

Potenciales Aplicaciones de la Minería de Datos en Ecuador

Roberto Camana

Carrera Tecnológica de Análisis de Sistemas, Instituto Tecnológico Superior Aloasi, García Moreno S4-35 y Ambato (Centro Histórico), Quito, Ecuador
robertocamana@yahoo.es

Resumen. El presente artículo, tiene el propósito de describir las potenciales aplicaciones de la minería de datos en Ecuador. En la primera parte de este artículo se conocerán sus orígenes, se definirá su concepto, ejemplo y se explicaran técnicas de minería de datos (redes neuronales, regresión lineal, árboles de decisión, reglas de asociación y agrupamiento), con el propósito de conocer sus inicios, desarrollos y aproximarnos a las aplicaciones de la minería, para ello, la aplicación de una metodología, útil para el desarrollo de todo proyecto de minería de datos. En la segunda parte, se centrará en tres casos prácticos, con resultados sorprendentes y notorios: La aplicación de la minería de datos, sobre datos meteorológicos, para obtener información inferida para la elaboración de boletines meteorológicos. En las elecciones presidenciales, descubrir patrones en grandes volúmenes de datos, para realizar sondeos en elecciones presidenciales posteriores. Finalmente, en la educación para determinar interacciones de estudiantes en un entorno virtual de aprendizaje, con el propósito de conocer la influencia de las interacciones en las tareas, exámenes, recursos, situación laboral y estado civil del estudiante. Estas aplicaciones permitieron conocer de cerca sus desarrollos en el país y cuáles de ellas continúan evolucionando con la aplicación de la minería de datos.

Palabras Clave: Conocimiento, elecciones, historia, información, metodología.

1 Introducción

En los últimos años, el crecimiento del volumen de información ha rebasado nuestras expectativas por la capacidad de generar miles de datos producto de diversas transacciones o hábitos. Al comprar en el supermercado pasamos la tarjeta de afiliado, enviamos mensajes por celular o mail, navegamos por el internet, leemos un par de periódicos o compramos tickets aéreos [1]. En todos ellos quedamos registrados, de este modo conocerán con cuánta gente nos relacionamos.

Sin embargo, en el interior de estas cantidades de datos, es posible encontrar información oculta que a simple vista del ojo humano no es posible encontrar. Por ello es necesario, utilizar técnicas de recuperación de la información que consiste en la búsqueda de información útil en grandes volúmenes de datos [2]. Por lo que se requieren, de herramientas capaces de extraer información de grandes volúmenes de datos, de esta forma; nos conduce a la minería de datos.

En este contexto, la minería de datos tiene varias aplicaciones en el mundo. En Estados Unidos, el FBI, desde el 2002, el Departamento de Justicia comenzó a introducirse en enormes cantidades de datos comerciales en los que revelan hábitos y costumbre de la población, con el fin identificar a potenciales terroristas antes que

cometan un atentado [3]. Con esta aplicación, el tema de privacidad y libertad de los individuos entra en vigilancia permanente todos nuestros movimientos, pudiendo ser utilizados los datos para otros fines distintos a los que fueron diseñados.

No lejano al terrorismo, y muy cerca a hechos delictivos, se encuentra la detección de fraudes en tarjetas de crédito. En 2011, las instituciones financieras a escala mundial perdieron 2.000 millones de dólares en fraudes cometidos con tarjetas de crédito [3]. Para enfrentar el fraude, utilizaron Falcon Fraud Manager, software con capacidades de examinar transacciones, a propietarios de tarjetas y sus datos financieros, de este modo intentaron frenar fraudes millonarios en tarjetas de crédito. Este software, según sus creadores, implementaba modelos de redes neuronales, para analizar datos de pagos por medio de tarjetas de crédito.

En Sudamérica, Argentina, una de las novedosas aplicaciones de la minería de datos es la exploración y detección de patrones delictivos en la ciudad de Buenos Aires. Esta aplicación trata, de un análisis de información criminal en Argentina y comprobar su efectividad y valor agregado, en base herramientas estadísticas descriptivas básicas, considera variables y relaciones primarias [4]. Esta aplicación va a permitir optimizar el recurso humano policial y detectar grupos delictivos organizador.

En Ecuador, el tema de minería de datos, está en fase de evolución, y son pocos los estudios realizados en diferentes campos de aplicación, sin embargo existen algunas aplicaciones, que están en continuo desarrollo [5], [25].

El objetivo de esta investigación, fue conocer la aplicación de la minería de datos en Ecuador, donde se utiliza y cuáles son sus beneficios. Además, el aporte de este artículo, permitirá el desarrollo conceptual y práctico de la minería de datos.

2. Minería de datos: Orígenes y Conceptos

2.1 Orígenes de la Minería de Datos

Conocido también como Data Mining, en inglés, tiene sus orígenes desde los años 60's, considerada como una ciencia para extraer la información de grandes volúmenes de datos. Los estadísticos utilizaban términos como Data Fishing (Pesca de datos) o Data Dredging (Filtración de datos), para referirse a lo que consideraban la "mala práctica" de analizar datos sin una hipótesis a priori [6].

Desde entonces, la minería de datos ha ido evolucionando de tal forma que acuñó varias frases: La arqueología de datos, el descubrimiento y recolección de la información, la extracción del conocimiento, la minería de base de datos, entre otras frases. Para finalmente, en 1990, varios investigadores acuerdan que la base de datos, la información, extracción y conocimiento deben resumirse en la minería de datos.

En 1989, Gregory Piatetsky-Shapiro acuñó el término "Knowledge Discovery in Databases (KDD)" (Descubrimiento de Conocimiento en Bases de Datos) para el primer taller que se realizó sobre el mismo tema, y este término se hizo más famoso dentro de la comunidad científica y académica [6]. De este modo permitió identificar la exploración de datos con el conocimiento, para encontrar información útil.

Para finalmente, los términos minería de datos y KDD estén estrechamente relacionados, el primero por su conjunto de técnicas, que utiliza para la explotación de datos, y el segundo por la metodología de desarrollo aplicado, para cualquier

proyecto de minería de datos. De esta forma dio origen a varias extensiones: La minería web, minería de opiniones y la minería de textos.

2.2. Concepto y Ejemplo

Aproximándonos al concepto de minería de datos, rápidamente nos imaginamos en una enorme mina en cuyo interior se encuentran grandes volúmenes de datos, que con herramientas de pico y pala, permiten cavar en las grandes montañas de datos y remover las piedras de oro, que sería la información útil para nosotros. Además, puede definirse inicialmente la minería de datos como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias tras examinar grandes cantidades de datos [5], [7].

La acumulación masiva de datos por años, en empresas u organizaciones ha dado lugar al uso del análisis de datos, por ello; se han requerido de herramientas de alto rendimiento capaces de procesar ingentes cantidades de datos. Para llevar a cabo tal efecto se usan técnicas especializadas, identificadas como algoritmos que son quienes realizan el trabajo sobre el conjunto de datos.

Las técnicas de minería de datos, persiguen el descubrimiento automático del conocimiento contenido, en la información almacenada de modo ordenado en grandes bases de datos [7]. La entrada de la automatización en las empresas y la aparición de aplicaciones informáticas específicas (contabilidad, nómina, cuentas por pagar, entre otros), han cubierto sus necesidades básicas de la empresa, pero cada vez se exigen sistemas a más de la gestión, que permitan la toma de decisiones. Para entender mejor este último enunciado; nos centraremos en el siguiente problema:

Supongamos que un Director de marketing desea conocer que clientes abandonararán un plan de pospago de celular contratado.

Se detectaron los siguientes problemas:

- Continuos abandonos de clientes en cualquier plan contratado.
- Bajas comisiones por cliente que deciden cambiarse de plan.
- Intentos para retener a sus clientes.
- Promociones sin expectativas

Posibles soluciones al problema:

- Predecir qué clientes estarían por abandonar el plan contratado.
- Clasificar a clientes buenos y de aquellos que no presentan interés por la operadora de celular.
- Ofrecer promociones a los posibles desertores, antes que decidan irse.

El ejemplo demostrado, nos muestra una posible solución, pero ¿cómo lo hacemos? No basta con tomar las decisiones rápidas o medidas que nos conduzca a fallar, sino, necesitamos de herramientas (minería de datos), que hagan el trabajo de análisis.

2.3. Técnicas de Minería de Datos

Como hemos comentado, las técnicas de minería de datos (una etapa dentro del proceso KDD (Proceso de Extracción del Conocimiento, en inglés, Knowledge Discovery in Databases), intentan obtener patrones o modelos a partir de los datos

recopilados [8]. Las técnicas más representativas, para el desarrollo de cualquier proyecto de minería de datos son:

- **Redes neuronales:** Son capaces de detectar y aprender patrones y características de los datos, además, una vez adiestradas las redes puede hacer previsiones, clasificaciones y segmentación [9]. Se comportan de forma parecida a nuestro cerebro aprendiendo de la experiencia y del pasado, y aplicando tal conocimiento a la resolución de problemas nuevos. Este aprendizaje se obtiene como resultado del adiestramiento ("training") y éste permite la sencillez, la potente adaptación y evolución, ante una realidad cambiante muy dinámica [15]. Por ejemplo, se utiliza para el reconocimiento de imágenes, búsqueda de información en la web, entre otras.

- **Regresión lineal:** Es la más utilizada para formar relaciones entre datos, es decir; más de dos variables, además permite identificar relaciones entre variables numéricas y construir modelos de regresión: Una variable de salida y múltiples entradas numéricas [9], [15]. Es decir, la regresión lineal, se modelan los datos usando una línea recta.

En el ejemplo, de la Fig. 1, muestra 35 marcas de cerveza, se estudia la relación entre el grado de alcohol de las cervezas y su contenido calórico, y se representa un pequeño conjunto de datos.

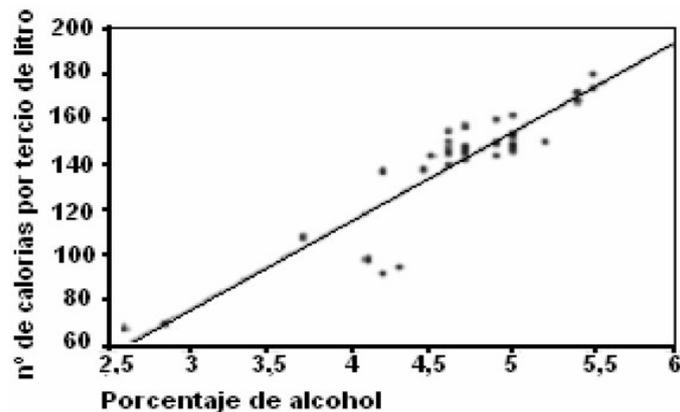


Figura 1. Regresión lineal simple

- **Árboles de decisión:** Son herramientas analíticas empleadas para el descubrimiento de reglas y relaciones, están dadas en un diagrama de flujo, con estructura de árbol, en donde los nodos internos representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones, y los "nodos hoja" representan las clases [9], [12]. Los árboles de decisión, parte de la inteligencia artificial es utilizado para la predicción, por medio de series condiciones sucesivas (reglas). Un típico ejemplo, de un árbol de decisión se muestra en la Fig. 2, ahí se representa el concepto de comprar computadoras, esto es, el modelo predice si un cliente compra o no una computadora en CompuRob. En la figura los nodos internos

están denotados por rectángulos, mientras que los “nodos hoja” están denotados por óvalos [13].

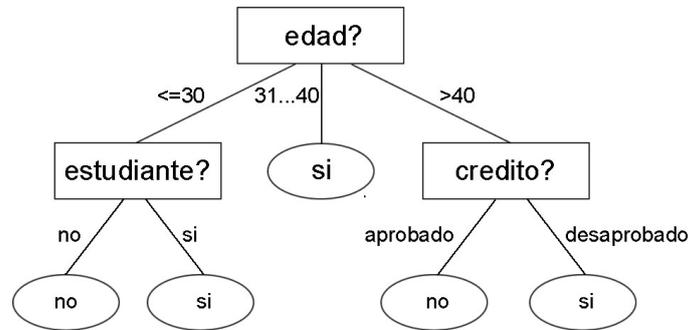


Figura 2. Árbol de decisión para la compra de computadoras.

- **Reglas de Asociación:** Se evalúan de acuerdo al soporte y a la confianza de los datos, se utilizan para encontrar hechos que ocurren en común dentro de un conjunto de datos. Dicho de otra manera deben ocurrir ciertas condiciones para que se produzca cierta condición, también para buscar por medio de conjunto de datos, reglas que revelan la naturaleza de las relaciones o asociaciones entre datos de las entidades [17], [18].

Se aplican en el análisis de la canasta de mercado, marketing cruzado con correo, diseño de catálogos, segmentación de clientes respecto a las compras y el soporte para la toma de decisiones [17].

- **Agrupamiento (Clustering):** Se basa en intentar responder, como es que ciertos objetos (casos) pertenecen o “caen” naturalmente en cierto número de clases o grupos, de tal manera que estos objetos comparten ciertas características [10].

Algunos ejemplos de la aplicación de las técnicas de clustering, en el contexto de la extracción del conocimiento, incluyen la identificación de subgrupos homogéneos de clientes en bases de datos de marketing, así como la identificación de subcategorías de medidas del espectro infrarrojo del cielo [14]. Es decir, la similitud puede medirse mediante funciones de distancia especificadas por los usuarios o por expertos, como se muestra en la Fig. 3.

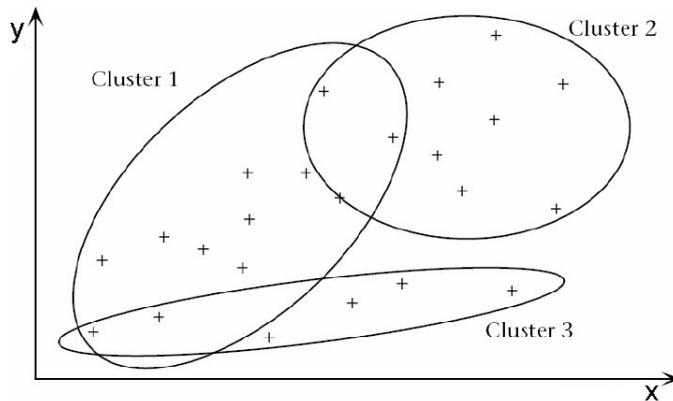


Figura 3. Clustering de datos bancarios.

3. Metodología para la Creación de un Proyecto de Minería de Datos

Para llevar a cabo el proceso de minería de datos, se necesita de fases, que permitan descubrir patrones interesantes y potencialmente útiles de información. Por ello existe la extracción del conocimiento, está principalmente relacionado con el proceso de descubrimiento conocido como Knowledge Discovery in Databases (KDD), se refiere al proceso no-trivial de descubrir conocimiento e información, potencialmente útil dentro de los datos contenidos en algún repositorio de información [11]. En la Fig. 4, se muestran las etapas del KDD:



Figura 4. Proceso KDD.

A continuación, se explica cada una de las etapas del proceso KDD [5], [19].

A. Selección de datos

Recopilación de datos relevantes y tipo de información obtenida de diferentes fuentes, para ser utilizadas en el preprocesado.

B. Preprocesado

Consiste en una exploración, ya que al venir de diferentes fuentes de datos, es necesario una limpieza, es decir; eliminar o corregir datos incorrectos, necesaria para la siguiente etapa.

C. Transformación

Esta etapa consiste en la transformación de los datos, es decir, la creación de nuevas variables a partir de las ya existentes y la normalización de datos preparados para su posterior análisis.

D. Minería de datos

Consiste en la búsqueda de patrones de interés, en una determinada forma de representación, en función al problema a solucionar.

E. Interpretación y evaluación

Se evalúan patrones que serán analizados por expertos, y si es necesario se vuelve a las fases anteriores para una nueva iteración.

4. Aplicaciones y resultados de la minería de datos en Ecuador

El rápido crecimiento de datos y las necesidades de convertir en información útil, ha permitido a centros de investigación y universidades de Ecuador, utilicen datos históricos almacenados, y que estos aporten al conocimiento, en la toma de decisiones.

Podemos formular, como hipótesis, el siguiente razonamiento: ¿Por qué la minería de datos en el Ecuador, se ha aplicado en unos campos, más que otros?. Se establecieron posibles causas que expliquen este fenómeno estudiado:

- Intereses de investigadores, para aplicar la minería de datos, en campos poco o nada explotados.
- Saturación de investigaciones, en un mismo campo de estudio.
- Aparición de nuevos campos de aplicaciones de la minería de datos.
- Investigaciones previas en un mismo campo, sin continuos procesos de investigación.

Para confirmar experimentalmente, la utilidad de la minería de datos, se puede dar dentro de los siguientes aspectos:

- **Sistemas parcialmente desconocidos:** Si el modelo del sistema que produce los datos, es bien conocido, entonces no necesitamos de la minería de datos ya que todas las variables son de alguna manera predecibles.
- **Enorme cantidad de datos:** Al contar con mucha información, en algunas bases de datos es importante para una empresa encontrar la forma de analizar "montañas" de información (lo que para un humano sería imposible) y que ello le produzca algún tipo de beneficio.
- **Potente hardware y software:** Muchas de las herramientas presentes en la minería de datos están basadas en el uso intensivo de la computación, un software eficiente, aumentará el desempeño del proceso de buscar y analizar información, algo humanamente imposible.

Con estas consideraciones, se seleccionaron los siguientes campos de aplicación de la minería de datos en Ecuador:

A. Meteorología

- **Aplicación de la Minería de Datos para Establecer Patrones de Comportamiento de Datos Meteorológicos:** Consiste en el análisis de la información meteorológica recopilada desde el año 1995 hasta el 2005, usando para ello los datos meteorológicos que posee el Observatorio Astronómico de Quito (OAQ) e implementa un Data Warehouse [23]. De este modo, por medio de la herramienta de minería de datos, permiten obtener patrones de comportamiento meteorológico. En la Fig. 5, se ilustra uno de los resultados:

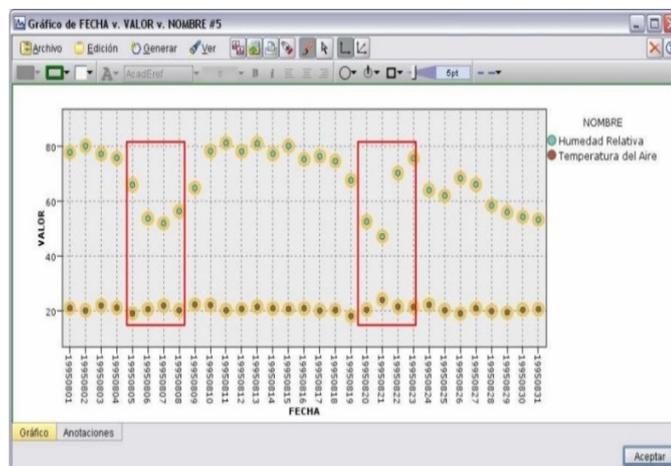


Figura 5. Temperatura y humedad.

En la Fig. 5, se puede apreciar que la humedad relativa disminuye cuando la temperatura aumenta, comprobando de esta forma un patrón de comportamiento [23].

Para poder realizar este estudio, utilizaron la temperatura máxima, media, y mínima y el modelo CHAID. Este realiza una detección automática de interacciones mediante el chi-cuadrado, es un método de clasificación para generar árboles de decisión mediante estadísticos de chi-cuadrado.

Los beneficios de esta aplicación radican en la construcción de modelos, con la finalidad de obtener patrones de comportamiento, en base a grandes cantidades de información, que las mismas permitieron ser usadas para la toma de decisiones.

B. Elecciones presidenciales

- **Selección de Juntas Receptoras del Voto más Representativas:** Las elecciones en Ecuador, ha marcado un hito histórico, por el incremento de los datos electorales presidenciales cada cuatro años, debido a estas montañas de datos, permitieron encontrar patrones de comportamiento o como los datos pueden incidir en una determinada elección presidencial. El objetivo de esta investigación, fue utilizar herramientas de minería de datos, para la selección de las juntas receptoras del voto (JRV) más representativas, a nivel nacional durante las elecciones presidenciales de 2009 [19], [25], [26].

Para la selección de juntas receptoras del voto (JRV) más representativas, utilizaron la técnica de agrupamiento (en inglés, clustering) [19], [22]. La distribución de las juntas receptoras del voto en función de la Distancia Euclídea, puede observarse en la Fig. 6.

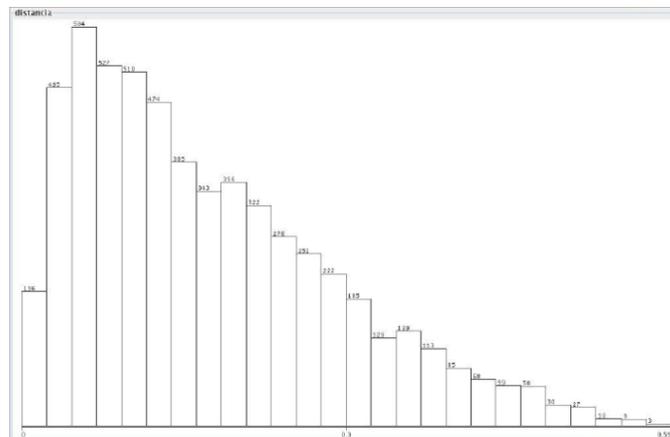


Figura 6. Distribución de las distancias del resultado de cada una de las JRV.

En la Fig. 6, el algoritmo de clustering utilizado fue el K-MEDIAS, con el objetivo de seleccionar las Juntas Receptoras del Voto más parecidas al resultado final. Cuanto mayor sea el número de grupos elegidos, menor número de muestras aparecerán en cada uno de ellas. Se eligieron 20 clusters, mostrados en la Tabla 1:

Tabla 1. Clusters seleccionados.

Cluster	%	Cluster	%
0	63 (1%)	1	429 (7%)
2	362 (6%)	3	409 (7%)
4	159 (3%)	5	380 (6%)
6	311 (5%)	7	339 (6%)
8	236 (4%)	9	232 (4%)
10	382 (7%)	11	304 (5%)
12	352 (6%)	13	190 (3%)
14	262 (4%)	15	133 (2%)
16	318 (5%)	17	350 (6%)
18	366 (6%)	19	287 (5%)

En el cluster 13 se ubican aquellas Juntas Receptoras del Voto con resultados más cercanos en la Distancia Euclídea. Este cluster agrupa un 3% del total, es decir 190 JRV, que podrían ser útiles, para efectuar cualquier clase de sondeo [19].

Este análisis es muy importante, ya que brinda una clara señal para realizar sondeos a “pie de urna” o “boca de urna” de forma fiable, reduciendo el costo de sondeos tradicionales.

- **Distribución del Voto:** Para realizar el análisis de distribución del voto, se utilizó un árbol de decisión DecisionStrump, obteniendo una sola entrada: el porcentaje de voto [19]. Los árboles de decisión son muy intuitivos y fáciles de comprender. Este estudio se realizó con datos de electores de las provincias de la Zona N°3 (Chimborazo, Cotopaxi, Pastaza y Tungurahua. Según se observa en la Fig. 7.

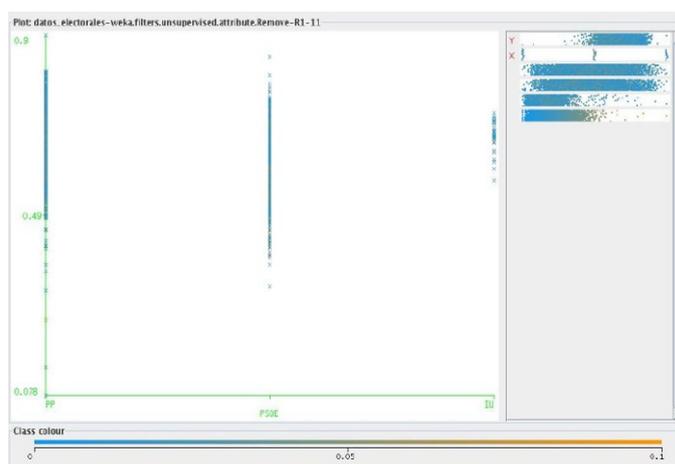


Figura 7. Distribución de las juntas ganadoras en función de porcentaje de votos de cada una.

En la Fig. 7, se puede comprobar que el Movimiento Alianza País necesita un porcentaje algo mayor en una junta para proclamarse ganador en la misma. Además, parece el que gane una u otra opción política, puede estar relacionado con los ciudadanos que votan en dicha junta. Dicha información puede ser utilizada por el partido político al cual perjudique la abstención. Como era de esperarse, las Juntas Receptoras del Voto más representativas, están centradas en las provincias, cantón, parroquia mostradas en la Tabla 2:

Tabla 2. Provincias con JRV mayormente representativas.

Provincia	Cantón	Parroquia	JRV	Sufrag.
Chimborazo	Riobamba	Lizarzaburo	171	30.659
Chimborazo	Riobamba	Velasco	121	22.437
Tungurahua	Ambato	La Matriz	117	19.907
Cotopaxi	Pujili	Pujili	104	20.884
Pastaza	Pastaza	Puyo	87	20.883

Como se puede observar en la Tabla 2, en una sola provincia, Cotopaxi, la cabecera cantonal (capital) no tiene mayor representación de JRV. Además estos resultados despiertan gran interés de empresas especializadas en sondeos a “boca de urna”, que si bien logra ser una estrategia para conocer resultados parciales, apunta obtener de aquellas JRV mayoritarias de forma específica una buena cantidad de información.

La distribución del voto, es la más importante en toda elección, principalmente presidencia, ya que estos resultados servirán de base para las próximas elecciones, es decir conociendo los puntos de concentración de los electores por algún movimiento o partido político, determinará efectuar campañas con mayor éxito, sobre este electorado que será decisivo en las elecciones.

C. Educación

- **Entornos virtuales de aprendizaje:** En la Universidad Nacional de Loja, se realizó un estudio, para determinar las interacciones de sus estudiantes del curso virtual del idioma Inglés, de la modalidad de Estudios a Distancia (MED), correspondiente al periodo académico 2013-2014 [16]. La educación en entornos virtuales, en todo el mundo, se ha tornado en una fuente de generación de grandes volúmenes de información, sin ser analizadas, por este motivo, se necesita de herramientas capaces de descubrir información útil y comprensible.

Para este estudio, utilizaron la minería de datos, con su técnica de clasificación supervisada, para realizar predicciones [20]. De acuerdo con las características del proyecto, la técnica de clasificación pudo generar modelos a través de algoritmos. Este estudio consistió en un análisis comparativo tomando en cuenta características de las herramientas de minería de datos, mediante algunos algoritmos conjuntamente con los datos de los estudiantes, dividieron en dos conjuntos, para entrenamiento y validación, siendo el algoritmo Decisión Tree, el de mejor resultado [21].

Este clasificó las instancias correctamente con un margen de error mínimo, así mismo presentó un árbol con los diferentes atributos dando las mejores reglas de las interacciones de los estudiantes, de tal forma pudieron generar el modelo mediante el cual se determinó que en gran mayoría los estudiantes tienen un nivel de interacción

medio en el curso virtual de inglés, donde los factores que más influyen son las interacciones en las tareas, exámenes, recursos, situación laboral y estado civil del estudiante [16], según se observa en la Tabla 3.

Tabla 3. Porcentaje de los factores, atributos con sus respectivos pesos.

Atributo	Porcentaje del atributo (%)
Interacciones tareas	12.196
Interacciones recurso	10.946
Interacciones exámenes	13.299
Genero	4.562
Estado civil	9.509
Edad	8.671
Servicios	8.299
Carrera	9.137
Número hijos	5.346
Trabajo	8.126
Ciudad	9.908

Luego de obtener el peso de cada uno de los atributos pertenecientes a los datos de los estudiantes, los que más inciden en el objetivo principal del presente proyecto es determinar el nivel de interacción de los estudiantes del curso virtual inglés de la MED, son intercepciones en las tareas con un 12%, en los recursos con el 11%, y en los exámenes el 13%.

Para determinar el nivel de interacción en el curso de inglés se aplicó diferentes algoritmos de clasificación, presentando los mejores resultados el Decision Tree, ya que este algoritmo obtuvo el menor margen de error durante la clasificación de los datos de las interacciones en el curso (tareas, exámenes, recursos), datos personales, institucionales y socioeconómicos [16], [21].

El beneficio de esta investigación, fueron que las interacciones de los estudiantes, en un nivel intermedio, en el curso virtual, influyeron las interacciones en las tareas, exámenes, recursos, situación laboral y estado civil del estudiante.

5. Conclusiones

La capacidad para almacenar datos ha crecido en los últimos años a velocidades exponenciales. En el otro extremo, nuestra capacidad para procesar esta enorme cantidad de datos para utilizarlos eficazmente no ha ido a la par. Por este motivo, la minería de datos se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido de grandes volúmenes de datos. Descubrir nuevos caminos que nos ayuden en la identificación de interesantes estructuras en los datos es una de las tareas fundamentales en la minería de datos.

En el ámbito del comportamiento climatológico de la ciudad de Quito, resulta interesante encontrar patrones ocultos e identificar información almacenada en base de datos dimensionales o datawarehouse. De esta forma, podría obtener información, para las personas interesadas en el tema, asimismo esta información podría ser utilizada para la elaboración de informes meteorológicos.

El análisis realizado para los datos electorales puede ser útil, primero para realizar sondeos en elecciones posteriores, donde se puede tener la certeza de gozar de cierta fiabilidad y por concentración un ahorro en costos. Los resultados obtenidos en esta investigación son confiables y para futuras elecciones servirán de base los algoritmos utilizados. Como ya existe una base de datos el cual sirvió para la explotación, solo restaría agregar datos y variables de las elecciones presidenciales futuras.

La minería de datos, es muy importante dentro del campo de la educación ya que ayudó a extraer información que se encuentra oculta en los datos de tal forma permitió el análisis y la generación de nuevo conocimiento para poder determinar en nivel de interacción de los estudiantes.

Agradecimientos

El presente artículo forma parte de investigaciones, concernientes a las elecciones presidenciales, y aplicación de la minería de datos y visualización de la información, para el análisis y modelamiento de la información electoral en Ecuador.

El autor expresa un agradecimiento sincero, a la carrera de Tecnología en Análisis de Sistemas del Instituto Tecnológico Superior “Aloasi”, por el interés de estas investigaciones.

Referencias

1. Camana, R. 2012. Digitalizando Nuestra Vida Cotidiana. Generar Conocimiento el Pensar y Hacer; Un Aporte a la Educación para la Ciencia y la Tecnología. Ambato: Corporación Editora Nacional, p. 73-74.
2. Galán, M. 2009. Minería de Datos para Recuperación de la Información. (En línea). (Consultado: 18 de septiembre de 2015). Disponible en: <http://www.oocities.org/es/mineria.datos/>
3. Virseda, F. & Ramón, J. 2011. Minería de Datos y Aplicaciones. (En línea). (Consultado el: 24 de septiembre de 2015). Disponible en: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>
4. Perversi, I. 2007. Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina. Tesis de Ingeniería Industrial. Instituto Tecnológico de Buenos Aires. Argentina.
5. Camana, R. 2014. Una Experiencia Personal: Pico y Pala en la Exploración y Visualización de Datos Electorales. Tecnológica ESPOL –RTE, 27: 1-13.
6. Webmining. 2011. ¿Qué es Data Mining?. (En línea). (Consultado: 24 de septiembre de 2015). Disponible en: <http://www.webmining.cl/2011/04/que-es-data-mining/>
7. Pérez, C.; Santín, D. 2008. Minería de Datos; Técnicas y Herramientas. Madrid, Thomson Ediciones Paraninfo. 774 p.
8. Molina, J. & García, J. 2013. Capítulo 3: Técnicas de Minería de Datos Basadas en Aprendizaje Automático. (En línea). (Consultado: 28 de septiembre de 2015). Disponible en: <http://santiagozapatakdd.files.wordpress.com/2011/03/curso-kdd-full-cap-3.pdf>
9. Aguilar, J. & Estrada, C. 2012. Minería de Datos (Data Mining). (En Línea). (Consultado: 20 de octubre de 2015). Disponible en: <http://www.slideshare.net/miriam1785/mineria-de-datos-8768313>

10. Fernández, H. 2013. Clustering (Agrupamiento). (En línea). (Consultado: 15 de octubre de 2015). Disponible en: <http://www.uccor.edu.ar/paginas/seminarios/Cursos/DM-Medicine/Clustering.ppt>
11. Han, J. & Kamber M. 2006. Data Mining: Concepts and Techniques. San Francisco, Morgan Kaufmann Publishers, 743 p.
12. Macías, M. 2008. Técnicas de Minería de Datos para la Retención de Clientes en el Sector Asegurador. "Trabajo Presentado para el XV Premio de Investigación sobre Seguros y Finanzas 2008," pp. 33.
13. Han, J. & Kamber, M. 2000. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers.
14. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine, 17 (3), 37-54.
15. Molina, J. & García, J. 2006. Técnicas de Análisis de Datos: Aplicaciones Prácticas Utilizando Microsoft Excel y Weka. pp. 153-154.
16. Jaramillo, A. & Paz, H. 2015. Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. Tecnológica ESPOL –RTE, 28: 64-90
17. García, J. & Acevedo A. Análisis para Predicción de Ventas Utilizando Minería de Datos en Almacenes de Ventas de Grandes Superficies, Universidad Tecnológica de Pereira - Facultad de ingenierías: eléctrica, electrónica, física y Ciencias de la computación - Ingeniería de sistemas y computación. (En línea). (Consultado: 18 de octubre de 2015). Disponible en: <http://repositorio.utp.edu.co/dspace/bitstream/11059/1339/1/006312G216.pdf>
18. Velandia, R. & Hernández, F. Evaluación de Algoritmos de extracción de reglas de decisión para el diagnóstico de huecos de tensión, Universidad Industrial de Santander. (En línea). (Consultado: 18 de octubre de 2015). Disponible en: <http://tangara.uis.edu.co/biblioweb/tesis/2010/134742.pdf>
19. Camana, R. 2012. Aplicación de Técnicas de Minería de Datos para la Indagación y Estudio de Resultados Electorales. CienciAmerica, 1(1): p. 85-94.
20. Moreno, M. & López, V. 2015. Uso de Técnicas No Supervisadas en la Construcción de Modelos de Clasificación de Ingeniería de Software. Departamento de Informática y Automática. Universidad de Salamanca.
21. Ordoñez, K. Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes de primer ciclo de la Modalidad Abierta y a Distancia de la UTPL, Universidad Técnica Particular de Loja – Area Técnica. (En línea). (Consultado: 20 de octubre de 2015). Disponible en: <http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-%20Informatica.pdf>
22. Jiawei, M. 2006. Data Mining: Concept and Techniques. California, Estados Unidos: Morgan Kaufmann.
23. Amores, B. & Cruz, D. 2008. Aplicación de Data Mining para Establecer Patrones de Comportamiento de Datos Meteorológicos para la Ciudad de Quito. Facultad de Ingeniería en Sistemas, Escuela Politécnica Nacional, Quito.
24. Camana, R. 2013. Visualización de Datos Electorales Presidenciales: Herramientas para la Visualización de Grandes Volúmenes de Datos. CienciAmerica, 2, 24-32.
25. Camana, R. 2014. Las Bases Electorales. (En línea). (Consultado: 23 de noviembre de 2015). Disponible en: <https://robertocamana.wordpress.com/2015/02/18/las-bases-de-datos-electorales/>
26. Camana, R. 2015. Acceso Libre a Base de Datos. (En línea). (Consultado: 23 de noviembre de 2015). Disponible en: <https://robertocamana.wordpress.com/2015/11/10/acceso-libre-base-datos-elec/>