetsky-Shapiro, 1990). El proceso de descubrimiento tiene la característica de ser automático, plantea la fusión de descubrimiento y análisis de datos, una vez extraídos los patrones o tendencias en forma de reglas o funciones, para que el usuario haga los análisis (Timarán-Pereira et al., 2016); permitiendo de esta manera encontrar el mejor modelo que se ajuste a los fines del presente artículo.

El proceso de KDD aplicado en esta investigación contiene una secuencia ordenada, interactiva e iterativa de pasos, expuestos en la Figura 1. Según Han et al. (2012) permite el descubrimiento de patrones o tendencias en unas series de datos, como la detección de los cantones del país con niveles altos de delitos de robos y homicidios y sus tipos.

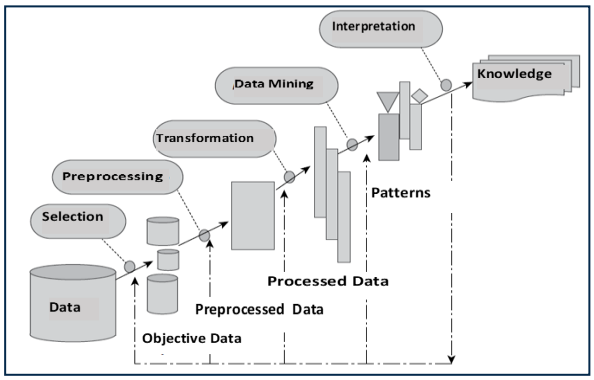

Figura 1. Etapas del proceso de KDD
Fuente. Timarán-Pereira et al., 2016

RESULTADOS Y DISCUSIÓN

Aplicación del Proceso de Minería de Datos

Para la demostración de la aplicación de minería a las fuentes de datos abiertas extraídas del Ministerio de Gobierno del Ecuador se completó las cinco etapas del proceso de KDD, las cuales comprenden:

Etapa 1. Selección

En esta primera etapa, previo a la selección del conjunto de datos, se lleva a cabo la definición de los objetivos desde el punto de vista del cliente, así como el conocimiento previo y la comprensión del área en que se realizará el análisis (Umair Shafique & Haseeb Qaiser, 2014).

Dentro de los dominios de aplicación, y el objetivo de la presente demostración, se encuentra el descubrir relaciones implícitas entre las variables: cantones, tipos de delito, año, mes, número de delitos, entre otros; con la finalidad de determinar patrones o tendencias de un tipo de delito en un lugar y periodo de tiempo. Luego, se realiza la elección, integración y recopilación de datos, se establecen las fuentes de información útiles y la ubicación de las mismas, se identifican y seleccionan las variables relevantes en los datos y se aplican las técnicas de muestreo cuando se requieren (Pérez & Santín, 2007).

La única alternativa del portal web del Ministerio de Gobierno de la República del Ecuador, que permitió obtener libremente datos, con posibilidad de reutilización, se encontró en la llamada opción de "Indicadores de Seguridad Ciudadana"; la cual, cuenta con una sección de descargas, desde donde se pudo obtener las series históricas de "Robos" y "Homicidios Intencionales", dentro de rangos de

fechas.
La aplicación de la minería de datos se desplegó en el ámbito de extraer y analizar los datos obtenidos del Ministerio de Gobierno, específicamente en los indicadores de seguridad ciudadana, las series históricas de robos y homicidios intencionales tomados de los años 2014 al 2021, publicados en el sitio web oficial de esta institución.

Se procedió una vez indicadas las series históricas (mes y año) a la descarga de los archivos dispuestos en formato de Microsoft Excel, los cuales constituyen las fuentes de datos, información base a ser procesada en el análisis. Una de las fuentes de datos utilizadas es el archivo descargado de "Robos.xls" contiene 7.333 registros cuyas variables son: tipo de delito, provincia, cantón, mes, año y número de delitos. En tanto que, el archivo de "Homicidios_Intencionales.xls", tuvo 40.737 registros, sus variables fueron: tipo de muerte, provincia, cantón, mes, año, tipo de arma, rango de edad, sexo, y número de homicidios.

Etapa 2. Preprocesamiento
Esta fase se centra en la limpieza y el preprocesamiento de fuentes de datos para completar datos que se requieren en el posterior análisis, para ello se desarrollan estrategias que coadyuven a la eliminación del ruido, las inconsistencias e incoherencias (Shafique & Qaiser, 2014), y en esa medida mejoran la calidad de los datos, la precisión y eficiencia del proceso de minería de datos (Han et al.,2012).

Se procedió a la limpieza de los archivos fuentes utilizando el programa de Microsoft Excel eliminado información no útil, como imágenes alusivas al Ministerio de Gobierno del Ecuador, registros que contienen nombre de la institución, identificación de la fuente, fecha de corte del histórico y datos del departamento encargado de la elaboración de dichos documentos.

Etapa 3. Transformación
La tercera etapa comprende, entre sus actividades operaciones básicas de transformación, la ordenación de los datos en la forma deseada, por ejemplo, convirtiendo un tipo de datos en otro, definiendo nuevos atributos, reduciendo la dimensionalidad de los datos, removiendo ruidos y valores atípicos, normalizando los datos, y decidiendo estrategias para manejar datos perdidos (Arteaga, Remigio & Calderón, 2018).
Los procesos ETL (extracción, transformación y carga) extraen datos de fuentes internas y externas de una institución, limpian y transforman estos datos, y los cargan en un almacén de datos. Estos procesos son muy complejos y costosos, comparados con otras etapas (Awiti et al., 2020). Pentaho, plantea una solución completa para la inteligencia de negocios, integrando importantes módulos; tiene uno para ETL, con una gran mantenibilidad y flexibilidad para realizar las requeridas transformaciones (Parra et al., 2016).

Estudios de tratamiento de datos comerciales utilizan Pentaho como herramienta para procesos de ETL. En la propuesta de un modelo de referencia de Data Governance, para agilizar los procesos de la cadena de suministro en las PYMES mediante técnicas de Integración de Datos y Business Analytics, de Barrenechea et al. (2019), se utilizó la herramienta Pentaho Data Integration - Spoon (Kettle). Otro ejemplo es, cómo Harvy et al. (2019) la utilizaron exitosamente para dar solución a las limitaciones de procesamiento de una gran cantidad de datos generados mensualmente, en la venta de libros.

Muchos autores coinciden que el proceso de implementación de una plataforma como Pentaho, está al alcance de las pequeñas y medianas empresas, y que, con poca inversión en soporte técnico, pueden alcanzar los beneficios que ofrece el análisis de datos (Leite et al., 2019).

Para las operaciones de transformación se utilizó el software Spoon de la ETL Pentaho Data Integration, realizando gráficamente cada una de las operaciones de depuraciones de los datos. Cada fuente de datos extraída siguió cuatro pasos de transformaciones de datos, en el caso de "Homicidios_Intencionales.xls" representados en la Figura 2 se tiene: (1) Se transformó los valores del campo MES al valor numérico correspondiente del trimestre en el año. (2) Se crea un campo denominado TRIMESTRE a partir de la operación de concatenación de los campos MES y AÑO. (3) Se eliminan los campos PROVINCIA, MES, AÑO y RANGO DE EDAD, el campo SEXO cuyos valores son "NO DETERMINADOS", pues no aportan al análisis de este estudio. (4) Se filtran los nombres de los cantones, tomando del conjunto de datos los primeros veinte cantones del Ecuador con mayor

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

número de habitantes tomados del Instituto Nacional de Estadística y Censo (INEC) del país.

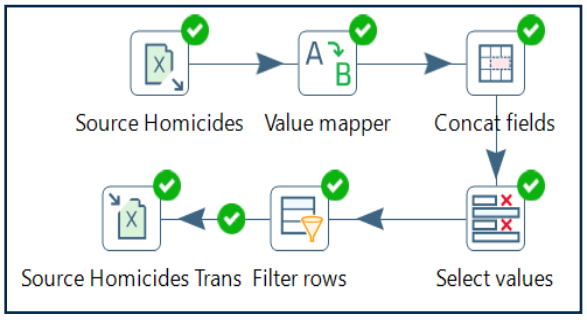

Figura 2. Transformación de fuente de datos: "Homicidios_Intencionales.xls"
Fuente. Elaboración propia, a partir de los datos procesados en Spoon.

Con la fuente de datos "Robos.xls" los pasos del 1 al 4 ilustrado en la Figura 3, (1) Se transformó los valores del campo MES al valor numérico correspondiente del trimestre en el año. (2) Se crea un campo denominado TRIMESTRE a partir de la operación de concatenación de los campos MES y AÑO. (3) Se eliminan los campos PROVINCIA, MES, AÑO pues no aportan al análisis de este estudio. (4) Se filtran los nombres de los cantones, tomando del conjunto de datos los primeros veinte cantones del Ecuador con mayor número de habitantes tomados del INEC.

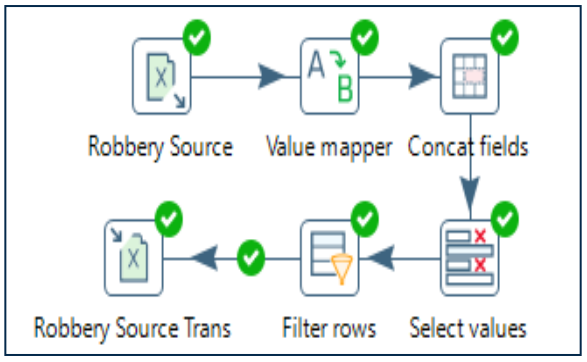

Figura 3. Transformación de fuente de datos "Robos.xls"
Fuente. Elaboración propia, a partir de los datos procesados en Spoon.

En resumen, todas las transformaciones e integración de datos de las Figuras 2 y 3 dan un paso importante en la construcción del modelo de predicción (Arteaga, Remigio, Calderón, 2018). Esta etapa se considera activa de investigación, debido a la gran cantidad de inconsistencias o datos sucios y la complejidad del problema (Han et al.,2012) que dan garantía de éxito en el análisis predictivo.

Etapa 4. Minería de datos

Una vez transformados y limpiados los datos de cada fuente, se procedió a la identificación del modelo de minería que se ajuste al tema escogido, ubicándose dentro del proceso supervisado o predictivo, pues la intencionalidad de la investigación gira en torno a la predicción de valores, a partir de un atributo o variable de un conjunto de datos. En este sentido, el proceso comprende la predicción de datos a partir de la inducción de una relación, entre una variable objetivo y otra serie de variables (Rodríguez & Díaz, 2009), es decir se pretende sea identificada la variable más influyente del problema, sin sacrificar la calidad del modelo de conocimiento obtenido del proceso de minería.

El aprendizaje supervisado por su parte, se ciñe en la inferencia de un modelo del conjunto de datos de entrenamiento, detectando las relaciones entre las variables en los datos representado en la función Y=f(X), donde se conoce a (X) como una variable de entrada u objetivo que sea capaz de predecir a la variable de salida (Y) (Belsini Gladshiya & Dr. K. Sharmila, 2021).

Esta fase comprende la aplicación de un conjunto de técnicas para analizar el rendimiento de los clasificadores; para el proceso de minería, el uso de algoritmos predictivo caracterizados principalmente por ser de asociación, ellos son: regresión logística, CN2 Rule Induction, Naive Bayes y el k- NN (Nearest Neighbour). En la evaluación de resultados se utilizó el análisis de la curva ROC (receiver operating Characteristic) y AUC (Area Under the ROC Curve), los cuales permitieron la extracción de conocimientos o patrones y relaciones ajustados a los objetivos del negocio. Todo ello, con el propósito de apoyar a la toma de decisiones en el Ministerio de Gobierno ecuatoriano o cualquier organismo interesado en aspectos relacionados a la seguridad y a la lucha contra la delincuencia, a través de información útil y valiosa.

Las variables correspondientes al estudio se presentan en Tabla 2, indicando la principal variable para predecir, variable objetivo.

Tabla 2
Identificación de variables objetivos del estudio.

Documento fuente	Variable objetivo (X)	Otras variables
Homicidios_Intencionales.xls	Canton	Tipomuerte, Trimestre, Tipoarma, Nrohomicidios
Robos.xls	Canton	Tipodelito, Trimestre, Nrodelitos

Fuente. Elaboración propia

En la actualidad, existen muchas herramientas de minería de datos; entre aquellas que son de código abierto: KNIME, RapidMiner, Orange, Weka y R Programming. Numerosos trabajos de investigación han evaluado dichas herramientas en diferentes casos. Ratra & Gulia (2020), seleccionaron a las herramientas Weka y Orange, para hacer un análisis comparativo, en la búsqueda de la mejor herramienta de minería de datos, según los requisitos. Raykar & Shet (2020) presentaron las características de las aplicaciones Weka, RapidMiner, Tanagra, Orange, R tool, KNIME, y finalmente tomaron a Weka y Orange, para realizar un análisis de rendimiento de las técnicas de clasificación de minería de datos, utilizando conjuntos de datos de atención médica.

Así mismo, Verma et al. (2019) en sus estudios "Latest Tools for Data Mining and Machine Learning" incluyen a Orange junto con Weka, RapidMiner y KNIME dentro de su evaluación de herramientas disponibles para la minería de datos y el análisis predictivo. De igual forma lo hacen SangeethaLakshmi & Jayashree (2018), en su trabajo "Comparative Analysis of Various Tools for Data Mining and Big Data Mining", quienes evaluaron KNIME, Orange, Rapid Miner y Weka.

Padmavaty et al. (2020), hacen la comparación de siete herramientas para minería de datos, utilizando ocho características y parámetros, y determinan finalmente que la herramienta de minería de datos Orange funciona bien y es de fácil uso. Es un software perfecto para el aprendizaje automático; es útil para la programación visual y el análisis de datos exploratorios; está escrito en Python, y tiene varios componentes conocidos como widgets. Esta herramienta puede ser utilizada en sistemas operativos macOS, Windows y Linux.

La herramienta de minería utilizada, para el proceso de extracción de conocimiento se ubica dentro del grupo de método de descubrimiento, facilita la detección de patrones potencialmente interesantes de forma automática (Rodríguez & Díaz, 2009). Los datos transformados obtenidos del Ministerio de Gobierno del Ecuador, fueron procesados inicialmente en el software Spoon de Pentaho, y luego por el software Orange, finamente se realizó el análisis correspondiente.

Este software brinda facilidad de uso, provee funcionalidades para leer datos, visualiza y analiza; contiene algoritmo de clasificación múltiple y de regresión utilizados para el aprendizaje automático (Belsini Gladshiya & Dr. K. Sharmila, 2021). Se identificaron los algoritmos de clasificación que posee Orange como la regresión logística, CN2 rule induction, y Naive Bayes exclusivos para clasificación; mientras que otros en cambios tanto clasificación como regresión y el k- NN (Nearest Neighbour), SVM, Random Forest (orange.com, 2021).

Se carga al programa Orange los datos transformados de "Homicidios_Intencionales.xls" (Figura 4) y de "Robos.xls" (Figura 5), inicialmente definiendo como variable objetivo CANTON para cada análisis. Dentro de las primeras acciones se procedió a seleccionar un subconjunto del total de 221 cantones del Ecuador (CEPAL, 2021), es decir se seleccionaron 20 cantones con la mayor población según datos del INEC; estos fueron: Ambato, Cuenca, Daule, Durán, Esmeraldas, Guayaquil, Ibarra, La Libertad, Latacunga, Loja, Machala, Manta, Milagro, Pasaje, Portoviejo, Quevedo, Quito, Samborondón, Santo Domingo.

Con los datos procesados, se exploran diversos modelos o algoritmos de minería que tiene Orange, en cada una de las fuentes de datos, se eligieron aquellos que demuestren mejor capacidad predictiva o rendimiento de las tres categorías de clasificación, de la variable dependiente de tipo ordinal, utilizando el widget "Test & Score" de Orange cuyo valor de medida tomado como referencia fue el AUC, debido a su capacidad invariante del umbral de clasificación elegido, y la evaluación de qué tan bueno es el test para discriminar si es un delito o no a lo largo de todo el rango de puntos de corte posibles, considerando aceptable AUC para mayor a 0,75, pues se encuentra a medio camino entre la no-discriminación (AUC = 0,50) y la discriminación perfecta (AUC = 1,00) (Cerda & Cifuentes, 2012).

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

El flujo de trabajo resultante de la fuente de datos "Homicidios_Intencionales.xls" (Figura 4), se observan como aceptables los modelos regresión logística, el k- NN (Nearest Neighbour) y CN" Rule Induction.

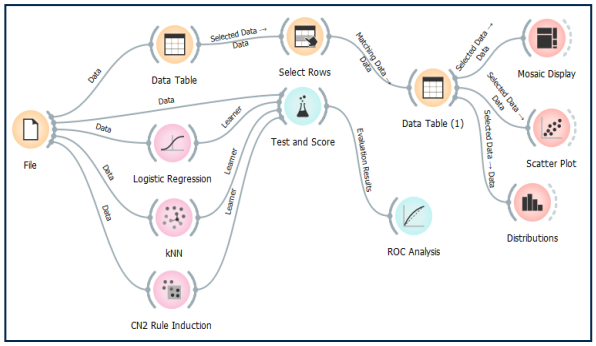

Figura 4. Flujo de trabajo con "Homicidios_Intencionales.xls"
Fuente. Elaboración propia, a partir de los datos procesados en Orange.

Por su parte, en el flujo de trabajo resultante de la fuente de datos "Robos.xls" (Figura 5), se muestra como modelos aceptables el de regresión logística, el Naive Bayes y CN" Rule Induction.

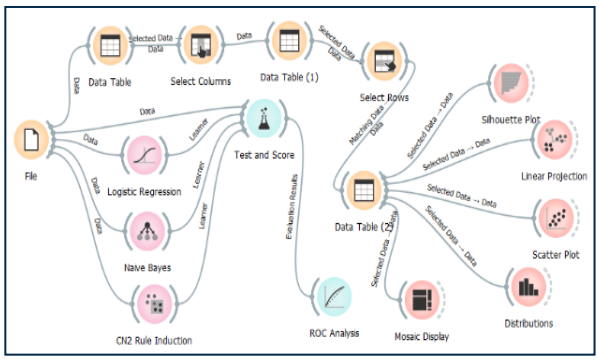

Figura 5. Flujo de trabajo con "Robos.xls"
Fuente. Elaboración propia, a partir de los datos procesados en Orange.

En este sentido, en el estudio se utilizaron una serie de clasificadores de las técnicas supervisadas de minería de texto, las cuales permitieron determinar si un atributo o clase, pertenecía o no a un determinado concepto (Haro et al., 2018). Entre ellos está, la regresión logística, modelo lineal potente, permite que las estrategias interpretativas de la relación funcional de las variables explicativas con la probabilidad de la dependiente sean ampliamente conocidas y bien fundamentadas (Temesio, García & Pérez, 2021). Se ha elegido el algoritmo denominado CN2 Rule Induction, técnica de clasificación diseñada para la sugestión eficiente de reglas sencillas y comprensibles de forma "if cond then predict class" (orange.com, 2022).

El otro algoritmo seleccionado el k-NN (siglas en inglés Nearest Neighbor), método de clasificación popular, debido a su simplicidad y sus buenos rendimientos. Consiste en un clasificador apoyado en la proximidad de sus medidas basadas en la distancia para realizar la clasificación (Allahyari, et al., 2017). En este método, la evidencia de todos los k-NN, más cercanos a las muestras de una prueba se combinan para decidir su clase. (Faziludeen & Sankaran, 2016).

Así mismo, se empleó el modelo Naive Bayes, clasificador probabilístico rápido y simple basado en el teorema de Bayes, con el supuesto de independencia de características (orange.com, 2022). Este proceso de clasificación se vale del algoritmo Bayesiano para ofrecer una solución óptima de la probabilidad de pertenencia de cada muestra a todas las clases (Haro et al., 2018).

Etapa 5. Interpretación / evaluación

Esta etapa comprende la identificación, interpretación y evaluación de los patrones, que verdaderamente representan el conocimiento basado en la variable objetivo de interés (Fayyad et al., 1996), para ello se apoya en las técnicas de visualización y representación del conocimiento para presentar lo extraído a los usuarios (Umair Shafique & Haseeb Qaiser, 2014). Se presentan los resultados de la aplicación de los diferentes clasificadores implementados a los datos de entrenamiento de cada fuente de datos, los cuales son evaluados y probados con el "Test and Score" de Orange; a partir de allí se procedió a la interpretación de los mismos.

Fuente de datos: Homicidios Intencionales

Se emplearon los modelos de regresión logística, CN2 rule induction y el Naive Bayes; posteriormente con la herramienta de visualización de resultados, el "ROC Analysis" de Orange, gráfica que mide la relación entre la sensibilidad (la tasa de positivos verdaderos) y 1 menos la especificidad (la tasa de falsos positivos) (Steyerberg, et al., 2011), se muestra el rendimiento del modelo para seis de los veinte cantones seleccionados en Tabla 3. En los resultados de los tres modelos aplicados, se observa a la regresión logarítmica con los valores de mejor capacidad predictiva. Se enfatiza en los resultados de ese algoritmo que los cantones con una buena puntuación ROC están en orden: (f)

Quito - 0.815, (e) Quevedo - 0.716 y (c) Guayaquil - 0.695.

Tabla 3
Resultados Test de prueba y ROC Analysis de los tres modelos implementados en la fuente de datos "Homicidios_Intecionales".

	Durán	**Esmeraldas**	**Guayaquil**	**Machala**	**Quevedo**	**Quito**
CN2 rule indu-	0.603	0.641	0.622	0.615	0.648	0.762
Naive	0.66	0.672	0.687	0.624	0.713	0.81
Regresión logística	0.684	0.676	0.695	0.639	0.716	0.815

Fuente. Elaboración propia

La Figura 4 muestra el clasificador de los tres modelos implementados, fijando en cada aplicación como variable objetivo el nombre del cantón, confirmando el funcionamiento o sensibilidad de cada modelo, se evidencia una buena puntuación ROC en los cantones (f) Quito (e) Quevedo y (c) Guayaquil, en ese orden. Por lo tanto, con estos resultados se infiere que estos modelos son capaces de predecir con buena precisión para esos cantones la cantidad de homicidios según los tipos en un periodo de tiempo.

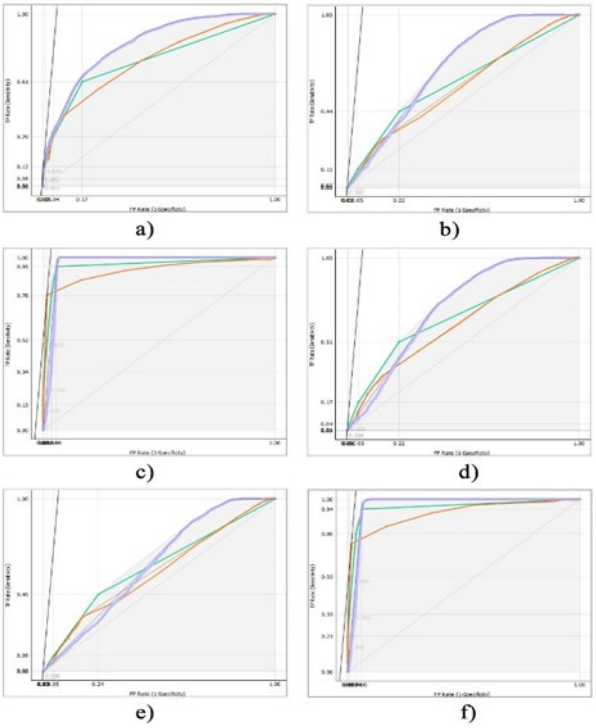

Figura 4. Resultados de Test de prueba y la curva ROC de Analysis de los tres modelos implementados. "Homicidios_Intecionales", variable objetivo cantones a) Durán, b) Esmeraldas, c) Guayaquil, d) Machala, e) Quevedo y f) Quito.
Fuente. Elaboración propia de los datos procesados en Orange.

Se aplicó los modelos de regresión logística, CN2 rule induction y el k-NN (Nearest Neighbour); conjuntamente con una herramienta de visualización de resultados, el "ROC Analysis" de Orange, para seis de los veinte cantones seleccionados. La Tabla 4, presenta los resultados de la ejecución de los tres modelos aplicados, observándose valores con mejor capacidad predictiva en la "regresión logarítmica". Se destaca en ese algoritmo que los cantones con una buena puntuación ROC en los cantones (e) Quito - 0.965, (c) Guayaquil - 0.964 y (f) Santo Domingo - 0.815, en ese orden; sus resultados son satisfactorios en las pruebas realizadas mostrando una buena capacidad predictiva.

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

Tabla 4
Resultados Test de prueba y la curva ROC de los tres modelos implementados en la fuente de datos "Robos"

	Cuenca	Durán	Guayaquil	Manta	Quito	Santo Domingo
CN2 rule inducer	0.736	0.587	0.925	0.584	0.924	0.718
KNN	0.728	0.605	0.960	0.588	0.952	0.813
Regresión logística	0.822	0.682	0.964	0.655	0.965	0.815

Fuente. Elaboración propia.

La Figura 5 muestra el clasificador de los tres modelos implementados, mostrando el buen desempeño de los modelos, se fija en cada aplicación como variable objetivo los cantones de a) Durán, b) Esmeraldas, c) Guayaquil, d) Machala, e) Quevedo y f) Quito; se obtiene una buena puntuación ROC en los cantones (e) Quito, (c) Guayaquil y (f) Santo Domingo, en esa disposición. Por lo tanto, se deduce que son capaces de predecir por cantón los tipos de delitos de robo en un periodo de tiempo con buena precisión para todos los cantones indicados.

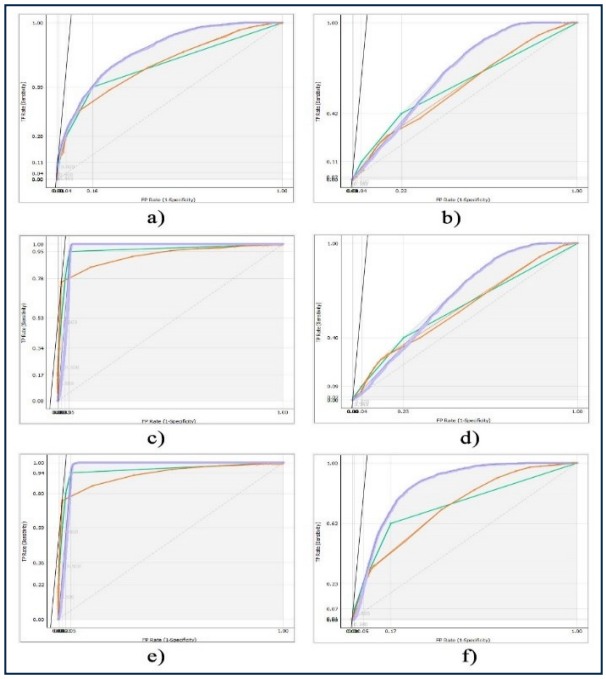

Figura 5. Resultados de Test de prueba curva ROC de los tres modelos implementados en la fuente de datos "Robos", variable objetivo cantones (a) Cuenca, (b) Guayaquil, (c) Quito.
Fuente. Elaboración propia utilizando Orange.

En base a los resultados presentados desde la fuente de datos "Homicidios_Intencionales", por el clasificador regresión logarítmica, se refleja que los mejores resultados, o que el mayor nivel de fiabilidad, corresponde al cantón de Quito con valor de 0.815; y para la fuente de datos de "Robos" por encima de 0.90, tanto en Quito como en Guayaquil. En los otros cantones las predicciones que se realicen, no serían eficaces.

Visualización de relación de variables

La representación gráfica a partir de la fuente de datos correspondiente a "Homicidios_Intencionales" con las variables declaradas en el análisis inicial y su relación, se muestran en la Figura 6, la cual permite visualizar el comportamiento de los tipos de muertes, tipos de armas, de los últimos ocho trimestres de los cantones seleccionados, ratificando la capacidad predictiva arrojada de los mejores resultados de la evaluación del clasificador, para Guayaquil y Quito, respectivamente.

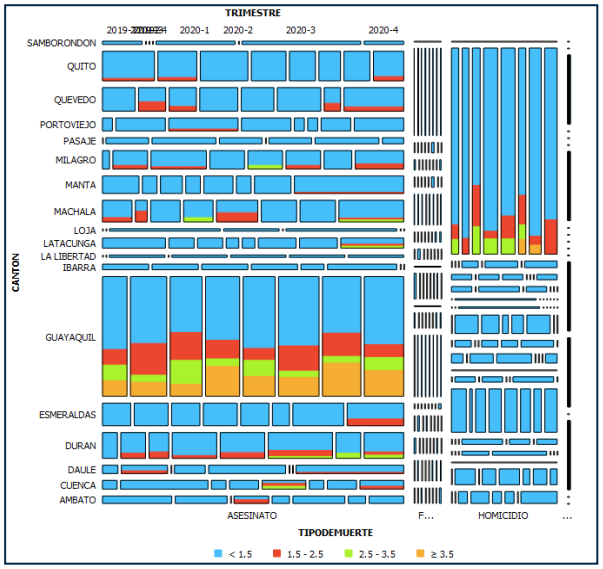

Figura 6. Visualización de la relación entre variables tipomuerte, cantón, trimestre y número de homicidios, de la fuente de datos "Homicidios_Intencionales.xls" del 2019 y 2020.
Fuente. Elaboración propia utilizando Orange.

Esta vista ayuda a la toma de decisiones en la elaboración de perfiles de homicidios en un periodo de tiempo durante el año y en un cantón, identificando, por ejemplo, qué el tipo de homicidios más frecuentemente es el "Asesinato", en el cantón Guayaquil, en los últimos trimestres de los años 2019 y 2020. Se complementa el análisis con la Figura 7, la cual permite revelar de forma específica los tipos de armas utilizadas en los delitos.

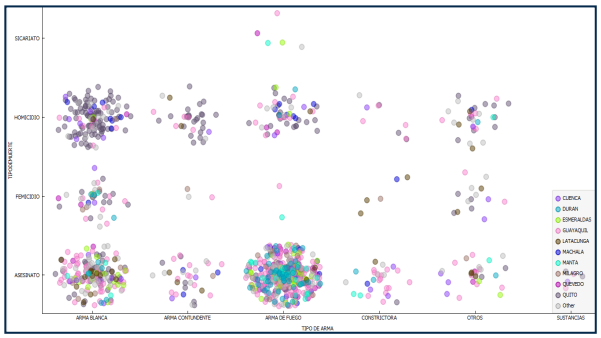

Figura 7. Visualización de la relación entre variables tipomuerte, tipoarma, cantón, trimestre y número de homicidios de la fuente de datos "Homicidios_Intencionales.xls" del 2019 y 2020.
Fuente. Elaboración propia utilizando Orange.

Estos escenarios de demostración de las herramientas de análisis predictivo son de utilidad, para los organismos públicos encargados de la seguridad ciudadana, permitiéndoles la comprensión, abordaje y predicción de delito de tipo "Asesinato", armas frecuentemente utilizadas y en esta medida diseñar políticas y estrategias efectivas contra estos actos delincuenciales. (Sangeeta Lal et al., 2020).

Un segundo ejemplo de aplicación, de la herramienta de minería de datos, se visualiza en la Figura 8; con ayuda de la fuente de datos "Robos"; se pudo observar el comportamiento de los tipos de robos, y su número, de los ocho últimos trimestres, por los cantones seleccionados, confirmando con ello, el poder de predicción del clasificador en la generación de los mejores resultados para Santo Domingo, Guayaquil y Quito, respectivamente. Este ejercicio refleja a Quito y Guayaquil con los mayores índices en este tipo de delito y sus diferentes variantes, permitiendo a las autoridades diseñar planes de prevención de manera específicas para esos cantones.

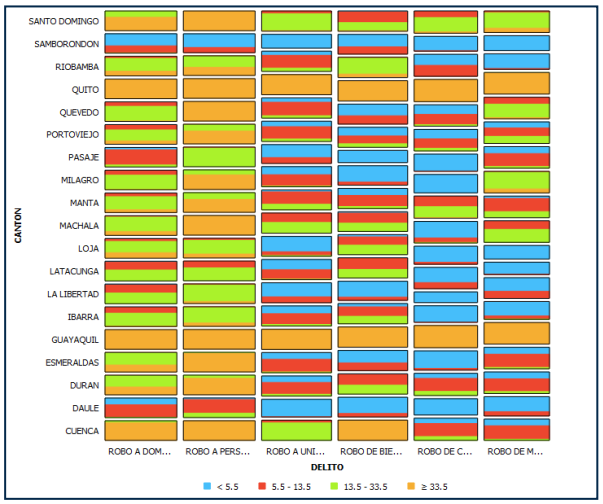

Figura 8. Visualización de la relación entre variables de la fuente de datos "Robos.xls" en los años 2019 y 2020.
Fuente. Elaboración propia utilizando Orange.

CONCLUSIONES

De la literatura revisada en torno a la realidad de los datos abiertos en el Ecuador se procedió a la identificación de los principales acuerdos y guías disponibles que sirven de referente importante para la aplicación por parte de la Administración pública a la publicación de datos abiertos, los cuales se basan en la Constitución de la República del Ecuador. Se llevó a cabo el proceso de búsqueda de datos abiertos relacionados con delitos en el Ecuador, ligado a la información de la clasificación nacional de delitos que se utiliza con fines estadísticos.

El portal web del Ministerio de Gobierno, contiene información estadística relacionada a la seguridad, pero mayoritariamente a modo de informes y con una sola una opción de descarga de datos con posibilidad de reutilización, en la sección de "Indicadores de Seguridad Ciudadana", desde la cual se extrajo dos series de datos para la aplicación de las herramientas de minería de datos.

Al complementarse la demostración desarrollada, con un reconocimiento teórico del problema que es la delincuencia y sus tipos, se facilitó el uso de la herramienta tecnológica de análisis, pues ayudó a la comprensión y conocimiento detallado del objeto de estudio.

Se plantearon y analizaron algunas herramientas

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

para la minería de datos, realizando finalmente una demostración con Spoon de Pentaho y Orange, aplicando cuatro tipos de clasificadores, cuyas medidas estadísticas de determinación de la exactitud diagnóstica de prueba para escalas continuas arrojaron valores aceptables.

Se identificó de las técnicas de clasificación de minería de datos dentro del aprendizaje supervisado utilizadas en el estudio, el modelo de "regresión algorítmica", como el algoritmo cuyos valores superan el 90% de probabilidad de detección de patrones delictivos, en este caso delitos de homicidios y robos, principalmente.

Este estudio permite confirmar la poca disponibilidad de datos abiertos de interés social, pero al mismo tiempo emerge la realidad de que no está siendo exigida la publicación de datos abiertos; tampoco, se está aprovechado las herramientas de código abierto para la extracción y descubrimiento de patrones relacionados con incidentes delictivos, que pudieran apoyar la toma de decisiones a partir de la definición de estrategias de prevención de problemas sociales, a través de la identificación temprana y caracterización de una amenaza presente.

REFERENCIAS

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2018). Indicadores de calidad de datos abiertos: El caso del portal de datos abiertos de Barcelona. *El Profesional de la Información*, *27*(2), 375. https://doi.org/10.3145/epi.2018.mar.16

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *ArXiv:1707.02919 [Cs]*. http://arxiv.org/abs/1707.02919

Alves, A. J. A., Neves, D. F., Santos, L. C., Rodrigues, M. C., & do Nascimento, R. P. C. (2018). Open Government Data Usage Overview: A Systematic Literature Mapping. *Proceedings of the Euro American Conference on Telematics and Information Systems*, 1-8. https://doi.org/10.1145/3293614.3293619

Arteaga, D., Remigio, R., & Calderón, D. (2018). Minería de Datos Aplicado al Marketing. *Número Especial de la Revista Aristas: Investigación Básica y Aplicada*, 6.

Awiti, J., Vaisman, A. A., & Zimányi, E. (2020). Design and implementation of ETL processes using BPMN and relational algebra. *Data & Knowledge Engineering*, *129*, 101837. https://doi.org/10.1016/j.datak.2020.101837

Barrenechea, O., Mendieta, A., Armas, J., & Madrid, J. M. (2019). Data Governance Reference Model to streamline the supply chain process in SMEs. *2019 IEEE XXVI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, 1-4. https://doi.org/10.1109/INTERCON.2019.8853634

Birks, D., Coleman, A., & Jackson, D. (2020). Unsupervised identification of crime problems from police free-text data. *Crime Science*, *9*(1), 18. https://doi.org/10.1186/s40163-020-00127-4

Britos, P., Fernández, E., Merlino, H., Pollo-Cataneo, F., Rodríguez, D., Procopio, C., Rancan, C., & García-Martínez, R. (2008, octubre). *Explotación de información aplicada a inteligencia criminal en Argentina*. XIV Congreso Argentino de Ciencias de la Computación.

CEPAL. (2021, diciembre). Ecuador—Sistema político electoral. Ecuador - Sistema político electoral. https://oig.cepal.org/es/paises/12/system

Cerda, J., & Cifuentes, L. (2012). Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. *Revista chilena de infectología*, *29*(2), 138-141. https://doi.org/10.4067/S0716-10182012000200003

Cerda y Cifuentes—2012—Uso de curvas ROC en investigación clínica Aspect.pdf. (s. f.).

Coimbra, L. O., & Briones, Á. (2019). Crimen y castigo. Una reflexión desde América Latina. *URVIO. Revista Latinoamericana de Estudios de Seguridad*, *24*, 26-41. https://doi.org/10.17141/urvio.24.2019.3779

De Juana-Espinosa, S., & Luján-Mora, S. (2019). Open government data portals in the European Union: Considerations, development, and expectations. *Technological Forecasting and Social*

Change, *149*, 119769. https://doi.org/10.1016/j.techfore.2019.119769

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 18.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27-34. https://doi.org/10.1145/240455.240464

Faziludeen, S., & Sankaran, P. (2016). ECG Beat Classification Using Evidential K - Nearest Neighbours. *Procedia Computer Science*, *89*, 499-505. https://doi.org/10.1016/j.procs.2016.06.106.

Feijoó, E., Gutiérrez, N., Torres, D., & Orellana, M. (2018). *Costos de la delincuencia y su impacto socio-económico en el Ecuador*. 11. https://intercostos.org/wp-content/uploads/2018/01/FEIJOO-GONZALEZ.pdf

García-García, J., & Curto-Rodríguez, R. (2018). Divulgación de información pública de las comunidades autónomas españolas (2013-2017): Portal de datos abiertos, portal de transparencia y web institucional. *El Profesional de la Información*, *27*(5), 1051. https://doi.org/10.3145/epi.2018.sep.09

Gladshiya, V. B., & Sharmila, D. K. (2021). Analyzing the risk factors and predicting the learning ability of students during pandemic and comparing machine learning algorithms using Orange tool. *Turkish Journal of Physiotherapy and Rehabilitation*, 8.

Gobierno de la República del Ecuador. (2021). *Misión/Visión—Ministerio de Gobierno del Ecuador*. https://www.ministeriodegobierno.gob.ec/valores-mision-vision/

Han, J., & Kamber, M. (2012). *Data Mining* (Morgan Kaufmann Publishers). Elsevier.

Haro, S., Zúñiga, L., Meneses A., Vera, L., & Escudero, A. (2018). Métodos de clasificación en minería de datos meteorológicos. Perfiles, 2(20), 107-113. https://doi.org/10.47187/perf.v2i20.40

Harvy, I., Matitaputty, G. A., Girsang, A. S., Michael, S., & Isa, S. M. (2019). The Use of Book Store GIS Data Warehouse in Implementing the Analysis of Most Book Selling. *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, 1-5. https://doi.org/10.1109/CITSM47753.2019.8965404

Herrera-Melo, C. A., & González Sanabria, J. S. (2019). Proposal for the Evaluation of Open Data Portals. *Revista Facultad de Ingeniería*, *29*, e10194. https://doi.org/10.19053/01211129.v29.n0.2020.10194

Instituto Nacional de Estadísticas y Censos de Ecuador. (2018). *Clasificación Nacional de Delitos con Fines Estadísticos. Versión Provisional*. https://www.cepal.org/sites/default/files/presentations/septima-reunion-gtci-clasificacion-nacional-delitos-con-fines-estadisticos-inec-ecuador.pdf

IX Conferencia Iberoamericana de Ministros de Administración Pública y Reforma del Estado. (2007). *Carta Iberoamericana de Gobierno Electronico*. https://clad.org/wp-content/uploads/2020/07/Carta-Iberoamericana-de-Gobierno-Electronico.pdf

Kosorukov, A. A. (2017). Digital government model: Theory and practice of modern public administration. *Journal of Legal, Ethical and Regulatory Issues*, *20*(3), 10.

Lausch, A. (2014). Data mining and linked open data – New perspectives for data analysis in environmental research. *Ecological Modelling*, 13. https://doi.org/10.1016/j.ecolmodel.2014.09.018

Leite, N., Pedrosa, I., & Bernardino, J. (2019). Open Source Business Intelligence on a SME: A Case Study using Pentaho. *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 1-7. https://doi.org/10.23919/CISTI.2019.8760740

Máchová, R., Hub, M., & Lnenicka, M. (2018). Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. *Aslib Journal of Information Management*, *70*(3), 252-268. https://doi.org/10.1108/AJIM-02-2018-0026

Ministerio de Gobierno del Ecuador. (2019). *Plan Nacional de Seguridad Ciudadana y Convivencia Social Pacífica 2019-2030*. https://www.ministeriodegobierno.gob.ec/wp-content/uploads/2019/08/PLAN-NACIONAL-DE-SEGURIDAD-CIUDADANA-Y-CONVIVENCIA-SOCIAL-PACI%CC%81FICA-2019-2030-1_compressed.pdf

Ministerio de Gobierno del Ecuador. (2021). *Indicadores de Seguridad Ciudadana*. http://cifras.ministeriodegobierno.gob.ec/comisioncifras/inicio.php

Ministerio de Telecomunicaciones y de la Sociedad de la Información. (2020a). *Acuerdo Ministerial No. 011-2020*.

ISSN IMPRESO: 2305-8552
ISSN ELECTRÓNICO: 2414-8695

https://www.gobiernoelectronico.gob.ec/wp-content/uploads/2020/04/Acuerdo-Poli%CC%81tica-Datos-Abiertos-17.04.20-v4-signed.pdf

Ministerio de Telecomunicaciones y de la Sociedad de la Información. (2020b). *Acuerdo Ministerial No. 035-2020*. https://www.gobiernoelectronico.gob.ec/wp-content/uploads/2021/02/Acuerdo-35-2020-Guia-Datos-Abiertos-20201211-signed-signed-signed-signed.pdf

Nascimento, F. R. A., Cesar da Rocha, J., & Garcia, A. C. B. (2018). Automated Evaluation of Open Government Data Portals: A Case Study. *International Journal of Electronic Government Research*, *14*(3), 57-72. https://doi.org/10.4018/IJEGR.2018070105

Naser, A., & Rosales, D. (2016, noviembre). Panorama regional de los datos abiertos. Avances y desafíos en América Latina y el Caribe. *Naciones Unidas*.

Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC). (2015). *Clasificación Internacional de Delitos con Fines Estadísticos*. Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC).

Padmavaty, V., Geetha, C., & Priya, N. (2020). Analysis of data mining tool Orange. *International Journal of Modern Agriculture*, *9*(4), 5.

Parra, V., Syed, A., Mohammad, A., & Halgamuge, M. (2016). Pentaho and Jaspersoft: A Comparative Study of Business Intelligence Open Source Tools Processing Big Data to Evaluate Performances. *International Journal of Advanced Computer Science and Applications*, *7*(10). https://doi.org/10.14569/IJACSA.2016.071003

Pérez, C., & Santín, D. (2007). *Minería de datos: Técnicas y herramientas* (Ediciones Paraninfo, S.A.).

Piatetsky-Shapiro, G. (1990). Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop. *AI Magazine*. https://doi.org/10.1609/aimag.v11i4.873

Ratra, R., & Gulia, P. (2020). Experimental Evaluation of Open Source Data Mining Tools (WEKA and Orange). *International Journal of Engineering Trends and Technology*, *68*(8), 30-35. https://doi.org/10.14445/22315381/IJETT-V68I8P206S

Raykar, S. S., & Shet, V. N. (2020). Cognitive Analysis of Data Mining Tools Application in Health Care Services. *2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, 1-7. https://doi.org/10.1109/ic-ETITE47903.2020.442

Registro Oficial de Ecuador. (2021). *Registro Oficial Suplemento Nº 371 del 15 de enero de 2021*. https://www.registroficial.gob.ec/index.php/registro-oficial-web/publicaciones/suplementos/item/download/13451_91cc67cffde9b156b6ae042bf07cf966

República del Ecuador. (2018). *Constitución de la República del Ecuador*. https://www.ambiente.gob.ec/wp-content/uploads/downloads/2018/09/Constitucion-de-la-Republica-del-Ecuador.pdf

Rodríguez, Y., & Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, *3*, 73-80.

Royo-Montañés, S., & Benítez-Gómez, A. (2019). Portales de datos abiertos. Metodología de análisis y aplicación a municipios españoles. *El Profesional de la Información*, *28*(6). https://doi.org/10.3145/epi.2019.nov.09

SangeethaLakshmi, & Jayashree. (2018). Comparative Analysis of Various Tools for Data Mining and Big Data Mining. *International Journal of Engineering Research And*, *V7*(11), IJERTV7IS110039. https://doi.org/10.17577/IJERTV7IS110039

Saxena, S. (2018). Open government data (OGD) in six Middle East countries: An evaluation of the national open data portals. *Digital Policy, Regulation and Governance*, *20*(4), 310-322. https://doi.org/10.1108/DPRG-10-2017-0055

Schauppenlehner, T., & Muhar, A. (2018). Theoretical Availability versus Practical Accessibility: The Critical Role of Metadata Management in Open Data Portals. *Sustainability*, *10*(2), 545. https://doi.org/10.3390/su10020545

Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, *12*(1), 6.

Steyerberg, E. W., Van Calster, B., & Pencina, M. J. (2011). Medidas del rendimiento de modelos de predicción y marcadores pronósticos: Evaluación de las predicciones y clasificaciones. *Revista Española de Cardiología*, *64*(9), 788-794. https://doi.org/10.1016/j.recesp.2011.04.017

Temesio, S., García, S., & Pérez, A. (2021). Rendimiento estudiantil en tiempo de pandemia: Percepciones sobre aspectos con mayor impacto. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*,

28, e45. https://doi.org/10.24215/18509959.28.e45

Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. https://doi.org/10.16925/9789587600490

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A., & Alvarado-Pérez, J. C. (2016). Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional. *Ediciones Universidad Cooperativa de Colombia*. https://doi.org/10.16925/9789587600490

Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., & Britos, P. (2008). Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. *VII Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*, 10.

Verma, K., Bhardwaj, S., Arya, R., Salim, M., Bhushan, M., Kumar, A., & Samant, P. (2019). Latest Tools for Data Mining and Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, *8*(9S), 18-23. https://doi.org/10.35940/ijitee.I1003.0789S19

VIII Cumbre de las Américas. (2018). *Compromiso de Lima*. http://www.summit-americas.org/LIMA_COMMITMENT/LimaCommitment_es.pdf

Villalta, C. J., Castillo, J. G., & Torres, J. A. (2016). *Violent Crime in Latin American Cities*. Inter-American Development Bank. https://doi.org/10.18235/0000428

Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, *35* (4), 644-656. https://doi.org/10.1016/j.giq.2018.10.006

XVII Conferencia Iberoamericana de Ministras y Ministros de Administración Pública y Reforma del Estado. (2016). *Carta Iberoamericana de Gobierno Abierto*. https://clad.org/wp-content/uploads/2020/07/Carta-Iberoamericana-de-Gobierno-Abierto-07-2016.pdf

Zhu, X., & Freeman, M. A. (2019). An evaluation of U.S. municipal open data portals: A user interaction framework. *Journal of the Association for Information Science and Technology*, *70*(1), 27-37. https://doi.org/10.1002/asi.24081