



Aplicación de técnicas de aprendizaje automático para disminuir las observaciones en las lecturas de medidores

Application of machine learning techniques to reduce observations in meter readings

Aplicação de técnicas de aprendizado de máquina para reduzir observações em leituras de medidores

Gustavo Adolfo Noboa Franco ^I
gnoboa@mgs.ecotec.edu.ec
<https://orcid.org/0000-0002-6907-4863>

Carlos Alex Valle Chiriboga ^{II}
carlitos.alex.valle@gmail.com
<https://orcid.org/0000-0003-0090-378X>

Jorge Eduardo Cevallos Zhunio ^{III}
jorge.cevallosz@ug.edu.ec
<https://orcid.org/0000-0001-8976-2973>

Correspondencia: gnoboa@mgs.ecotec.edu.ec

Ciencias de la Educación
Artículo de Investigación

* **Recibido:** 23 de mayo de 2022 * **Aceptado:** 12 de junio de 2022 * **Publicado:** 11 de julio de 2022

- I. Magíster en Sistemas de Información mención en Inteligencia de Negocios, Universidad de Guayaquil, Guayaquil, Ecuador.
- II. Magíster en Sistemas de Información mención en Inteligencia de Negocios, Tecnológico ARGOS, Guayaquil, Ecuador.
- III. Magíster en Sistemas de Información mención en Inteligencia de Negocios, Universidad de Guayaquil, Guayaquil, Ecuador.

Resumen

El siguiente trabajo consiste, en la búsqueda de solucionar inconveniente en la hora de tomar decisiones con las observaciones de lectura de los equipos de medición mediante técnicas de minería de datos. En la actualidad las técnicas de aprendizaje automático de diferentes algoritmos son herramientas muy utilizadas para la toma de decisiones en diferentes modelos de negocios.

Para este tipo de modelo de negocio se procede a utilizar las técnicas de algoritmo de Regresión Lineal de acuerdo al análisis de los objetivos del negocio. Estas técnicas se basan en tener datos cuantitativos para poder evaluar diferentes algoritmos supervisados, para el entrenamiento y evaluación de los modelos a utilizar, para con esto evaluar de acuerdo al error aceptable y el error cuadrático de menor impacto para la elección del modelo de las técnicas de aprendizaje automático. En este proyecto se utilizó algoritmos para el aprendizaje supervisado como el de regresión lineal, el cual inicialmente se utilizó con todas las variables, y posterior fue reducido manualmente, también se utilizó la función STEP y RandomForest. Se investigó varias metodologías de minería de datos y la cual se las alineó de acuerdo al modelo de negocio es CRISP-DM.

Se generó un dashboard con la herramienta R Studio para visualizar el algoritmo seleccionado del error más bajo del algoritmo de STEP, donde se podrá observar los sectores con mayor cantidad de observaciones de lectura de los equipos de medición y tomar acciones en función de cada sector para disminuir las observaciones.

Palabras Clave: Aprendizaje automático; regulación energética; algoritmo STEP; CRISP-DM.

Abstract

The following work consists in the search to solve a problem when making decisions with the reading observations of the measurement equipment through data mining techniques. At present, machine learning techniques of different algorithms are widely used tools for decision making in different business models.

For this type of business model, Linear Regression algorithm techniques are used according to the analysis of the business objectives. These techniques are based on having quantitative data to be able to evaluate different supervised algorithms, for the training and evaluation of the models to be used, in order to evaluate according to the acceptable error and the squared error of least impact for the choice of the model of the techniques. machine learning.

In this project, algorithms for supervised learning were used, such as linear regression, which was initially used with all the variables, and later was reduced manually, the STEP and RandomForest functions were also used. Several data mining methodologies were investigated and the one that was aligned according to the business model is CRISP-DM.

A dashboard was generated with the R Studio tool to visualize the selected algorithm of the lowest error of the STEP algorithm, where it will be possible to observe the sectors with the greatest number of reading observations of the measurement equipment and take actions based on each sector to decrease observations.

Keywords: machine learning; energy regulation; STEP algorithm; CRISP-DM.

Resumo

O trabalho a seguir consiste na busca de resolver um problema ao tomar decisões com as observações de leitura do equipamento de medição através de técnicas de mineração de dados. Atualmente, técnicas de aprendizado de máquina de diferentes algoritmos são ferramentas amplamente utilizadas para tomada de decisão em diferentes modelos de negócios.

Para este tipo de modelo de negócio, são utilizadas técnicas de algoritmos de Regressão Linear de acordo com a análise dos objetivos de negócio. Essas técnicas baseiam-se em ter dados quantitativos para poder avaliar diferentes algoritmos supervisionados, para o treinamento e avaliação dos modelos a serem utilizados, a fim de avaliar de acordo com o erro aceitável e o erro quadrado de menor impacto para a escolha do modelo das técnicas, aprendizado de máquina.

Neste projeto, foram utilizados algoritmos de aprendizado supervisionado, como a regressão linear, que inicialmente foi utilizada com todas as variáveis, e posteriormente foi reduzida manualmente, também foram utilizadas as funções STEP e RandomForest. Diversas metodologias de mineração de dados foram investigadas e a que se alinhou de acordo com o modelo de negócio é a CRISP-DM.

Foi gerado um dashboard com a ferramenta R Studio para visualizar o algoritmo selecionado de menor erro do algoritmo STEP, onde será possível observar os setores com maior número de observações de leitura do equipamento de medição e realizar ações com base em cada setor para diminuir as observações.

Palavras-chave: aprendizado de máquina; regulação energética; algoritmo PASSO; CRISP-DM.

Introducción

La evolución de los sistemas informáticos y con el objetivo de brindar un mejor servicio a los clientes, se ha ido optimizando los sistemas de la toma de lectura, con la finalidad de disminuir la cantidad de reclamos que llegan al área de atención al cliente. Con la información que se cuenta de las observaciones de los medidores eléctricos mensualmente tales como: medidores dañados, medidores altos, tapas sucias, medidores no localizados, no se toman decisiones para los correctivos, donde se pueda disminuir el índice, por la falta de conocimiento de herramientas de inteligencia de negocio que brinda facilidades para el análisis de los correctivos correspondientes. La Regulación energética se constituye en un elemento del Estado para garantizar la eficiencia energética de los clientes, principalmente de la parte económica al ser una empresa estratégica, donde las fuerza están constituida en brindar un servicio de calidad a los Ecuatorianos (Pereira, 2015). Las empresas encargadas de la distribución de energía eléctrica son reguladas por la ley Orgánica de Servicio Público de Energía Eléctrica, y supervisadas o controladas por la Agencia de Regulación y Control de Electricidad (ARCONEL), que se encuentran amparada por la parte jurídica en el artículo 14, con las actividades que debe cumplir del Servicio público de energía eléctrica (ARCONEL, 2018).

La empresa reguladora de energía eléctrica realiza la contratación de una compañía de toma de lectura de los equipos de medición para brindar un mejor servicio de energía eléctrica a los clientes, con esto tener una información correcta del área analizar y tomar decisiones para mejorar los indicadores de servicio de la empresa reguladora de energía (CNEL, 2018).

El problema que existe con las observaciones de medidores eléctricos, radica que no existe un tratamiento de los datos, no se realiza seguimiento a la información. Con la aplicación de técnicas de limpieza, procesamiento y transformación de datos, se puede obtener una visualización y clasificación de las observaciones de lecturas en los medidores eléctricos, permitiendo entregar la información a los responsables de la toma de decisiones, para el adecuado seguimiento y solución de los problemas, optimizando el tiempo y calidad en el servicio. (Uselli, 2014). Cada mes se tiene medidores que deben ser reemplazados, reubicados, o medidores no encontrados en la toma de lectura, por lo que se debe realizar un trato a la información, analizar algoritmos, que permitan aportar a la toma de decisiones y correcciones de las observaciones.

Por lo tanto, se propone utilizar técnicas de aprendizaje automático para disminuir la cantidad de observaciones generadas en la lectura de los medidores eléctricos, el aprendizaje automático es una

rama de la inteligencia artificial que se encarga de componer algoritmos que tienen la posibilidad de aprender, sin tener que programarlos de manera explícita, lo que se debe hacer es mantener el algoritmo con una magnitud gigantesca de datos para que el algoritmo aprenda y sepa qué formar en conjunto de los casos expuesto (Sandoval, 2018). El aprendizaje automático consiste en utilizar herramientas donde permitan el procesamiento de los datos, tratamiento de la información para aplicar modelos de algoritmos matemáticos, para realizar tareas de identificación de patrones y análisis de los datos (Mueller & Massaron, 2016).

Este es un proyecto de aprendizaje automático y no de inteligencia de negocios, considerando que la inteligencia de negocios es de naturaleza retrospectiva, mientras que el Aprendizaje automático se trata del futuro, de predecir lo que sucederá (Labs, 2019). En esta investigación se va a realizar un análisis predictivo de las observaciones de los medidores, donde se va a evidenciar el comportamiento de los patrones a evaluar en el modelo, utilizando herramientas de inteligencia de negocios como es R studio que es un lenguaje de programación de código abierto, dedicado a realizar algoritmos de aprendizaje automáticos, estadística descriptiva y gráficos, contiene consolas para editar sintaxis, también herramienta para realizar trazos, mantener histórico, depuración y gestionar los proyectos de estudio (Pimpler, 2018). La herramienta R es un descubrimiento para el análisis datos, por las bondades que permiten aplicar los conocimientos de la ciencia de los datos a través de un software libre de código abierto que permite sistemáticamente analizar la información y vincularse con el aprendizaje automático para impartir el conocimiento de los datos (Williams, 2017).

Existen 3 tipos de modelos de machine learning: modelos lineales, modelos de árbol y modelos neuronales (Castillo Rojas, Medina Quispe, & Vega Damke, 2017). En este trabajo se utilizó el modelo lineal que según Sandoval (2018) estos tratan de hallar un trazo que se ajuste bien a los puntos que se disponen, estos destacan a partir de modelos conocidos y usados como la regresión lineal, la logística y los dos modelos tienen el inconveniente del que se ajustan demasiado los datos disponibles, con el peligro que esto tiene para nuevos datos que pudieran alcanzar y son modelos relativamente simples; el otro modelo utilizado es el de árbol que son modelos precisos, estables y sencillos de interpretar básicamente se construyen unas reglas de resolución que se pueden constituir como un árbol ya que, a oposición de los modelos lineales, pueden imitar relaciones no lineales para resolver problemas de modelos, destacan los árboles de resultados y los random forest.

Al ser precisos y elaborados, obviamente ganamos en capacidad predictiva y perdemos en beneficio (Jaramillo & Paz Arias, 2015)

La investigación incluyó técnicas y procedimientos de minería de datos que es la obtención de datos implícitos, desconocidos o anticipadamente ignorados, que puede ser potencialmente válidos. Se puede pensar a la minería de datos como una colección de diferentes técnicas que sirven para incitar la comprensión e información de un modo estructurado de un gran conjunto de datos. A la minería de datos se la conoce con sus siglas en inglés como Data Mining y además se la relaciona con la revelación de la comprensión de las bases de datos conocidas como Knowledge Data Discovery (KDD) (Camana, 2016). Adicional se utilizó Shiny para realizar el dashboard, este paquete de R Studio, está diseñado para facilitar la formación de aplicaciones interactivas de aplicaciones web. Proporciona un enlace instantáneo "reactivar" entre entradas y salidas: cuando el interesado visualiza el modelo encontrado mediante el aplicativo ejecutado (Horton & Kleinman, 2015).

La información será utilizada por el área comercial como soporte a toma de decisiones relacionadas a las correcciones a ejecutar, para así contribuir con la disminución de las observaciones de lectura evitando consumos generados, clientes no localizados, medidores reubicados y la facturación se realice acorde a observado en el medidor. Esto se traduce en beneficios para la empresa, porque se disminuiría las pérdidas energéticas y por ende económicas, y reduciría los porcentajes de las observaciones; brindando un servicio de calidad a los clientes, para que exista conformidad del servicio ofrecido para la empresa reguladora de energía eléctrica, lo cual concuerda con Zielesny (2016) que menciona que a través del análisis predictivo de los datos y aplicación de un modelo de regresión lineal se facilita el análisis de la información contribuyendo con la toma de decisiones y seguimiento en los departamentos pertinentes para su adecuada solución, disminuyendo así el porcentaje de observaciones reportadas.

Métodos o metodología

“La metodología se refiere a la serie de métodos y técnicas de rigor científico que se aplican sistemáticamente durante un proceso de investigación para alcanzar un resultado teóricamente válido” (Vera et al., 2015). El presente trabajo de desarrollo tecnológico que concluye con la aplicación de técnicas de aprendizaje automático a través de un modelo regresión lineal con un análisis de datos predictivos, para identificar patrones en las observaciones de lecturas en los

medidores eléctricos ,donde a partir de características como código de identificación del sector, marca del medidor, fecha de la observación reportada, consumo de energía levantado, y tarifa se identifiquen comportamientos que contribuyan de manera efectiva a la toma de decisiones en la ciudad de Babahoyo.

En el proyecto se utilizó metodologías de minería de datos y técnicas de aprendizaje automático, además de realizar el tratamiento, limpieza, modelamiento de los datos en el entrenamiento en el cual se evalúan diferentes algoritmos, a fin de seleccionar aquel que cumpla con los parámetros para aplicar un modelo predictivo en la herramienta estadística R Studio (Gandrud, 2015), donde se realice el análisis de las observaciones de lecturas, identificando los factores que inciden en los sectores de la ciudad de Babahoyo, ayudando a la toma de decisiones (Karthik Ramasubramanian, 2017).

El proyecto de desarrollo busca determinar un modelo predictivo de los datos utilizando técnicas de aprendizaje automático, para lo cual a través de técnicas de minería de datos se identificará un algoritmo que aporte con una solución a las observaciones de lectura de los equipos de medición para la toma de decisiones. El enfoque de la investigación es cuantitativo, que de acuerdo con Fernández Collado, Hernandez Sampieri, & Baptista Lucio (2017), representa un conjunto de procesos, es secuencial y probatorio. Se utilizan datos basados en la medición numérica y el análisis es estadístico, con el fin de establecer pautas de comportamiento o probar teorías.

El alcance es exploratorio, descriptivo y explicativo. Los estudios exploratorios tienen como objetivo familiarizarse con un tema desconocido o poco estudiado (Fernandez Collado, Hernandez Sampieri, & Baptista Lucio, 2017), como es el caso del análisis predictivo utilizando técnicas de aprendizaje automático que es un área que recién se está explotando. Los estudios descriptivos tienen como detallar las propiedades de la forma o fenómeno que se va a aprender y facilitar una perspectiva del crecimiento puntual permitido. Es inevitable, por la proporción, elegir los rasgos o conceptos del fenómeno y determinarlos en conjunto de modo independiente, con gran exactitud (Diaz Narvaez & Calzadilla Nunez, 2016).

El tipo de investigación explicativa intenta recorrer más al detalle del tipo de investigación exploratoria y descriptiva para detectar las causas reales de un inconveniente. En estos casos, un investigador explicativo estaría ambicioso en las razones posteriormente de estos hechos (Begueri & Malberti, 2017).

La metodología de desarrollo a utilizarse es CRISP-DM (Cross-Industry Standard Process for Data Mining), que incluye fases normales de un proyecto y cumple con el ciclo de vida de la minería de datos:

Modelo de negocio

Entendimiento de los datos

Preparación de los datos

Modelado

Evaluación y Despliegue Crisp-DM es flexible y simplifica los pasos para la implementación del proyecto.

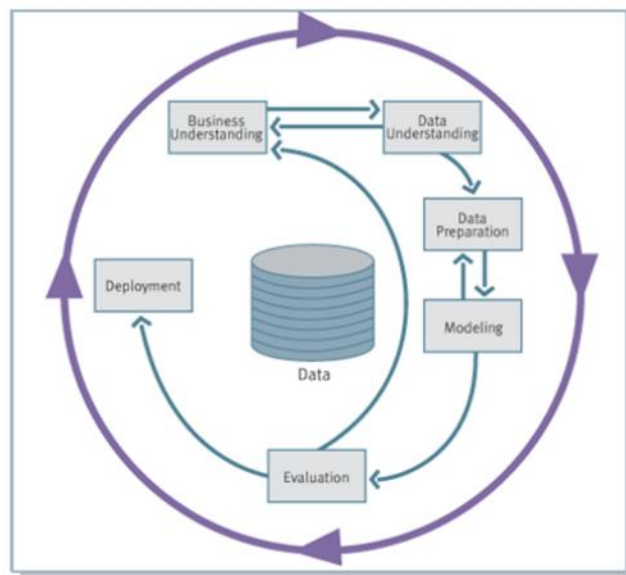


Figura 1 Metodología de Minería de Datos CRISP – DM

Magerit es una de las metodologías más utilizadas que permite analizar la gestión del riesgo de los Sistemas de Información, y será utilizada para valorar los riesgos que se podrían presentar en el proyecto (Molina Miranda, 2017).

Resultados y discusión

Resultados

Los resultados al seguir la metodología CRISP-DM son:

En la empresa reguladora de energía eléctrica de la ciudad de Babahoyo, existen inconvenientes relacionados a las observaciones de lecturas de los equipos de medición, que impiden realizar un adecuado análisis que contribuya a la toma de decisiones. Siguiendo el principio de confidencialidad de los datos de la empresa reguladora de energía, se presentan de manera global las principales observaciones de lectura analizadas durante los últimos 3 meses:

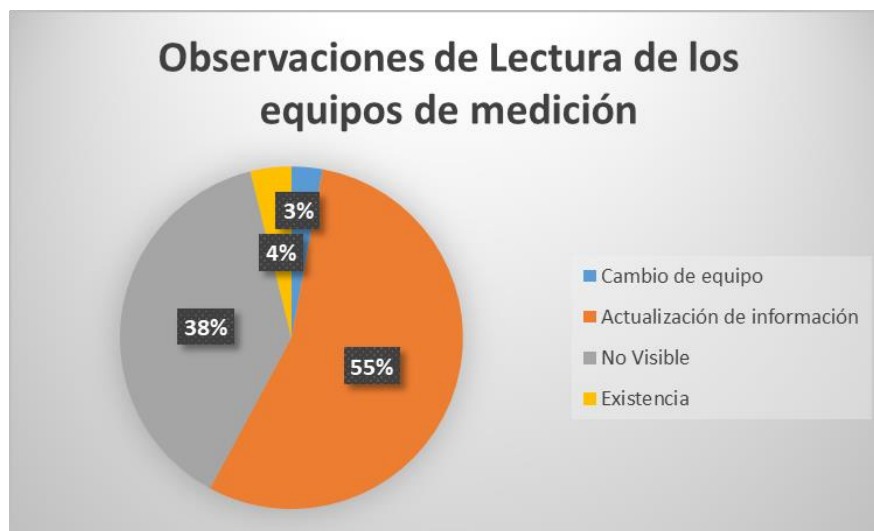


Figura 2 Observaciones de Lectura de los Equipos de Medición

De la base de datos se obtuvieron los siguientes campos, en relación a las observaciones de la toma de lectura del equipo de medición, se consideraron las siguientes observaciones para realizar el respectivo análisis:

Tabla 1 Descripción de Datos

Nombre	Tipo de dato	Descripción
PRO	Integer	Detalle del código de la Provincia
CAN	Integer	Detalle del código de la Cantón
SEC	Integer	Detalle del código de la Sector
Consumo.Promedio	Float	Detalle del Consumo Promedio
Minutos	Integer	Detalle de los minutos
Hora	Integer	Detalle de los hora

Novedad	Integer	Detalle de las Novedad
Año	Integer	Detalle del año
Tarifa	Varchar	Detalle de las tarifas
DiaN	Varchar	Detalle de los días
MesN	Varchar	Detalle de los meses
NovedadN	Varchar	Detalle de las novedades

Elaborado: Autores

Se realiza una exploración de los datos de la toma de lectura de los equipos de medición de las variables del archivo plano que recibe mensualmente la empresa reguladora de energía eléctrica con la función dim y análisis a la Base de datos. Se explora la cantidad de registros que se obtiene de la base de datos de la toma de lectura y la cantidad de columnas que se cuenta para la obtención de resultados.

```
> dim(Bd_nov)
[1] 99837 30
```

Figura 3 Cantidad de Registros

Los sectores se encuentran estructurados de acuerdo a los límites periféricos del interior de la ciudad, e identificados de forma numérica (1 al 6). Al analizar los datos de la toma de lectura se puede observar la concurrencia en los sectores de acuerdo a las observaciones reportadas.

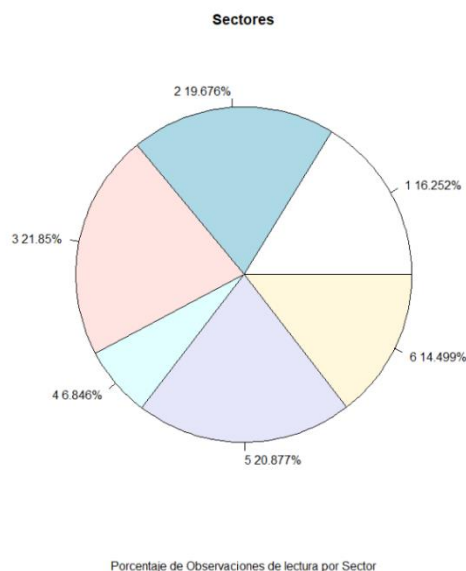


Figura 4 Porcentaje de observaciones en los Sectores

Se toma en consideración un periodo de tiempo compuesto por tres meses, para realizar el análisis de las observaciones de lectura de los equipos de medición de la empresa reguladora de energía eléctrica, de acuerdo a los porcentajes de la información levantada y la cual solo esta graficada lo que respecta a las observaciones que van a ser consideradas para la toma de decisiones.

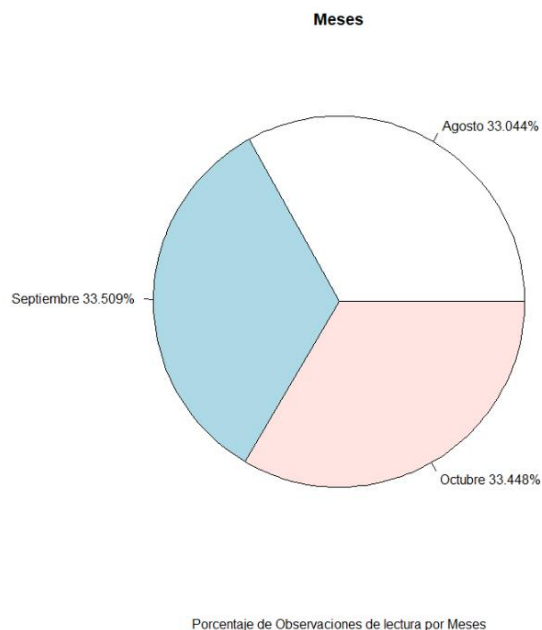


Figura 5 Porcentaje de Observaciones de los meses analizados

La función *boxplot()* en R studio, se utilizó para identificar las variables que contienen datos que se consideran aislados, los cuales no pueden formar parte del modelo por que causarían ruido. Estos se pueden evidenciar en la figura siguiente, donde aquellos puntos que se encuentran distantes son los que no se consideran.

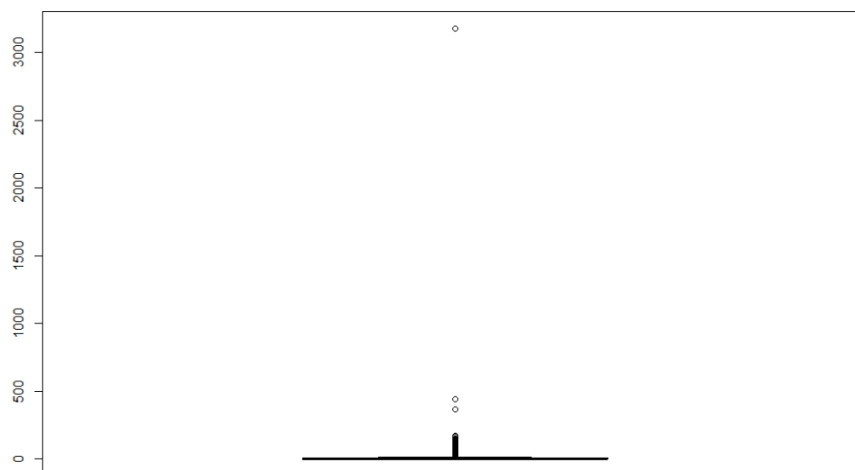


Figura 6 Outlier de Consumo

Para la selección de los datos se considera los objetivos y el entendimiento del negocio donde se va a realizar el análisis de las observaciones de los equipos de medición, se obtienen 99837 registros de la toma de lectura de los equipos de medición, donde se encuentran incluidos aquellos que no presentan observaciones, por lo cual serán aislados al realizar el entrenamiento de los modelos y la evaluación respectiva. Se realiza la visualización general de los datos con la función *summary()* con el nombre de la base de datos para el análisis descriptivo de las variables:

```
> summary(Bd_nov)
      PRO      CAN      SEC
Min.   :12   Min.   :1   Min.   :1.000
1st Qu.:12   1st Qu.:1   1st Qu.:2.000
Median :12   Median :1   Median :3.000
Mean   :12   Mean   :1   Mean   :3.393
3rd Qu.:12   3rd Qu.:1   3rd Qu.:5.000
Max.   :12   Max.   :1   Max.   :6.000
```

Figura 7 Summary de la Base de Datos

Se ejecuta la limpieza de los datos basura, para en lo posterior continuar con la aplicación del algoritmo para realizar el análisis respectivo.

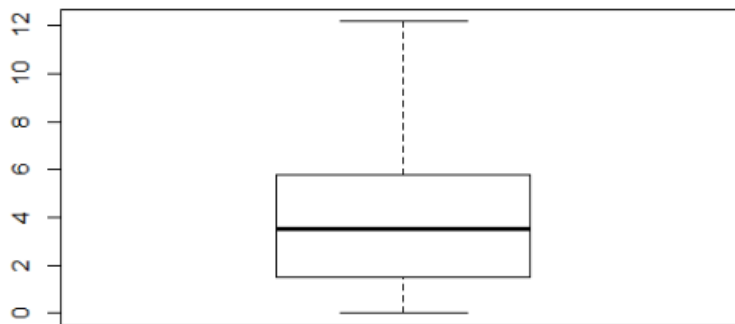


Figura 8 Boxplot después de la limpieza de los Consumos

Se procede a transformar los datos que son cualitativos a cuantitativos para el análisis de la información de las observaciones de lectura de los equipos de medición, para proceder después al entrenamiento de los algoritmos de minería de datos.

```
model_Bdnov_T=model.matrix(~Tarifa+0,Bd_nov)
model_Bdnov_D=model.matrix(~DiaN+0,Bd_nov)
model_Bdnov_M=model.matrix(~MesN+0,Bd_nov)
model_Bdnov_S=model.matrix(~Sector+0,Bd_nov)
model_Bdnov_N=model.matrix(~NovedadN+0,Bd_nov)
```

Figura 9 Construcción de datos cualitativos a cuantitativos

Después de la construcción de los datos, se procede a la integración de los mismos en un nuevo dataset, donde va a contemplar los datos que se procedió a la construcción de datos cualitativos a cuantitativos, además de las variables iniciales de la base de datos.

```
data_Bdnov=data.frame(Bd_nov$PRO,Bd_nov$CAN,Bd_nov$SEC,Bd_nov$Dia,
Bd_nov$Consumo.Promedio,Bd_nov$Consumo,Bd_nov$Año,
model_Bdnov_T,model_Bdnov_D,model_Bdnov_M,model_Bdnov_S,
model_Bdnov_N)
```

Figura 10 Integración de la Base de datos

Se procede a seleccionar técnicas de regresión lineal las cuales son las siguientes: *Lm (todas las variables)*, *Step*, *Lm (reducción manual)*, *RandomForest*. Para la elección del modelo ideal se basa en función del error más bajo que se generó en de los modelos evaluados, dando como resultado el algoritmo de la función STEP con el error más bajo de todos los modelos que se han realizado el respectivo entrenamiento y evaluación de los datos.

```
#Porcentaje del test & train
nr=nrow(data_Bdnov)
s=sample(nr,nr*0.7)
#train
dataBdnov_train=data_Bdnov[s,]
#test
dataBdnov_test=data_Bdnov[-s,]
dim(dataBdnov_train)
dim(dataBdnov_test)
head(dataBdnov_train)
```

Figura 11 Generar datos Supervisados

Para la elección del modelo ideal se basa en función del error más bajo que se generó en de los modelos evaluados, dando como resultado el algoritmo de la función STEP con el error más bajo de todos los modelos que se han realizado el respectivo entrenamiento y evaluación de los datos.

Tabla 2 Descripción del error de cada uno de los modelos evaluados

Algoritmo	Error cuadrático	Error absoluto
LM	No	No
STEP	3.642379e-24	1.907971e-12
LM (manual)	0.05115917	0.1716271
RandomForest	1.637066e-05	0.0001162791

En la herramienta shiny se muestra la gráfica de los 6 sectores, donde en 2 sectores como es el 3 y 5 se encuentran concentrado el mayor número de los dos tipos de observaciones que se generaron en el algoritmo de STEP, además se puede ir deslizando observándose el comportamiento de los sectores en función de las observaciones de la ecuación encontrada.

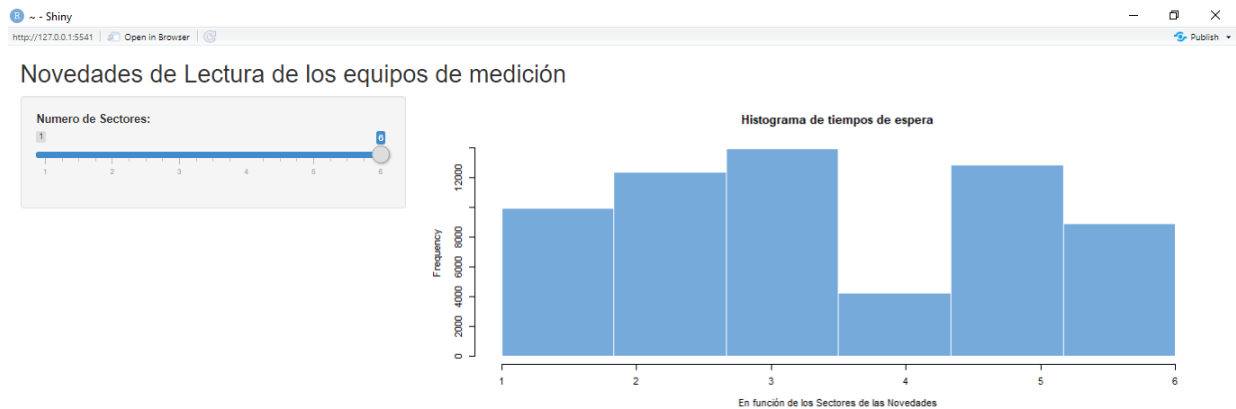


Figura 12 Dashboard de los Sectores de las Observaciones de Lectura de los Equipos de medición

Discusión

De acuerdo a los modelos entrenado con los diferentes técnicas de algoritmos ejecutados, encontramos 4 tipos de observaciones de los equipos de medición planteadas inicialmente en la situación actual de la empresa, se visualiza que se tiene el rango más alto de incidencia en la actualidad en la empresa reguladora de energía eléctrica son consideradas en la generación del algoritmo con el error más bajo que es STEP, por lo tanto a medida que se dé el trato correctivo por el área de comercial responsable se irá disminuyendo las observaciones, a medida que se tome los correctivos correspondientes irán disminuyendo las observaciones con porcentaje alto, la cual se incorporará paulatinamente los 2 tipos de observaciones en el algoritmo a medida que se mejore las observaciones para una mejor solución del problema inicial en la empresa reguladora de energía eléctrica.

Con el algoritmo STEP escogido al realizar el entrenamiento de los datos, con la ecuación lineal seleccionada por la evaluación y verificación del error por medio de las técnicas de minería de datos, con lo cual los dos tipos de observaciones seleccionadas son la no visible y existencia de los equipos de medición, en función con los sectores que se incursionan en la ciudad de Babahoyo para la toma de decisiones del área de pérdida de energía, balcón de servicios e instalación de medidores, para la solución en base de la ecuación encontrada colaborativa.

Al finalizar se realizó una encuesta que permitió conocer el alto nivel de aceptación por parte del área comercial de la empresa reguladora de energía eléctrica respecto a las técnicas de aprendizaje automático de la minería de datos que permitirá disminuir las observaciones y el dashboard en

función de la ecuación encontrada para la ilustración de los sectores con los dos tipos de observaciones que fueron consideradas en el algoritmo elaborado.

Referencias

1. ARCONEL. (21 de Octubre de 2018). Regulación Eléctrica. Obtenido de Regulación Eléctrica: https://www.regulacionelectrica.gob.ec/introduccion_regulacion/
2. Begueri, G. E., & Malberti, M. A. (2017). Minería de datos y una aplicación en la educación superior. Red de Universidades con Carreras en Informática (RedUNCI), 1276-1280.
3. Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. Revista Tecnológica ESPOL, 170-183.
4. Castillo Rojas, W., Medina Quispe, F., & Vega Damke, J. (2017). Esquema de Visualización para Modelos de Clústeres en Minería de Datos. Revista Ibérica de Sistemas e Tecnología de la Información, 67-84.
5. CNEI. (18 de Septiembre de 2018). Atención de reclamo Comercial. Obtenido de Atención de reclamo Comercial: <https://www.gob.ec/cnel/tramites/atencion-reclamo-comercial>
6. Diaz Narvaez, V. P., & Calzadilla Nunez, A. (2016). Artículos científicos, tipos de investigación y productividad científica en las Ciencias de la Salud. Scielo, 115-121.
7. Fernandez Collado, C., Hernandez Sampieri, R., & Baptista Lucio, P. (2017). Metodología de la Investigación. Mc graw Hill.
8. Gandrud, C. (2015). Reproducible Research with R and R Studio, Second Edition. Chapman and Hall/CRC.
9. Horton, N., & Kleinman, K. (2015). Using R and RStudio for Data Management, Statistical Analysis and Graphics. London: Chapman and Hall/CRC.
10. Jaramillo, A., & Paz Arias, H. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. Revista Tecnológica ESPOL , 64-90.
11. Karthik Ramasubramanian, A. S. (2017). Machine Learning Using R. Apress.
12. Labs, G. (23 de 05 de 2019). Planeta Chat Bot. Obtenido de Planeta Chat Bot: <https://planetachatbot.com/claves-para-diferenciar-inteligencia-empresarial-de-inteligencia-artificial-e494ece43970>

13. Molina Miranda, M. (2017). ANÁLISIS DE RIESGOS DE CENTRO DE DATOS BASADO EN LA HERRAMIENTA PILAR DE MAGERIT. *Revista Espirales*, 1-9.
14. Mueller, J. P., & Massaron, L. (2016). *Machine Learning for Dummies*. John Wiley & Sons, Incorporated.
15. Pereira, M. (2015). *La regulación Energética*. Universidad Libre Seccional, 91.
16. Pimpler, E. (2018). *Data Visualization and Exploration with R*. Geospatial Training Services.
17. Sandoval, L. J. (2018). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y. *ITCA-FEPADE*, 36-40.
18. Usuelli, M. (2014). *R Machine Learning Essentials*. Birmingham B3 2PB, UK.: PacktLib.
19. Williams, G. (2017). *The Essentials of Data Science: Knowledge Discovery Using R*. CRC Press LLC.
20. Zielesny, A. (2016). *From Curve Fitting to Machine Learning: An Illustrative Guide to Scientific Data Analysis and Computational Intelligence*. Recklinghausen: Springer.