

Modelo Predictivo de Regalías Mineras aplicando Técnicas de Analítica Predictiva con R

Carla Gabriela Calisaya Choque
Postgrado en Informática
Universidad Mayor de San Andrés
La Paz - Bolivia
carla.calisaya24@gmail.com

Resumen—El presente trabajo de investigación hace referencia a la carencia de patrones de comportamiento que permitan tener un análisis integral de la información de la regalía minera, formulándose la siguiente pregunta de investigación ¿Cuál es la incidencia de un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva, en el nivel de incertidumbre de la toma de decisiones gerenciales u operativas? Consecuentemente el objetivo de la investigación, es desarrollar un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva para reducir el nivel de incertidumbre en la toma de decisiones gerenciales u operativas. Para ello se implementa el prototipo del modelo predictivo de regalías mineras utilizando técnicas de analítica predictiva con la metodología CRISP-DM, utilizando datos de regalías mineras y las variables que forman parte de su cálculo delimitada a cuatro minerales. El análisis de información y creación de modelos predictivos es realizado con el lenguaje de programación R, se valida el resultado de los modelos aplicando el método de validación cruzada de k-iteraciones o k-fold cross validation, y se selecciona el modelo que se ajusta mejor a los datos en base a las métricas de evaluación.

Palabras clave—Modelo, predictivo, analítica predictiva, regalía minera, toma de decisión, árboles de decisión, k-vecinos más cercanos, random forest, redes neuronales, CRISP-DM.

I. INTRODUCCIÓN

La minería es la segunda industria de extracción de Bolivia, entre las riquezas mineras del país, las mayores se concentran en los departamentos de Potosí, La Paz y Oruro, productores de estaño, zinc, plata, cobre, wólfam, oro, antimonio entre otros. Santa Cruz cuenta con los yacimientos más importantes de hierro y piedras semipreciosas como la bolivianita y amatista; los departamentos de Pando y Beni producen principalmente oro; Cochabamba, Tarija y Chuquisaca producen minerales no metálicos, como piedra caliza, yeso, arcillas y otros.

El Estado a través de la Constitución Política del Estado, Ley N° 535 de Minería y Metalurgia y Decreto Supremo N° 29577 regulan la Regalía Minera que surge con el propósito de compensar la explotación de los recursos minerales y metales no renovables, aplicada obligatoriamente a las actividades de explotación, concentración y/o comercialización, fundición, refinación e industrialización, y en la prospección y exploración minera.

La presente investigación se enfocará en estudiar y analizar la información histórica de la regalía minera durante el periodo 2010 a 2018, así como las variables peso, cotización y alícuota,

que permitirán desarrollar un modelo predictivo de regalías mineras para optimizar la toma de decisiones y minimizar el nivel de incertidumbre, mejorar el conocimiento sobre el tema y obtener la capacidad de poder predecir sus acciones futuras.

II. MARCO DEL PROBLEMA

A. Planteamiento del problema

La carencia de patrones de comportamiento que permita tener un análisis integral de la información de la regalía minera, puede ocasionar una toma de decisiones inadecuada o extemporánea en la distribución de recursos que los Gobiernos Autónomos Departamentales y Municipales deben destinar a actividades como la prospección y exploración minera, reactivación productiva y monitoreo ambiental; y proyectos de inversión pública en el área de impacto de la operación minera, respectivamente; así como una toma de decisiones inoportuna en la aplicación de políticas públicas, a nivel gerencial u operativo.

B. Formulación del Problema de Investigación

De acuerdo a lo definido anteriormente el problema ha sido formulado de la siguiente manera:

¿Cuál es la incidencia de un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva, en la toma de decisiones gerenciales u operativas?

C. Planteamiento del objetivo

1) Objetivo general

Desarrollar un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva para reducir el nivel de incertidumbre en la toma de decisiones gerenciales u operativas.

2) Objetivos específicos

Los objetivos específicos son los siguientes:

- Recolectar, procesar y analizar la información de las variables que intervienen en el cálculo de la regalía minera.
- Desarrollar y/o construir un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva.
- Validar el modelo predictivo de regalías mineras.

D. Planteamiento de hipótesis

El desarrollo de un modelo predictivo de regalías mineras a través de técnicas de analítica predictiva reduce el nivel de

Para referenciar este artículo (IEEE):

[N] C. Calisaya, «Modelo Predictivo de Regalías Mineras aplicando Técnicas de Analítica Predictiva con R», *Revista PGI. Investigación, Ciencia y Tecnología en Informática*, n° 8, pp. 36-40, 2020.



incertidumbre de la toma de decisiones a nivel gerencial u operativo.

1) Variables dependientes

La variable dependiente está definida de la siguiente manera:

Modelo predictivo de regalías mineras a través de técnicas de analítica predictiva.

2) Variables independientes

Nivel de incertidumbre de la toma de decisiones a nivel gerencial u operativo.

3) Operacionalización de variables

La variable independiente modelo predictivo de regalías mineras, a través de técnicas de analítica predictiva, tiene como definición operacional una función matemática y/o secuencia de pasos desarrollado a través de técnicas de analítica predictiva que predecirán el comportamiento de los valores de la regalía minera. Y la variable dependiente nivel de incertidumbre de la toma de decisiones a nivel gerencial u operativo, surgirá en ocasiones donde no pueden asignarse probabilidades a los eventos posibles, a la hora de tomar una decisión.

III. MARCO TEÓRICO

A. Modelo predictivo

Según la definición de la Real Academia Española el modelo es el “esquema teórico, generalmente en forma matemática, de un sistema o de una realidad compleja, como la evolución económica de un país, que se elabora para facilitar su comprensión y el estudio de su comportamiento”. Ríos (1995) señala que “un modelo es un objeto, concepto o conjunto de relaciones que se utiliza para representar y estudiar de forma simple y comprensible una porción de la realidad empírica”.

B. Regalía Minera

Según la definición del diccionario de la Real Academia Española la regalía “es la participación en los ingresos o cantidad fija que se paga al propietario de un derecho a cambio del permiso para ejercerlo”.

La Ley No. 535 de Minería y Metalurgia promulgada el 28 de mayo de 2014, en su Título VII, Capítulo I, Artículo 223 establece que la “Regalía Minera – RM, por el aprovechamiento de los recursos minerales y metales no renovables, es un derecho y una compensación por su explotación”, regulada por la Constitución Política del Estado y la mencionada ley.

Asimismo, en su Artículo 226 determina “La base de cálculo de Regalía Minera – RM, es el valor bruto de venta, que resulta de multiplicar el peso del contenido fino del mineral o metal por su cotización oficial” [1].

Por lo mencionado anteriormente la Regalía Minera se define en base a la siguiente función:

$$RM(PF, CO, A) \quad (1)$$

Asimismo, la fórmula definida para el cálculo de la Regalía Minera es la siguiente:

$$RM = VBV * \frac{A}{100} \quad (2)$$

Y la fórmula para el Valor Bruto de Venta es:

$$VBV = PF[unidades cotización] * \frac{CO}{1[unidades cotización]} \quad (3)$$

Donde:

RM Regalía Minera.

PF Peso Fino en la unidad de cotización internacional del mineral o metal.

CO Cotización Oficial del mineral o metal en dólares americanos.

A Alícuota del mineral o metal.

VBV Valor Bruto de Venta.

C. Analítica predictiva

Según la International Business Machines Corporation (IBM), la analítica predictiva combina prestaciones de analítica avanzada, que incluyen análisis estadístico ad-hoc, modelado predictivo, minería de datos, análisis de texto, optimización, puntuación en tiempo real y machine learning. Estas herramientas ayudan a las empresas a descubrir patrones en datos y a dar un paso más allá de saber qué ha ocurrido a anticiparse a lo que probablemente ocurrirá a continuación [2].

La analítica predictiva es una forma de análisis avanzado que utiliza datos nuevos e históricos para predecir la actividad futura, el comportamiento y las tendencias. Implica la aplicación de técnicas de análisis estadístico, consultas analíticas y algoritmos automáticos de aprendizaje automático a conjuntos de datos para crear modelos predictivos que sitúen un valor numérico o puntuación en la probabilidad de que ocurra un evento particular [3].

Entre las técnicas aplicables a la analítica predictiva se encuentran las de análisis estadístico, consultas analíticas y algoritmos automáticos de aprendizaje, para la creación de modelos predictivos; entre estas herramientas se encuentran el modelo de regresión lineal, árboles de clasificación y regresión, redes neuronales y k-vecinos más cercanos entre otros.

Árboles de decisión (clasificación y regresión): La metodología CART utiliza datos históricos para construir árboles de clasificación o de regresión los cuales son usados para clasificar o predecir nuevos datos. Estos árboles CART pueden manipular fácilmente variables numéricas y/o categóricas. Son árboles de regresión cuando la variable dependiente es continua (numérica) y árboles de clasificación cuando la variable dependiente es cualitativa (categórica).

Esta metodología consiste en tres pasos: Construcción del árbol saturado, escogencia del tamaño correcto del árbol y clasificación de nuevos datos usando el árbol construido.

K-vecinos más cercanos: K-NN (K-Nearest Neighbor) es un algoritmo de clasificación de aprendizaje supervisado. Este algoritmo clasifica cada nuevo ejemplo calculando la distancia de ese ejemplo con todos los del conjunto de train. La clase predicha para este nuevo ejemplo vendrá dada por la clase a la que pertenezcan los ejemplos más cercanos del conjunto de train, el valor de la k es el que determina en cuantos vecinos debemos fijarnos para predecir la clase. Así, con un valor de k = 1, la clase predicha para cada nuevo ejemplo será la clase a la que pertenezca el ejemplo más cercano del conjunto de train [4].

Random forest: O Bosques Aleatorios es un método de aprendizaje automático, propuesto inicialmente por Kam Ho de Laboratorios Bell, y posteriormente desarrollado por Breiman. Random forest surge como combinación de las técnicas de Classification And Regression Tree (CART) y Bootstrap Aggregating (Bagging) para realizar la combinación de árboles predictores en la que cada árbol depende de los valores de un

vector aleatorio probado independientemente y con la misma distribución para cada uno de estos [5].

Redes neuronales: La red neuronal artificial es un algoritmo de aprendizaje supervisado, Haykin [6] define las Redes Neuronales Artificiales (RNAs) como un paradigma computacional en el cual una gran cantidad de unidades de cómputo simple (las neuronas) interconectadas en red, realizan tareas de procesamiento de datos. Una RNA puede definirse formalmente como un procesador paralelo y distribuido que tiene la capacidad de almacenar conocimiento basado en la experiencia. Además, se parece al cerebro humano en cuanto a que el conocimiento es adquirido a través de un proceso de aprendizaje y las conexiones entre las neuronas son usadas para almacenar ese conocimiento.

D. Lenguaje de programación R

La herramienta a utilizar en esta investigación será R, que es un lenguaje de programación interpretado, de distribución libre, bajo Licencia GNU y se mantiene en un ambiente para el cómputo estadístico y gráfico [7].

R proporciona una amplia variedad de técnicas estadísticas (modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupación, ...) y técnicas gráficas, y es altamente extensible. Disponible como Software Libre bajo los términos de la Licencia Pública General GNU de la Free Software Foundation en forma de código fuente. Compila y se ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluyendo FreeBSD y Linux), Windows y MacOS.

IV. MÉTODOS

A. Diseño Metodológico

El diseño a utilizar en el presente trabajo es no experimental, debido a que se trabajará con información histórica de las regalías mineras; ya que según Hernández [8] la investigación no experimental es aquella que se realiza sin manipular deliberadamente variables, es decir, observar fenómenos tal y como se dan en su contexto natural, para después analizarlos.

B. Tipo de investigación

El tipo de estudio de inicio será descriptivo para posteriormente utilizar el tipo de estudio correlacional. Considerando que el estudio descriptivo busca especificar propiedades, características y tendencias de las variables que intervienen en el cálculo de la regalía minera; asociando las mismas al diseño del modelo predictivo lo que llevará al tipo de estudio correlacional.

C. Método de investigación

El método a utilizar en el presente trabajo de investigación es el mixto, el método deductivo que parte de lo general a lo particular, aplicado en las técnicas de analítica predictiva a casos particulares de los minerales, y el método inductivo que va de lo particular a lo general, enfocado en los resultados de la aplicación de teorías a cada mineral para generalizar el modelo predictivo a obtener.

D. Fases Metodológicas

El trabajo de investigación empleará la metodología CRISP-DM [9] que en el contexto del proyecto comprende las siguientes fases: comprensión del negocio o problema, análisis y comprensión de los datos, preparación de los datos, modelado,

evaluación de resultados e implantación tal como se muestra en la Figura 1.

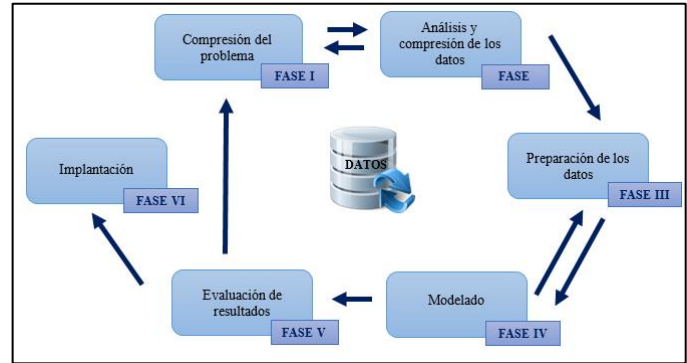


Fig. 1. Fases metodológicas

E. Técnicas de investigación

La realización del estudio implicará principalmente:

- Análisis documental y bibliográfico relacionado con el contenido y alcance de la investigación;
- Técnica de análisis de contenido, que consiste en sistematizar la información contenida en registros escritos, en este caso publicaciones realizadas por el Ministerio de Minería y Metalurgia en dossiers, anuarios y boletines estadísticos, la cual es organizada y almacenada en una base de datos diseñada en PostgreSQL.

F. Universo o población de referencia

Considerando que los parámetros de cálculo de la regalía minera están definidos para la comercialización interna y exportación de minerales, y que los minerales se clasifican en minerales metálicos y no metálicos; en el presente trabajo de investigación se manipulará únicamente información de las exportaciones de minerales metálicos de gestiones 2010 a 2018.

G. Muestra parametrizada

De acuerdo a la población obtenida se tiene que desde la gestión 2010 a 2018 se exportan en promedio anual 15 tipos de minerales metálicos. Según Suárez [10], afirma que la muestra es un subconjunto de la población, y que sus principales características son: Representativa, Adecuada y válida.

Por lo mencionado, se obtiene el tamaño de la muestra de la variable mineral aplicando la siguiente fórmula:

$$n = \frac{N\sigma^2 Z^2}{(N-1)e^2 + N\sigma^2 Z^2} \quad (4)$$

Donde:

n: Tamaño de la muestra.

N: Tamaño de la población.

σ : Desviación estándar.

Z: Nivel de confianza (1.96 al 95% de confianza).

e: Límite aceptable de error muestral.

Obteniendo como resultado $n = 4$, se realiza un análisis adicional donde se identifica que los principales minerales que aportan regalías mineras a los gobiernos autónomos departamentales y municipales del Estado son: plata, zinc, oro y estaño.

H. Delimitación

1) *Delimitación geográfica:* La delimitación geográfica en la investigación son las regalías mineras del Estado Plurinacional de Bolivia.

2) *Delimitación temporal:* Información de las regalías mineras de las gestiones 2010 a 2018.

V. RESULTADOS

La Tabla I muestra un resumen del desempeño de los modelos construidos con árboles de regresión, k-nearest neighbor, random forest y redes neuronales para los minerales zinc, plata, oro y estaño. Aplicando métricas de evaluación error porcentual absoluto medio (MAPE), error cuadrático medio (MSE) y R cuadrado (R2).

TABLA I. RESULTADO DE MÉTRICAS DE EVALUACIÓN DE LOS MODELOS CONSTRUIDOS POR MINERAL

MINERAL	MODELO	MSE	R2	MAPE
Zinc	K-Nearest Neighbor	4.08672E+11	0.989861104	3.19%
	Random Forest	3.10700E+12	0.945955793	6.23%
	Árboles de regresión	6.03568E+12	0.850194921	7.55%
	Redes Neuronales	2.14386E+14	0.973855533	9.55%
Plata	K-Nearest Neighbor	9.23711E+11	0.966112293	3.41%
	Árboles de regresión	3.49813E+12	0.858445904	5.80%
	Random Forest	1.18693E+13	0.878851908	7.19%
	Redes Neuronales	2.34082E+14	0.985039867	8.10%
Oro	K-Nearest Neighbor	16049007445	0.986483047	3.75%
	Árboles de regresión	98982684850	0.907190004	8.33%
	Random Forest	3.04144E+11	0.835666528	9.43%
	Redes Neuronales	3.21319E+12	0.988214895	9.85%
Estaño	K-Nearest Neighbor	25142975350	0.916707941	4.05%
	Árboles de regresión	39619823031	0.851315012	5.02%
	Random Forest	88368976332	0.792783402	6.67%
	Redes Neuronales	4.44387E+11	0.985560435	8.35%

Las técnicas de analítica predictiva empleadas en el proceso de modelado han permitido encontrar una respuesta a las preguntas planteadas desde la fase de conocimiento del negocio, determinando el modelo que mejor se ajusta a los patrones de comportamiento de la variable regalía minera para los minerales zinc, plata, oro y estaño. A continuación, se expone la Tabla II, con el resultado de la exactitud de los modelos construidos por mineral.

TABLA II. EXACTITUD DE LOS MODELOS POR MINERAL

Mineral	K-Nearest Neighbor	Árboles de regresión	Random Forest	Redes Neuronales
Zinc	3.19%	6.23%	7.55%	9.55%
Plata	3.41%	5.80%	7.19%	8.10%
Oro	3.75%	8.33%	9.43%	9.85%
Estaño	4.05%	5.02%	6.67%	8.35%

Tal como se observa en la tabla precedente, los modelos construidos con K-Nearest Neighbor son los que tienen mejor exactitud, por lo que nuestro modelo predictivo de regalías

mineras estaría basado en el algoritmo k- Nearest Neighbor, con valores de k igual a 3,5,2 y 2 para los minerales zinc, plata, oro y estaño respectivamente.

Para la fase de implantación se propone la construcción de un dashboard¹ en Power BI², que permite ver gráficamente el comportamiento histórico de las variables peso, cotización, alícuota y regalía minera, tal como se observa en la Figura 2.

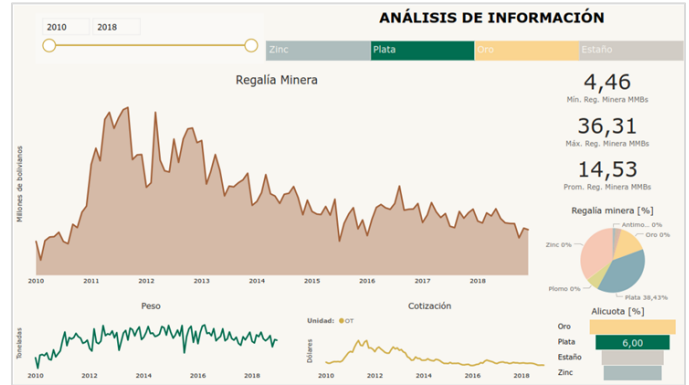


Fig. 2. Dashboard información histórica

Asimismo, se propone uno similar para mostrar las predicciones de las regalías mineras por mineral en la Figura 3 que servirá de apoyo a la toma de decisiones.

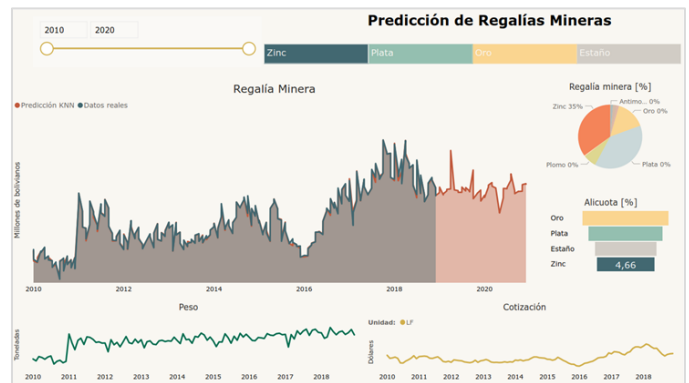


Fig. 3. Dashboard predicción de regalías mineras por mineral

VI. DISCUSIÓN

A. Conclusiones

Se realizó la recopilación, procesamiento y análisis de información de regalías mineras y variables que intervienen para su cálculo, tomando como muestra 4 minerales: zinc, plata, oro y estaño de las exportaciones de minerales.

Se desarrolló el prototipo de Modelo Predictivo de Regalías Mineras utilizando técnicas de analítica predictiva, que permitieron establecer el patrón de comportamiento de las variables que componen el cálculo de la regalía minera para reducir el nivel de incertidumbre en la toma de decisiones gerenciales u operativas.

¹ Un dashboard es una representación visual de la información más importante y necesaria para lograr uno o más objetivos, de forma consolidada y organizada en una sola pantalla [11].

² Power BI es un servicio de análisis empresarial de Microsoft, proporciona información detallada para permitir la toma de decisiones rápidas e informadas a través de las visualizaciones interactivas y capacidades de inteligencia empresarial [12].

Se procedió a la validación del modelo (prototipo) con el método de validación cruzada de *k-fold cross validation* y métricas de evaluación (error absoluto medio, error cuadrático medio y R cuadrado), con el objetivo de evaluar los resultados obtenidos en la aplicación de técnicas de analítica predictiva.

Se construyeron modelos predictivos con base en árboles de decisión, *k-nearest neighbor*, *random forest* y redes neuronales, de los cuales se obtuvo el modelo con mejor desempeño que fue el construido con *k-nearest neighbor*.

B. Recomendaciones

En cuanto a recomendaciones que surgen del presente trabajo se tendría las siguientes:

- Implementar técnicas de analítica predictiva en diferentes campos de estudio.
- Desarrollar e implantar modelos predictivos en otros campos de estudio con muestras cuantitativas relacionadas a cambios temporales utilizando herramientas informáticas como R, Python, Power BI, Tableau y otras tecnologías informáticas relacionadas.

C. Discusión

Por lo antes expuesto, se acepta la hipótesis de que el desarrollo de un modelo predictivo de regalía mineras a través de técnicas de analítica predictiva reduce el nivel de incertidumbre en la toma de decisiones a nivel gerencial u operativo. Considerando que, de los modelos construidos, aplicando las técnicas de modelado predictivo: árboles de decisión, *k-nearest neighbor*, *random forest* (bosques aleatorios) y redes neuronales, permiten obtener el modelo con mejor exactitud en relación a las otras técnicas mencionadas, con la construcción *K-Nearest Neighbor* se podrá realizar la mejor predicción a futuro del valor de la regalía minera.

REFERENCIAS

- [1] E. P. d. Bolivia, *Ley N° 535 de Minería y Metalurgia*, 2014.
- [2] IBM, «Analítica Predictiva.» [En línea]. Available: <https://www.ibm.com/es-es/analytics/predictive-analytics>.
- [3] M. Rouse, «Tech Target.» Abril 2017. [En línea]. Available: <https://searchdatacenter.techtarget.com/es/definicion/Analitica-predictiva-o-analisis-predictivo>.
- [4] Berastegui A. G. y Galar I. M., «Implementación del algoritmo de los k vecinos más cercanos (k-NN) y estimación del mejor valor local de k para su cálculo.» Pamplona, 2018.
- [5] J. Alarcón, «Modelos de minería de datos: random forest y adaboost, para identificar los factores asociados al uso de las TIC (internet, telefonía Fija y televisión de paga) en los hogares del Perú. 2014.» 2017. [En línea]. Available: https://cybertesis.unmsm.edu.pe/bitstream/handle/cybertesis/7404/Alarcon_fj.pdf?sequence=3&isAllowed=y.
- [6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, 2007.
- [7] Santana S., Farfán E., «El arte de programar en R: un lenguaje para la estadística.» 27 Noviembre 2014. [En línea]. Available: https://cran.r-project.org/doc/contrib/Santana_EL_arte_de_programar_en_R.pdf.
- [8] Hernández, R., Fernández, C. y Baptista M. P., *Metodología de la Investigación*, México: Mc Graw Hill, 2010.
- [9] IBM Corporation, «Guía de CRISP-DM de IBM SPSS Modeler.» 1994-2019. [En línea]. Available: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_modeling_phase.html.
- [10] M. Suárez, *Interaprendizaje Holístico de Matemática*, Quito-Ecuador, 2004.
- [11] D. Martínez, «Metodología para el diseño de Dashboards orientado hacia el registro de evidencias en el proceso de evaluaciones institucionales.» 24 Julio 2017. [En línea]. Available: https://reunir.unir.net/bitstream/handle/123456789/6171/MARTINEZ_%20ROBALINO%2C%20DANIEL%20ANDRES.pdf?sequence=1&isAllowed=y.
- [12] Microsoft Power BI, «Power BI.» 2019. [En línea]. Available: <https://powerbi.microsoft.com/es-es/>.

Breve CV de la autora

Carla Gabriela Calisaya Choque es Licenciada en Informática con mención en Ingeniería de Sistemas Informáticos por la Universidad Mayor de San Andrés. Actualmente realiza la Maestría en Alta Gerencia en Tecnologías de la Información y Comunicaciones e Innovación para el Desarrollo (MAG-TIC) del Postgrado en Informática de la UMSA.

Ejerce profesionalmente como Analista de Sistemas de Información en la Autoridad Jurisdiccional Administrativa Minera. Trabajó en el Servicio Nacional de Registro y Control de la Comercialización de Minerales y Metales como Administradora de Base de Datos.

Es miembro del WiDS La Paz (Women in Data Science La Paz). Entre sus intereses investigativos están el Análisis de Datos, Machine Learning y la Ciencia de Datos. Email: carla.calisaya24@gmail.com