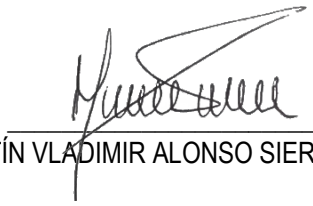
SISTEMA EMPRESARIAL PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE SOSTENIBILIDAD EN ORGANIZACIONES MINERAS

BRYAN JOSSER FERNÁNDEZ EBRATH; JHON ENRIQUE RAMÍREZ MANTILLA; ANDRÉS FELIPE ORTEGA LUCERO

Nota de Aceptación

Certificamos que el presente Trabajo de Grado Satisface, en alcances y calidad, todos los requisitos que demanda un Trabajo de Grado de Maestría.



MARTÍN VLADIMIR ALONSO SIERRA GALVIS

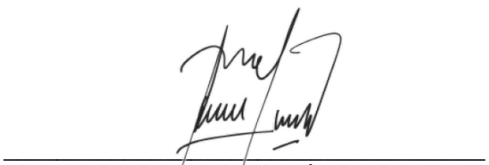
Director

LUIS EDUARDO TOBÓN – Jurado 1

VALENTINA CORCHUELO – Jurado 2

Aprobado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana Cali, para optar el título de Magister en Ciencia de Datos.

HERNÁN CAMILO ROCHA NIÑO Ph. D.
Decano Facultad de Ingeniería y Ciencias



JUAN CARLOS MARTÍNEZ ARIAS
Director Posgrados de Ingeniería y Ciencias

Santiago de Cali, 05 de julio de 2023

Acta de Correcciones al Documento de Trabajo de Grado

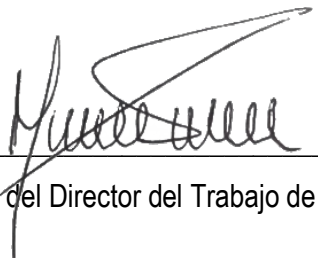
Santiago de Cali, 03 de julio 2023

Autor:
Jhon Enrique Ramírez Mantilla, Andrés Felipe Ortega Lucero, Bryan Josser Fernández Ebrath

Título del Trabajo de Grado: "SISTEMA EMPRESARIAL PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE SOSTENIBILIDAD EN ORGANIZACIONES MINERAS"

Director: Martín Vladimir Alonso Sierra Galvis

Como indica el artículo 2.13 de las Directrices para Trabajo de Grado de Maestría, he verificado que el estudiante indicado arriba ha implementado todas las correcciones que los Jurados del Proyecto de Trabajo de Grado definieron que se efectuaran, como consta en el Acta de Evaluación correspondiente.



Firma del Director del Trabajo de Grado

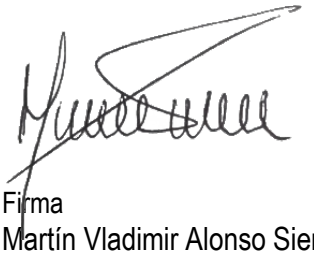
Santiago de Cali, 29 de mayo de 2023

Ingeniero:
Juan Carlos Martínez Arias
Director Posgrados de Ingeniería
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana - Cali

Con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el Trabajo de Grado y posteriormente optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto de Trabajo de Grado denominado SISTEMA EMPRESARIAL PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE SOSTENIBILIDAD EN ORGANIZACIONES MINERAS, el cual será realizado por el (los) estudiante(s) Bryan Josser Fernández Ebrath, Jhon Enrique Ramírez Mantilla, y Andrés Felipe Ortega Lucero con código 1048272259, 8971663, y 1082657705 pertenecientes al énfasis en Sistemas y Computación, bajo la dirección del profesor Martín Vladimir Alonso Sierra Galvis.

El suscrito director del Trabajo de Grado autoriza para que se proceda a hacer la evaluación de este Proyecto ante el Tribunal que para el efecto se designe, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado oficialmente.

Atentamente,



Firma
Martín Vladimir Alonso Sierra Galvis

C.C. 1144034673 de Cali

Firma
Bryan Josser Fernández Ebrath

C.C. 1048272259 de Malambo

Firma
Andrés Felipe Ortega Lucero

C.C. 1082657705 de Ancuya

Firma
Jhon Enrique Ramírez Mantilla

C.C. 1095821204 de Floridablanca

ALIANZA POR LA
MINERÍA RESPONSABLE

29 de mayo de 2023

SEÑORA:
Gloria Inés Álvarez V.
Directora de la Maestría en Ciencia de Datos
Pontificia Universidad Javeriana Cali
Cali-Colombia

Cordial saludo,

Por medio de la presente, desde **Alianza por la Minería Responsable (ARM)** expresamos nuestro total conocimientoyapoyodel proyecto denominado "**SISTEMA EMPRESARIAL PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE SOSTENIBILIDAD EN ORGANIZACIONES MINERAS"**, desarrollado por los estudiantes de la maestría en Ciencia de Datos: Bryan Josser Fernández Ebrath (Código 1048272259), Jhon Enrique Ramírez Mantilla (Código 8971663) y Andrés Felipe Ortega Lucero (Código 1082657705).

ARM facilitó los datos requeridos de las organizaciones mineras y otorgó el permiso de su tratamiento para fines académicos y poder así llevar a cabo satisfactoriamente el proyecto de grado de los estudiantes mencionados.

En el evento de requerir información adicional puedo ser contactado al correo jairocardenas@responsiblemines.org

Sin otro particular,

Muy atentamente,

JAIRO CÁRDENAS
Gerente de Sostenibilidad Minera
Fundación Alianza por la Minería Responsable

Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias

FICHA RESUMEN
TRABAJO DE GRADO DE MAESTRÍA

TITULO: "Sistema empresarial para la gestión y análisis de datos de sostenibilidad en organizaciones mineras"

1. ÉNFASIS: Sistemas y Computación
2. TIPO DE PROYECTO: Aplicado
3. ÁREA DE TRABAJO: Ciencia de datos
4. ESTUDIANTE (S): Bryan Josser Fernández Ebrath; Jhon Enrique Ramírez Mantilla; Andres Felipe Ortega Lucero
5. CORREO ELECTRÓNICO: bjfernandez@javerianacali.edu.co; jeramirezm@javerianacali.edu.co; afortegal@javerianacali.edu.co
6. DIRECCIÓN Y TELÉFONO: Cra 19 No 47 – 02, 3004369428; AK. 9 #127B-24 APTO 302, Bogotá D.C., 3103394830; Carrera 31 D # 65-103, Palmira, 3165311588.
7. DIRECTOR: Martín Vladimir Alonso Sierra Galvis
8. VINCULACIÓN DEL DIRECTOR (en la universidad): Planta
9. CORREO ELECTRÓNICO DEL DIRECTOR: martinv.sierra@javerianacali.edu.co
10. CO-DIRECTOR (ES) (Si aplica): NA
11. GRUPO O EMPRESA QUE LO AVALA (Si aplica): NA
12. OTROS GRUPOS O EMPRESAS: Alianza por la Minería Responsable, ARM.
13. PALABRAS CLAVE (al menos 5): ciencia de datos, algoritmos de regresión, regresión ponderada, inteligencia de negocios, aplicativo web, sostenibilidad minera.
14. ODS QUE APLICA EL PROYECTO (Agenda 2030): Objetivo 12 Producción y consumo responsables
15. FECHA DE INICIO (Desarrollo del proyecto): 1/06/2022

16. RESUMEN (máximo 400 palabras).

El monitoreo del flujo de datos generado por la comercialización diaria de oro y otros minerales por parte de los grupos mineros artesanales y de pequeña escala - MAPE, con énfasis en su conexión con el mercado formal, ha sido una prioridad clave para la Alianza por la Minería Responsable (AMR). No obstante, el manejo de grandes volúmenes de datos y su constante dinamismo ha dificultado el seguimiento en tiempo real, la toma oportuna de decisiones y la generación de observaciones que permitan la definición ágil de planes de acción. En consecuencia, el uso de técnicas de gestión efectiva de datos y la aplicación de inteligencia artificial para el análisis automatizado han despertado un gran interés en la AMR. Estas técnicas permiten identificar grupos mineros en situación de riesgo, asegurando su acceso a cadenas de comercialización con precios más competitivos y su vinculación efectiva al mercado. En una primera etapa, se llevó a cabo un análisis exploratorio de los datos disponibles para la AMR, con el objetivo de identificar las técnicas más apropiadas para la predicción y clasificación, así como definir un modelo de datos adecuado para su gestión. Posteriormente, se implementaron y validaron técnicas de inteligencia artificial, con el fin de seleccionar la que mejor se ajustara al modelo de datos proporcionado. Por último, se propuso el desarrollo de un prototipo de aplicación web, que brindara al equipo técnico de la AMR acceso en cualquier momento a los datos y la capacidad de gestionar los planes de acción de acuerdo con los indicadores establecidos. A pesar del volumen considerable de datos, las aplicaciones de la ciencia de datos en el sector minero aún están en discusión, existiendo un vacío evidente en términos de buenas prácticas y métodos de predicción, entre otros aspectos. Como resultado, este proyecto no solo generó aplicaciones prácticas para la AMR, sino que también sentó para la organización, las bases para futuras investigaciones y aplicaciones en el campo de la ciencia de datos.

Pontificia Universidad
JAVERIANA
Cali

SISTEMA EMPRESARIAL PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE SOSTENIBILIDAD EN ORGANIZACIONES MINERAS

Bryan Josser Fernández Ebrath
Código 8972772
Jhon Enrique Ramírez Mantilla
Código 8971663
Andrés Felipe Ortega Lucero
Código 8971663

Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos

Director(a)
Martín Vladimir Alonso Sierra Galvis

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, MAYO 29 DE 2023

TABLA DE CONTENIDO

1. INTRODUCCIÓN 7
2. DEFINICIÓN DEL PROBLEMA 9
2.1. PLANTEAMIENTO DEL PROBLEMA 9
2.2. FORMULACIÓN DEL PROBLEMA 10
3. OBJETIVOS DEL PROYECTO 12
3.1. OBJETIVO GENERAL 12
3.2. OBJETIVOS ESPECÍFICOS 12
4. MARCO TEÓRICO Y ANTECEDENTES 13
4.1. MARCO TEÓRICO 13
4.1.1. Entendimiento del negocio 13
4.1.2. Metodología de inteligencia de negocios 15
4.1.3. Metodología ágil de desarrollo de software 16
4.2. ANTECEDENTES 19
4.2.1. Gestión, Predicción y Visualización de Datos 20
4.2.2. Innovaciones en la minería sostenible alineadas a la trazabilidad de cadenas de suministro de minerales 21
5. RESULTADOS 23
5.1. DESCRIPCIÓN DE LAS FUENTES DE DATOS 23
5.2. DISEÑO DE LA BASE DE DATOS 25
5.3. ANÁLISIS EXPLORATORIO DE DATOS 29
5.3.1. Análisis de dimensiones y descripción estadística 30
5.3.2. Análisis de relaciones entre variables 37
5.4. APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL 43
5.4.1. Descripción de conjunto de datos 43
5.4.2. Revisión de técnicas implementadas y resultados 44
5.4.3. Resultados del modelo de regresión ponderada 49
5.5. APLICATIVO WEB PARA LA GESTIÓN DE DATOS 52
5.5.1. Equipo Scrum 52
5.5.2. Componentes BACKEND 53

5.5.3. Componentes FRONTEND 54
5.5.4. Tablero de control 60
6. CONCLUSIONES 62
7. RECOMENDACIONES 63
8. BIBLIOGRAFÍA 64
9. ANEXOS 67

LISTA DE FIGURAS

Figura 1 - Cadena de suministros de minerales. 14
Figura 2 - Esquema del modelo CRISP-DM [8]. 15
Figura 3- Metodología SCRUM. 18
Figura 4 - Diseño de tablas en la base de datos. 28
Figura 5 - Conteo de criterios por tipo de Estándar y Categoría. 31
Figura 6 - Conteo de auditorías realizadas por las dimensiones país, año-mes (fecha de auditoria), mina certificada, región, tipo de mina y de acuerdo al uso de mercurio. 31
Figura 7 - Análisis Exploratorio de datos. Diagrama de cajas. 34
Figura 8 - Media de la calificación por Mina. 35
Figura 9 - Media de la calificación por dimensiones País, Región, Estándar y Categoría. 36
Figura 10 - Histograma de frecuencia - Presupuesto ARM. 36
Figura 11 - Histograma de frecuencia - Tiempo Dedicación Personal. 37
Figura 12 - Mapa de calor de la matriz de correlación. 38
Figura 13 - Presupuesto ARM y total estimado por Mina. 39
Figura 14 - Distribución de criterios evaluados por categoría y estándar, comparados contra el promedio de presupuesto total $MM por categoría (línea azul claro). 40
Figura 15 - Diagrama de dispersión entre variable objetivo, Presupuesto ARM, y variables predictoras, Calificación parcial y Tiempo de dedicación de personal. 41
Figura 16 - Histograma de frecuencia - Calificación Parcial. 43
Figura 17 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos mediante el modelo de regresión lineal con regularización Ridge. 47
Figura 18 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de árbol de decisión. 48
Figura 19 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de máquina de soporte vectorial - SVR. 49
Figura 20 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de regresión ponderada. 50
Figura 21 - Ejemplo algoritmo de modelo de regresión ponderada. 52
Figura 22 – Prueba Backend en Insomnia. 54
Figura 23 - Modulo Login. 55
Figura 24 - Módulo del tablero de control. 56
Figura 25 - Módulo de usuario. 57
Figura 26 - Módulo de Mina. 57
Figura 27 - Módulo criterio mina. 58
Figura 28 - Módulo actividad. 59
Figura 29 - Módulo de predicción presupuesto ARM. 60

LISTA DE TABLAS

Tabla 1 - Equipo Scrum. 19
Tabla 2 - Cantidad de criterios evaluados por cada indicador de sostenibilidad y estándar. 24
Tabla 3 - Entidades fuertes y débiles del modelo Entidad-Relación. 26
Tabla 4 - Descripción del conjunto de datos previo al análisis exploratorio. 30
Tabla 5 - Descripción estadística 33
Tabla 6 - Descripción estadística en campos de Presupuesto 33
Tabla 7 - Factor de inflación de varianza VIF 41
Tabla 8 – Coeficientes de correlación Pearson y Análisis de significancia valor-p. 42
Tabla 9 – Descripción del conjunto de datos para implementación de técnicas de IA 44
Tabla 10 - Medidas de rendimiento. 45
Tabla 11 - Resultados modelo de regresión ponderada. 51
Tabla 12 - Roles dentro del desarrollo del producto. 53

LISTA DE ANEXOS

1. ANEXO 1. Diccionario de datos.
2. ANEXO 2. Modelo entidad relación.
3. ANEXO 3. Script de la base de datos.
4. ANEXO 4. Metodología Ágil.
5. ANEXO 5. Consulta SQL conjunto de datos para análisis exploratorio.
6. ANEXO 6. Algoritmos implementación técnicas de inteligencia artificial.

1. INTRODUCCIÓN

La Alianza por la Minería Responsable es una ONG que promueve la implementación de prácticas responsables por parte de pequeños mineros de oro alrededor del mundo facilitándoles la vinculación al mercado formal y garantizando así precios más justos. Sin embargo, las exigencias del mercado han ido incrementando, en el sentido de solicitar, con información fundamentada en datos, que estas buenas prácticas realmente se están presentando y a su vez manteniendo en el tiempo. Alineado con esto, los donantes de proyectos de cooperación internacional para apoyar a un grupo minero específico fueron más exigentes con La Alianza por la Minería Responsable, queriendo conocer con mayor precisión los posibles destinos de los recursos y el impacto en cada categoría de sostenibilidad. Bajo esta solicitud, la Alianza por la Minería Responsable evidencia la necesidad de contar con un sistema de gestión de datos donde se pueda canalizar todos los esfuerzos realizados por parte del equipo técnico en campo, tiempos de dedicación y presupuesto, así como las mejoras logradas por los grupos mineros en las diferentes categorías de sostenibilidad, y presentar esta información de manera concisa al mercado internacional para continuar incentivando el abastecimiento de estos grupos mineros de pequeña escala.

Para resolver el problema fue necesario utilizar varios componentes de ciencia de datos, desde gestión de bases de datos y consultas en lenguaje SQL, técnicas de análisis exploratorio de datos con Python, Javascript como lenguaje para la construcción del aplicativo web, técnicas de inteligencia artificial y herramientas para la visualización de datos que permitan un mejor aprovechamiento de estos.

Primero se analizaron todas las fuentes de datos existentes por parte de la Alianza por la minería responsable tanto de implementación en campo, como de profesionales, tiempos y presupuestos. Todos los datos existentes, se encontraban en archivos de Excel diferentes, sin unificar en una base de datos como tal y bajo diferentes formatos, inclusive cuando recolectaban la misma información. El primer paso fue ordenar la información, determinar los componentes claves y relaciones entre sí, con lo cual se estructuró un modelo relacional en el gestor PostgreSQL con todas las entidades, variables, relaciones y cardinalidades, lo cual fue validado con las directivas de la Alianza. Una vez definido el modelo relacional, se procedió al cargue de los datos bajo cada entidad y variables definidas, considerando que todos los datos estaban en hojas de Excel. Paralelamente, se construyó el aplicativo web para la gestión de los datos donde se buscó permitir a la organización un sistema amigable con el usuario final para reportar todos los nuevos datos generados en las auditorías a los grupos mineros. El aplicativo web se encuentra conectado a la base de datos, y adicionalmente, mediante Power BI, se construyó la visualización de estos, la cual fue embebida al aplicativo, visualización que permitió dar respuesta a las preguntas del contexto de negocio, enfocadas principalmente en abordar las inquietudes previamente planteadas por el mercado internacional. Adicionalmente, se habilitó un módulo de predicción en el aplicativo web, donde los usuarios pueden subir en un formato preestablecido los datos de tiempo de dedicación

del personal profesional y obtener de forma rápida la predicción del presupuesto a utilizarse para la implementación de apoyo a un grupo minero considerando los tiempos previamente destinados por cada profesional, responsable de cada temática de sostenibilidad.

Como resultados del proyecto se tiene un prototipo de sistema empresarial para la gestión y análisis de datos de sostenibilidad en organizaciones mineras vinculadas con la Alianza por la Minería Responsable, el cual contiene las funcionales específicas para almacenar nuevas fuentes de datos, presentar visualizaciones de desempeños y un módulo adicional de la predicción de los presupuestos para nuevos proyectos a implementarse.

2. DEFINICIÓN DEL PROBLEMA

2.1. PLANTEAMIENTO DEL PROBLEMA

En las últimas décadas, se ha observado un incremento exponencial en la cantidad de datos disponibles, lo cual ha generado un alto volumen de información y ha abierto nuevas oportunidades para el análisis de tendencias y toma de decisiones orientadas mediante tecnologías de vanguardia. En el campo de la ciencia de datos y el aprendizaje automático, se ha reconocido que el valor de los datos no solo radica en los datos individuales, sino en las correlaciones y patrones subyacentes que pueden descubrirse. Estos patrones pueden proporcionar información más precisa, identificar tendencias relevantes y revelar asociaciones ocultas. Sin embargo, el tratamiento eficiente y efectivo de esta enorme cantidad de datos plantea importantes desafíos.

Uno de los desafíos más relevantes en la ciencia de datos es la necesidad de trabajar con datos provenientes de diversas fuentes y enfoques, lo que abarca el manejo de grandes volúmenes de datos (por ejemplo, Big Data), la extracción de patrones y comportamientos mediante técnicas de minería y procesamiento de datos, y la visualización clara y comprensible de la información obtenida [1]. Estas tres áreas, en conjunto, permiten un análisis integral y significativo de los datos.

En el ámbito específico de la industria minera a nivel global, se ha observado un incremento considerable en la producción de minerales. Paralelamente, ha surgido la necesidad de implementar estándares de sostenibilidad para rastrear el origen de los minerales y evaluar el grado de responsabilidad ambiental, social y de gobernanza en los procesos de extracción y producción. En respuesta a esta necesidad, la Unión Europea ha establecido regulaciones que obligan a los importadores (refinadores internacionales) de minerales a cumplir con la "Guía de Debida Diligencia para Cadenas de Suministro Responsables de Minerales en Áreas de Conflicto o de Alto Riesgo" desarrollada por la Organización para la Cooperación y el Desarrollo Económico (OCDE). Estas regulaciones exigen la evaluación de proveedores de minerales en aspectos como el lavado de activos, el crimen organizado, el trabajo forzado e infantil, así como los impactos ambientales [2][3].

En este sentido, la fundación Alianza por la Minería Responsable (una ONG colombiana con alcance global que tiene como objetivo transformar la minería artesanal y de pequeña escala en una actividad social y ambientalmente responsable) ha creado estándares internacionales como el Código para la Mitigación de Riesgos en la Minería Artesanal y de Pequeña Escala, Formando Cadenas Transparentes y Legales (CRAFT, por sus siglas en inglés), basado en la Guía de Debida Diligencia de la OCDE. Además, han desarrollado el estándar FAIRMINED, cuyo objetivo es crear

oportunidades para los mineros artesanales y de pequeña escala de oro, promoviendo su organización progresiva, mejores condiciones laborales, gestión ambiental y contribución al bienestar de las familias y la comunidad, mediante el acceso a mejores mercados que reconocen un precio justo y brindan un premio económico para invertir en mejoras continuas y en el desarrollo social [4][5].

La Alianza por la Minería Responsable brinda apoyo a diversas organizaciones mineras en Latinoamérica y el Caribe, incluyendo países como Colombia, Ecuador, Perú, Guyanas, Suriname y Honduras. A pesar de este respaldo, surge un desafío significativo en el manejo de los datos recopilados en el terreno, así como de los criterios de sostenibilidad relacionados (ambientales, comerciales, laborales, de género, legitimidad, organizativos, de negocios, seguridad y salud, sociales y de trazabilidad). Actualmente, dichos datos se gestionan de forma manual, sin un sistema sistematizado ni una unificación efectiva, empleando hojas de cálculo independientes que carecen de una conexión entre ellas. Además, los informes periódicos también se elaboran manualmente, lo cual demanda una cantidad considerable de tiempo al equipo encargado y podría ser automatizado considerando que los datos se rigen por estándares comunes en cada grupo minero.

Esta situación dificulta el proceso de lectura, visualización continua, filtrado y análisis dinámico de los datos, así como la generación de informes para respaldar una toma de decisiones más informada en la organización. No se puede medir adecuadamente el progreso o los avances en las distintas categorías, ni realizar comparaciones significativas entre los grupos de cada país. Tampoco se pueden identificar con precisión las categorías más influyentes o aquellas que presentan mayores rezagos en su implementación. Abordar este problema resulta crucial para potenciar el valor del trabajo desarrollado. Desde la perspectiva de la ciencia de datos, se presenta una oportunidad para aplicar técnicas y metodologías específicas que permitan resolver las particularidades mencionadas y superar los desafíos presentes en la gestión de datos en este contexto.

2.2. FORMULACIÓN DEL PROBLEMA

De acuerdo con el planteamiento del problema, se realizó la siguiente formulación: ¿Cómo desarrollar una herramienta que permita la gestión y análisis de datos de desempeños en categoría de sostenibilidad de las organizaciones mineras vinculadas con la Alianza por la Minería Responsable?

Las preguntas de sistematización del problema son las siguientes: ¿Como preparar y depurar los datos de categorías de sostenibilidad suministrada por la empresa?, ¿cómo diseñar una base de datos a partir de los datos previamente depurados?, ¿cuáles técnicas de inteligencia artificial se

pueden utilizar para descubrir o predecir comportamientos o patrones de las categorías de sostenibilidad?, ¿cómo desarrollar el prototipo de aplicación web para la gestión de los datos?

3. OBJETIVOS DEL PROYECTO

3.1. OBJETIVO GENERAL

Desarrollar un prototipo de aplicación web que permita la gestión y análisis de datos de desempeño en categoría de sostenibilidad de las organizaciones mineras vinculadas con la Alianza por la Minería Responsable.

3.2. OBJETIVOS ESPECÍFICOS

- Preparar y depurar la información de categorías de sostenibilidad suministrada por la empresa para conocer los datos a procesar.
- Diseñar una base de datos a partir de la información previamente depurada para el almacenamiento de los datos.
- Obtener modelos de predicción mediante técnicas de inteligencia artificial para descubrir comportamientos o patrones.
- Desarrollar el prototipo de aplicativo web para la gestión de los datos.

4. MARCO TEÓRICO Y ANTECEDENTES

4.1. MARCO TEÓRICO

4.1.1. Entendimiento del negocio

La Alianza por la Minería Responsable es una organización colombiana que cuenta con 20 años de experiencia en el sector de la minería artesanal y de pequeña escala a nivel global donde su principal objetivo es vincular a los pequeños mineros (principalmente de oro) con mercados internacionales, destacándose refinadores suizos y la industria joyera americana y europea, quienes demandan cada vez más por el abastecimiento de productos que contemplen el cumplimiento de los derechos humanos, y políticas sociales y ambientales. Para este fin la Alianza apoya en la apropiación por parte de los mineros de prácticas responsables que a su vez están alineadas con estándares internacionales y que el cumplimiento de estos les facilita la inclusión del metal precioso en las cadenas de suministro formales bajo sellos de certificación de producción que traen consigo un precio significativamente superior al que comúnmente se consigue en el sector minero informal. En este sentido, el desarrollo de prácticas mineras responsables involucra la ejecución de una minería sostenible con el medio ambiente y con el entorno en general.

Sostenibilidad minera. de acuerdo con Roderick Eggert (experto internacional en temas de sostenibilidad minera y profesor emérito de la Escuela de Colorado de Minas, USA) el término sostenibilidad nace en el campo de la gestión de recursos renovables, partiendo de que el consumo de estos recursos debería ser gestionado de tal manera que no excede la razón misma de su regeneración, manteniendo así indefinidamente las reservas. Sin embargo, al ser los minerales recursos no renovables, es decir, que una vez explotados no pueden regenerarse, Eggert plantea abordar el término sostenibilidad minera desde cuatro principios fundamentales, los cuales son [6]:

1. Facilitar la medición del recurso mineral para garantizar el acceso al mismo en el largo plazo que permita contribuir a la sostenibilidad, considerando que es un recurso no renovable.

2. Asegurar que el proceso de extracción y procesamiento mineral se desarrolla de una manera social y ambientalmente eficiente.

3. Distribuir justamente los beneficios de la actividad minera de tal manera que ayuden a disminuir la pobreza.

4. mantener los beneficios de la actividad minera, incluso después de haber finalizado el proceso extractivo.

El presente proyecto fue enfocado en abordar una solución por medio de un sistema web empresarial, donde la gestión de los datos influye principalmente sobre el segundo principio de sostenibilidad minera referente a "asegurar que el proceso de extracción y procesamiento mineral se desarrolla de una manera social y ambientalmente eficiente", considerando que los datos suministrados por la Alianza por la Minería Responsable contempla una base de datos con indicadores que permiten medir el nivel de responsabilidad de las organizaciones mineras. Una vez se logra cumplimiento de buenas prácticas que puedan ser certificadas por estándares internacionales, se abre el mercado formal para los grupos mineros y se contempla el concepto de cadena responsable de suministro de minerales.

Cadena responsable de suministro de minerales. Consiste en la generación de impactos positivos y minimización de los efectos negativos a lo largo de toda la cadena de valor de una compañía, en este caso de las organizaciones mineras. Para esto es necesario incluir la evaluación y aplicación de criterios ambientales, sociales y de gobernanza en las actividades de explotación y procesamiento de minerales como tal, lo cual impacta positivamente la reputación que la organización minera tiene en la parte baja de la cadena de valor donde convergen las industrias refinadoras, transformadores (para el caso específico del oro, empresas de joyería) y los clientes finales como tal [7].

Para un mejor entendimiento de la cadena de valor en el sector minero, se presenta el siguiente esquema:

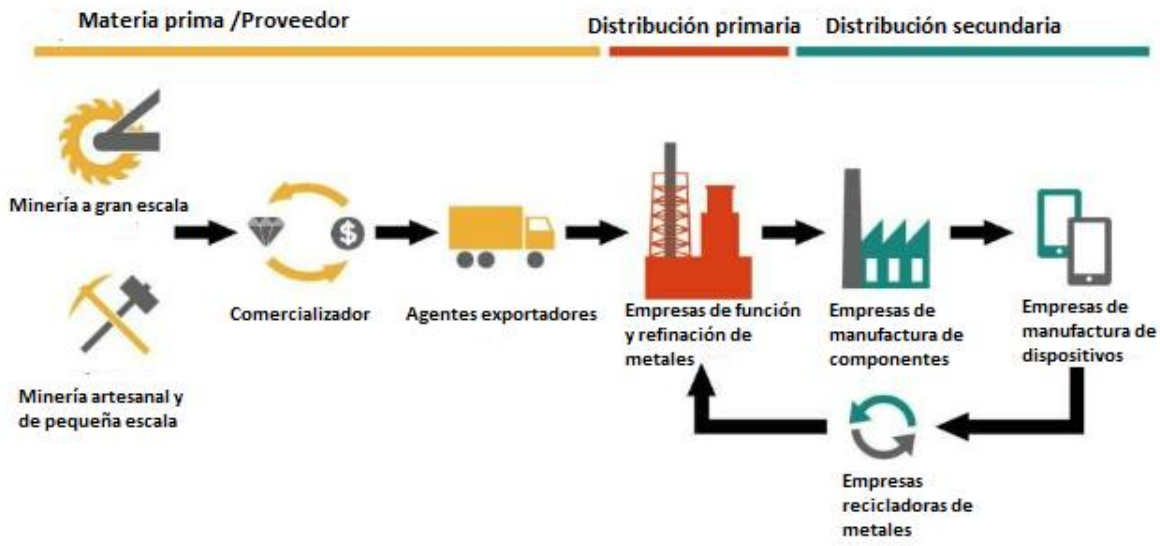


Figura 1 - Cadena de suministros de minerales [7].

4.1.2. Metodología de inteligencia de negocios

La inteligencia de negocios aplicada a la sostenibilidad de grupos mineros implica el uso de tecnologías avanzadas y herramientas de análisis de datos para garantizar el cumplimiento de los estándares internacionales para la exportación de minerales, al mismo tiempo que se asegura la sostenibilidad de la actividad minera en el largo plazo, lo que permite la toma de decisiones informadas y eficientes en relación con la gestión de recursos, la reducción de costos, la mitigación de riesgos ambientales y la mejora de la eficiencia energética. Además, la inteligencia de negocios puede ayudar a los grupos mineros a monitorear y cumplir con los estándares internacionales de sostenibilidad, lo que puede mejorar la reputación de la empresa y su capacidad para acceder a los mercados internacionales de minerales, al tiempo que se promueve el bienestar social y ambiental en las comunidades mineras.

Para el desarrollo de este proyecto, se tuvo en cuenta como guía la metodología de inteligencia de negocios CRISP-DM (Cross Industry Standard Process for Data Mining), por sus siglas en inglés, traducido al español como “Proceso Estándar Transversal para la Minería de Datos”. Teniendo como referencia la guía de aplicación definida por J. Han y M. Kamber [8], una referencia ampliamente utilizada en el campo de la minería de datos y la cual proporciona una guía detallada sobre las diferentes etapas de la metodología CRISP-DM, desde la comprensión del negocio y la recolección de datos hasta la evaluación y la implementación de los resultados.

CRISP-DM es una metodología basada en la práctica y experiencia real de analistas DM que han contribuido activamente al desarrollo de esta. CRISP-DM se organiza en fases, procesos, documentos entregables y actividades. El proceso se detalla en la figura 2.

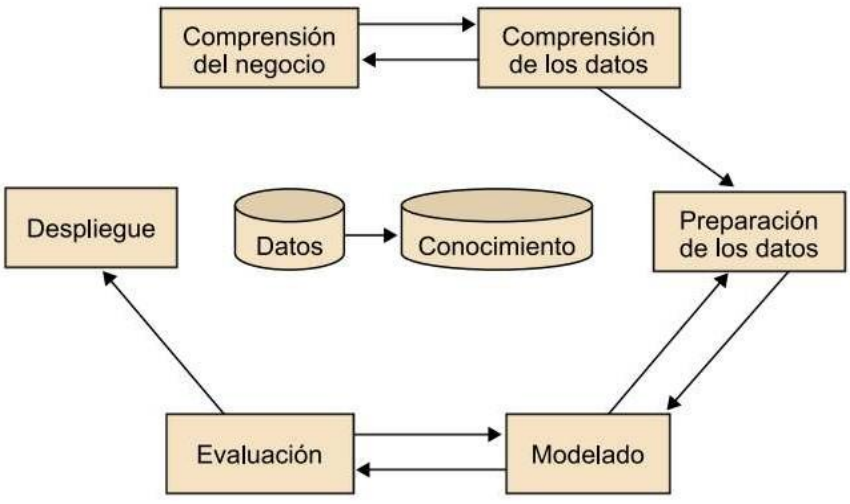


Figura 2 - Esquema del modelo CRISP-DM [8].

4.1.3. Metodología ágil de desarrollo de software

El término ágil surge como iniciativa de un conjunto de expertos en el área de desarrollo de software con el fin de optimizar el proceso de creación de este, el cual era caracterizado por ser rígido y con mucha documentación [9].

La metodología ágil también enuncia los doce principios que caracterizan un proceso ágil diferenciándolo de otro tradicional donde este enfoque no se había aplicado lo suficiente; siempre se había dejado implícito, pero sin hacer hincapié en ellos. El objetivo de un proyecto ágil no es controlar la ejecución para garantizar el cumplimiento de la planificación, sino proporcionar de forma continua el mayor valor posible al producto. Resulta por tanto más adecuada una relación de implicación y colaboración continua con el cliente, que una contractual de delimitación de responsabilidades.

Los 12 principios del manifiesto ágil. El manifiesto ágil, tras los postulados de estos cuatro valores en los que se fundamenta, establece estos 12 principios [9]:

1. Nuestra principal prioridad es satisfacer al cliente a través de la entrega temprana y continua de software de valor.
2. Son bienvenidos los requisitos cambiantes, incluso si llegan tarde al desarrollo. Los procesos ágiles se doblegan al cambio como ventaja competitiva para el cliente.
3. Entregar con frecuencia software que funcione, en periodos de un par de semanas hasta un par de meses, con preferencia en los periodos breves. Las personas del negocio y los desarrolladores deben trabajar juntos de forma cotidiana a través del proyecto.
4. Construcción de proyectos en torno a individuos motivados, dándoles la oportunidad y el respaldo que necesitan y procurándoles confianza para que realicen la tarea.
5. La forma más eficiente y efectiva de comunicar información de ida y vuelta dentro de un equipo de desarrollo es mediante la conversación cara a cara.
6. El software que funciona es la principal medida del progreso.
7. Los procesos ágiles promueven el desarrollo sostenido. Los patrocinadores, desarrolladores y usuarios deben mantener un ritmo constante de forma indefinida.
8. La atención continua a la excelencia técnica enaltece la agilidad.
9. La simplicidad como arte de maximizar la cantidad de trabajo que no se hace, es esencial.
10. Las mejores arquitecturas, requisitos y diseños emergen de equipos que se autoorganizan.

11. En intervalos regulares, el equipo reflexiona sobre la forma de ser más efectivo y ajusta su conducta en consecuencia.

Modelo Scrum. Scrum es un modelo de desarrollo ágil caracterizado por:

- Adoptar una estrategia de desarrollo incremental, en lugar de la planificación y ejecución completa del producto.
- Basar la calidad del resultado más en el conocimiento tácito de las personas en equipos autoorganizados, que en la calidad de los procesos empleados.
- Solapamiento de las diferentes fases del desarrollo, en lugar de realizarlas una tras otra en un ciclo secuencial o de cascada.

Scrum se basa en la teoría de control de procesos empírica o empirismo. El empirismo asegura que el conocimiento procede de la experiencia y de tomar decisiones basándose en lo que se conoce. Esta metodología emplea un enfoque iterativo e incremental para optimizar la predictibilidad y el control del riesgo, realiza entregas del proyecto en sí, figura 3 [10].

La metodología Scrum describe cuatro eventos importantes que componen cada una de las entregas:

- Reunion de planificación del sprint (Sprint Planning Meeting).
- Scrum Diario (Daily Scrum).
- Revision del Sprint (Sprint Review).
- Retrospectiva del Sprint (Sprint Retrospective).

Figura 3- Metodología SCRUM [10].

Pila del producto. La pila del producto es la lista ordenada de todo aquello que el propietario de producto cree que necesita el producto. Es el inventario de funcionalidades, mejoras, tecnología y corrección de errores que deben incorporarse al producto a través de los sucesivos Sprints. Representa todo aquello que esperan el cliente, los usuarios, y en general los interesados. Todo lo que suponga un trabajo que debe realizar el equipo debe estar reflejado en esta pila. La pila del producto nunca se da por completada; está en continuo crecimiento y evolución. Al comenzar el proyecto incluye los requisitos inicialmente conocidos y mejor entendidos, y evoluciona conforme avanza el desarrollo.

Sprint. El evento clave de scrum para mantener un ritmo de avance continuo es el sprint: el periodo de tiempo acotado ("time-box") de duración máxima de 4 semanas, durante el que se construye un incremento del producto. El incremento realizado durante el sprint debe estar terminado, esto es: completamente operativo y útil para el cliente, en condiciones de ser desplegado o distribuido.

Pila de Sprint. La pila del sprint (sprint Backlog) es la lista de las tareas necesarias para construir las historias de usuario que se van a realizar en un sprint. La pila Scrum es confeccionada por el equipo en la reunión de planificación del sprint, indicando para cada tarea el esfuerzo previsto

para realizarla. Para calcular el esfuerzo de cada tarea (en puntos o tiempo ideal), la pila del sprint descompone las historias de usuario en unidades de tamaño adecuado para monitorizar el avance a diario e identificar riesgos y problemas sin necesidad de procesos de gestión complejos. Es también, una herramienta para la comunicación visual directa del equipo.

Equipo Scrum. El equipo Scrum se detalla en la tabla 1, a continuación.

Tabla 1 - Equipo Scrum.

Rol	Responsabilidad
Scrum Máster	Es un líder que está bajo el servicio del equipo scrum, este miembro ayuda al equipo y a los clientes externos a comprender las interacciones que pueden ser de ayuda y cuáles no lo son, además él es el encargado de asegurar que el equipo adopte las teorías, prácticas y reglas de la metodología scrum
Gestor del producto	Es la persona responsable de transmitir al equipo de desarrollo la visión del producto que se desea crear, aportando la perspectiva de negocio
Desarrollador	Responsable de desarrollar y entregar el producto. Mantiene una organización horizontal en la que cada miembro del equipo se auto-gestiona y organiza libremente en la definición y ejecución de los distintos sprints
Verificador	Es la persona responsable de realizar las pruebas necesarias al software o aplicativo.

4.2. ANTECEDENTES

A continuación, se dan a conocer algunos trabajos relacionados tanto en materia de ciencia de datos como del contexto del problema, los cuales fueron referentes para la estructuración del proyecto.

4.2.1. Gestión, Predicción y Visualización de Datos

Diseño de un mecanismo de medición basado en tablero de control para la gestión de proyectos en la empresa Energizando S.A.S [11]. El proyecto mencionado se basó en la utilización de herramientas de visualización de datos con PowerBI debido a la problemática de la empresa Energizando S.A.S, la cual presentaba anomalías en los sistemas de medición como los indicadores claves de desempeños, además, no se habían implementado estrategias, actividades o tableros de control para obtener métricas de desempeño. Para el desarrollo de la propuesta, se diseñó un modelo relacional de base de datos a partir de fuentes de datos en archivos de Excel, se implementó la herramienta de Power BI bajo la metodología de abordaje Value-Chain Data Approach (basado en la cadena de valor de BI, focalizándose en los datos que generaran mayor valor para el negocio) generando un tablero de control conectado con el motor de base de datos y así, realizar un análisis general en la cual se aplican filtros por proyectos, fechas, costos, satisfacción. Como resultado se tiene un tablero de control que les permitió visualizar los datos, sus relaciones, calcular indicadores y brindar una visión más integral de los procesos y la compañía en general ya que son clave en el desarrollo empresarial de hoy, y constituye la base para la gestión estructurada de los procesos, áreas y equipos de negocio. Esta propuesta o antecedente fue base fundamental para la creación del tablero de la organización minera ya que, brindó modelos y ejemplos pertinentes para tener en cuenta al momento del diseño final de las visualizaciones con los datos de la organización minera.

Diseño de un tablero de control que permita la identificación temprana de la corrosión externa en los ductos de transporte de gas natural [12]. El objetivo de esta investigación fue la creación de una herramienta de visualización de la información relacionada a la corrosión (reacción electroquímica) que ha generado fallas en las líneas de transporte. Los mecanismos de protección para la corrosión externa poseen herramientas de inspección, con las cuales se monitorean la eficiencia de los sistemas de protección catódica y el recubrimiento. Como resultado de realizar las inspecciones se genera una gran cantidad de datos, los cuales no están procesados y representados para generar un análisis de estado actual en el que se encuentra la línea. Por lo anterior, se creó una herramienta informática de geoprocesamiento que permite representar los datos obtenidos para el control de la corrosión en los ductos haciendo posible la visualización de cada punto de inspección por medio de mapas y gráficos desarrollados con el lenguaje de programación Python. Esta herramienta proporciona la gestión de validar si una línea se encuentra protegida, sobreprotegida o desprotegida aumentando la vida útil de la misma y evitando situaciones de riesgo, ocasionadas por fugas debido a un control ineficiente de la corrosión externa. La propuesta anterior permitió dentro del marco del presente proyecto, conocer un poco más el diseño de tableros como herramienta de análisis y visualización de datos,

facilitando la gestión de seguimiento en actividades similares a las que abordamos en el sistema web para el manejo de datos de sostenibilidad minera.

Desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0 [13]: La propuesta de esta investigación tuvo como finalidad el monitoreo predictivo de los procesos de manufacturas para pronosticar el comportamiento del sistema. Anteriormente, se habían propuesto varias técnicas basadas en análisis de datos históricos, sin embargo, la investigación está orientada al proceso de manufactura de la industria 4.0. El sistema se implementó en los laboratorios de aprendizaje de manufactura de la Pontificia Universidad Javeriana de Bogotá bajo el modelo de desarrollo CRISP-DM (CrossIndustry Standard Process for Data Mining) utilizando el lenguaje de programación Python. Después de hacer la comparación de los resultados de las métricas de desempeño que se necesitaban, se realizó la comparación de diferentes modelos teniendo como resultado a elegir el modelo de agrupación. Finalmente, con base en el conjunto de datos históricos y la finalidad de las visualizaciones, se diseña un tablero de control monitoreo predictivo utilizando la librería Dash de Python. Con base a esta propuesta de investigación, se pudo conocer como es el proceso de aplicar la metodología CRISP-DM en proyecto relacionados a ciencia de datos específicamente, a la organización minera.

4.2.2. Innovaciones en la minería sostenible alineadas a la trazabilidad de cadenas de suministro de minerales

Métodos de trazabilidad de cobalto, litio y grafito utilizados en la producción de baterías. "Evaluación de la huella dactilar geológica para la trazabilidad de las materias primas de las baterías" [14]: Este proyecto nace de la necesidad de trazar los metales a incorporarse en la producción de baterías con la finalidad de construir y fortalecer una cadena responsable de suministro de minerales en este sector. Como parte de la transición energética y para depender cada vez menos de los combustibles fósiles, se ha incorporado en los últimos años la manufactura de vehículos eléctricos donde las baterías de ion de litio es la tecnología más aplicada en el mundo (a pesar del nombre, éstas también incluyen otros minerales como níquel, cobalto, manganeso, entre otros). Varias soluciones para trazar la cadena de suministro de baterías están siendo lideradas por compañías de vehículos híbridos y eléctricos; este tipo de soluciones abarcan el manejo de grandes volúmenes de datos y desarrollos digitales; y procuran certificar el origen del mineral para aumentar la transparencia en la producción de baterías y reducir los impactos ambientales y humanos asociados al proceso extractivo. En este sentido se presenta una discusión del método de huella dactilar geológica desarrollado conjuntamente por el Servicio Geológico de

Finlandia y el Centro de Investigación Técnica de la empresa VTT del mismo país bajo el proyecto denominado BATTRACE.

Los métodos tradicionales de trazabilidad hacen referencia a implementación de formatos en campo sobre los procesos de empacado, sellado y etiquetado del mineral, sin embargo, estos formatos pueden ser fácilmente alterados sin generar ningún tipo de alerta; algo que se podría corregir si se logran mapear los datos geológicos (propiedades intrínsecas del mineral) de cada lugar de extracción y realizar procesos de analítica comparativos con la información obtenida del lingote de metal ya introducido en la cadena de valor; generando así patrones de concordancia o discordancia entre el origen del lingote en los documentos de exportación y el origen de acuerdo a su huella digital geológica.

BATTRACE fue fundamental en el desarrollo de este proyecto, porque nos permitió tener un mejor entendimiento del problema a nivel global, considerando la creciente necesidad de demostrar trazabilidad en la inclusión de metales en las cadenas de suministro de minerales, especialmente en el sector de vehículos eléctricos. Actualmente aún más con el auge de la transición energética que requiere sustancialmente un aumento significativo de explotación de minerales estratégicos para la manufactura de equipos y dispositivos para la generación de energías renovables. Por otro lado, cabe destacar que los enfoques de los dos proyectos abordan la misma problemática desde dos puntos diferentes, siendo nuestro proyecto enfocado en medir varias categorías de sostenibilidad bajo la metodología propuesta, y, por otro lado, estando asociado el proyecto BATRACCE enfocado únicamente en el tema de trazabilidad con datos geológicos, lo cual marca una gran diferencia con nuestro origen de datos, sin embargo, evidenciándose a su vez un gran potencial de articulación complementaria entre sí.

Cadena de bloques para la trazabilidad en las cadenas de suministros de minerales y metales, oportunidades y desafíos [15]. Cada vez más los actores de la parte baja de la cadena de valor (joyeros, refinadores, sector de vehículos eléctricos, entre otros) están demandando información detallada a sus proveedores de minerales, esto debido a las presiones globales para garantizar una explotación y procesamiento justo y transparente con el medio ambiente y comunidades. Esta información requerida se fundamenta principalmente en demostrar el origen del mineral y dar a conocer los métodos responsables de producción que están siendo utilizados. En este sentido, el Consejo Internacional de Minería y Metales (organización internacional dedicada a promover una industria minera y metalúrgica segura, justa y sostenible) presenta el reporte" Cadena de bloques para la trazabilidad en las cadenas de suministros de minerales y metales: oportunidades y desafíos" donde evalúa el sistema de cadena de bloques (conocido ampliamente como Blockchain) como una alternativa a la problemática presentada. Se plantea la cadena de bloques como una similitud a una base de datos que provee una plataforma donde pueden ser registradas todas las transacciones de la cadena de valor desde la mina hasta el consumidor final. La base de

datos incluye peso del lingote, cantidad, pureza, pero además información sobre el origen y certificados de producción responsable, los cuales pueden ser cargador a la plataforma como evidencias de soporte y posteriormente validados en cualquier punto de la cadena para posteriormente ser vinculados al lingote usando cualquier tipo de tecnología como códigos de barra, etiquetas, entre otros; información que puede ser compartida con externos y actores interesados en conocer más sobre el producto vinculado a la cadena de valor. Uno de los grandes retos identificados en este sentido, es que el sistema de cadenas de bloques funciona una vez se obtenga el metal, y no garantiza en realidad que ese metal fue generado en un punto geográfico específico, es decir, es una solución parcial al problema y se ubica en la parte media a la parte baja del esquema de la cadena de valor presentado en la sección 4.1.1.

Este proyecto de cadenas de bloques para la trazabilidad de mineral nos proporcionó un aporte significativo para el desarrollo de nuestro proyecto, considerando que involucra una solución digital enfocada en la parte comercial del lingote como tal, brindando soluciones para el cierre de la cadena de valor, sin embargo, también existentes diferencias significativas, considerando que nuestro proyecto se enfoca en el manejo de datos en temáticas de sostenibilidad enfocados en las etapas de explotación y procesamiento, donde posteriormente se involucra la parte de comercialización, y por lo cual pueden ser notablemente complementarios.

5. RESULTADOS

5.1. DESCRIPCIÓN DE LAS FUENTES DE DATOS

Fuentes de datos. Los datos suministrados por la ARM se alojan en un Sharepoint de Microsoft en línea donde se guarda una colección de archivos de Excel, cada uno correspondiente a una auditoría realizada a un grupo minero específico sobre una mina específica. Estos archivos corresponden a una plantilla predefinida por la organización donde se detalla información sobre los criterios evaluados, la calificación parcial obtenida, y las actividades de mejora recomendadas, así como información asociada a la mina (ubicación geográfica, nombre, certificación, extracción de minerales, cantidad de trabajadores, entre otros datos relacionados). Se cuenta con una base de 24 auditorias, donde en cada una se evaluaron 241 criterios.

Indicadores de sostenibilidad. Se han establecido nueve indicadores de sostenibilidad minera, los cuales se encuentran registrados en la base de datos bajo el campo "Categoría". Estos indicadores han sido evaluados cuantitativamente mediante calificaciones que van desde cero (0) hasta tres (3). Una calificación de cero representa un cumplimiento nulo del criterio evaluado, mientras que una calificación de tres indica un cumplimiento satisfactorio. En la tabla 2 se presenta el número

de criterios evaluados para cada uno de los indicadores de sostenibilidad (categorías). Estos datos permiten conocer la extensión y el alcance de la evaluación realizada en relación con cada indicador, brindando una visión más detallada de los aspectos considerados en la medición de la sostenibilidad minera.

Tabla 2 - Cantidad de criterios evaluados por cada indicador de sostenibilidad y estándar.

INDICADOR-CATEGORIA	CRAFT	FAIRMINED	TOTAL
Ambiental	22	20	42
Condiciones de empleo	1	19	20
Género	4	1	5
Legitimidad	7	6	13
Mercurio	6	5	11
Organizativo y de negocios	17	14	31
Seguridad y salud	23	9	32
Social	37	14	51
Trazabilidad	8	4	12
Comercial	0	24	24
TOTAL	**125**	**116**	**241**

Limpieza de datos. Equipo interno de la organización ARM, realizó previamente un proceso de limpieza a partir de los archivos fuente (hojas de cálculo de Excel). Esta limpieza fue necesaria para que la empresa pudiera realizar análisis de datos preliminares en este tipo de archivos, ya que se requerían informes específicos. Como resultado, al migrar los datos al gestor de base de datos, se obtuvo una versión limpia de los datos.

El proceso de limpieza realizado en los archivos fuente fue descrito por el equipo interno de la ARM, e incluyó las siguientes acciones:

- Eliminar datos duplicados o datos irrelevantes.
- Corrección de errores estructurales y de digitación: particularmente algunos criterios que no aplicaban para cierta organización minera eran calificados como N/A, n/a, na, nulos, etc. Lo cual era un formato no permitido dentro de la formulación de los archivos fuente.
- Identificación y filtro de valores atípicos: los criterios de la base de datos deben tener una calificación de cero a tres, algunos de ellos tenían calificaciones por fuera de este rango, lo cual tuvo que ser corregido para no afectar los resultados de desempeño.
- Manejo de datos faltantes: algunas calificaciones fueron identificadas como faltantes o vacías, donde se realizó un riguroso detalle y se complementó con la información

específica para cada criterio. (La empresa tiene un responsable de cada categoría y calificación, por lo cual este tipo de “errores” tiene un responsable específico encargado de solventar el problema).

- Validación y control: todos los archivos fuente de cada organización minera son validados al finalizar el proceso de limpieza, las hojas de cálculo generan gráficos de diferentes tipos los cuales están a su vez formulados con la data y son la salida fundamental para hacer la validación de la base de datos.

Estas medidas garantizaron la integridad y calidad de los datos al eliminar inconsistencias y asegurar la coherencia en la información, lo cual facilita el análisis y generación de informes precisos. Este proceso se llevó a cabo con cada archivo de las auditorías realizadas a los grupos mineros.

5.2. DISEÑO DE LA BASE DE DATOS

Considerando que los datos de las auditorías de las minas suministrados por la Alianza presentan diferentes versiones de formatos con una estructuración diferente, y sin un orden lógico, se realizó el proceso ETL de forma manual, dónde se pudo garantizar la consistencia de los datos y realizar el cargue al gestor PostgreSQL por medio de lotes, el cual es un sistema de gestión de bases de datos relacional con características relevantes como ser potente, multipropósito, que garantiza la integridad de los datos y multiplataforma [16].Por otro lado, la interfaz gráfica para conectarse al servidor PostgreSQL fue PGAdmin y se desarrolló como una base de datos relacional dinámica, lo que permite ingresar continuamente los datos provenientes de los archivos fuente, que son hojas de cálculo de Excel, así como, directamente desde el aplicativo web.

Cada mina cuenta con un archivo de Excel que contiene diferentes hojas de cálculo, las cuales almacenan los datos que se necesita guardar en la base de datos. A partir de las necesidades y requerimientos de la empresa, se realizó el diseño de la base de datos. El cual, como primera medida se plantea la creación de las entidades tomando como referencia, que esta es un objeto o concepto que puede identificarse en el mundo real por ejemplo país, región, mina, actividad entre otros. Además, se pueden clasificar como entidades fuertes cuya característica principal es que no depende de la existencia de otra entidad y entidades débiles que se identifican porque depende de la existencia de otras entidades [17]

A partir de lo anterior y, teniendo en cuenta que la base de datos es de tipo relacional, se definieron las entidades fuertes y débiles, las cuales se especifican en la tabla 3.

Tabla 3 - Entidades fuertes y débiles del modelo Entidad-Relación.

Entidades fuertes	Entidades débiles
País	Región
Estándar	Modulo
Categoría	Criterio
Perfil	Mina
Grupo	Usuario
	Seguimiento_criterio
	Detalle_mina
	Actividad
	Seguimiento_plan
	ESG
	ODS
	Criterio_ODS

Una vez establecido el modelo entidad-relación (anexo 2), se definió el diccionario de datos (anexo 1) que describe los tipos de datos para cada campo de las tablas, incluyendo los campos clave y su descripción.

Normalización. Se tuvieron en cuenta en el diseño de la base de datos las tres reglas de normalización, detalladas a continuación:

- En la primera forma normal, se aseguró que los campos tengan valores atómicos y atributos únicos, así como se configuraron las tablas para no permitir duplicidad en las celdas.
- En la segunda forma normal, se verificó que los atributos no clave dependan completamente de la clave primaria.
- Finalmente, en la tercera forma normal, se aseguró que todos los atributos no clave sean independientes entre sí y no proporcionen información sobre otros atributos de la relación [18]

La codificación de las tablas se realizó utilizando el lenguaje SQL con el motor de bases de datos PostgreSQL. El script correspondiente se encuentra disponible en el anexo 3.

Con este proceso de diseño de la base de datos, se logró una estructura sólida y eficiente para almacenar los datos de las organizaciones mineras, facilitando su gestión y análisis en el aplicativo

web, así como su usabilidad para la generación de conjuntos de datos para análisis más complejos. La estructura de la base de datos se presenta en la figura 4, esto desde la vista del diseño de las tablas y sus relaciones.

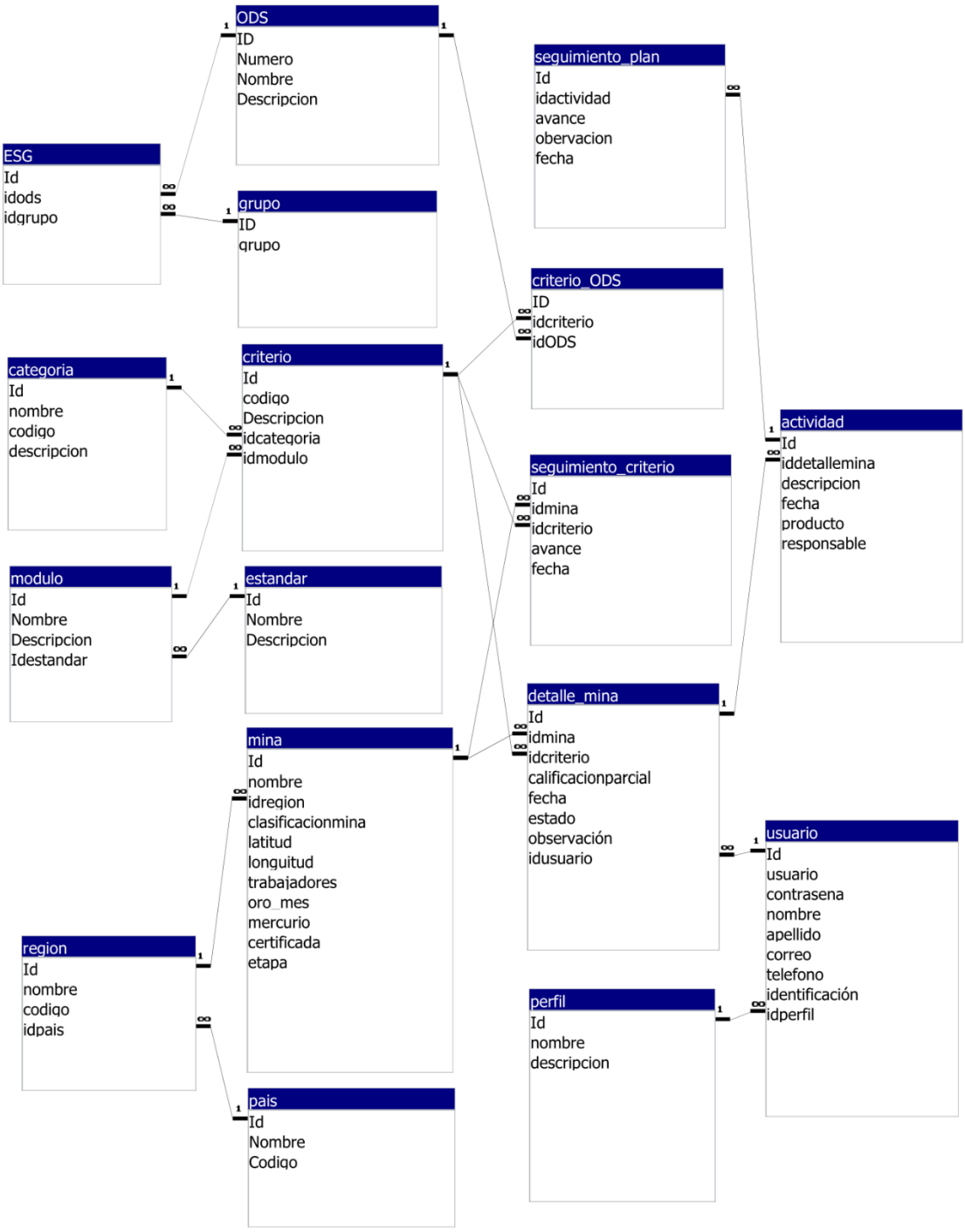


Figura 4 - Diseño de tablas en la base de datos.

5.3. ANÁLISIS EXPLORATORIO DE DATOS

A través de la aplicación de técnicas estadísticas y visuales, se buscó comprender la estructura y las propiedades inherentes de los datos (conjunto de datos generado mediante consulta SQL desde la base de datos creada, anexo 5), así como descubrir patrones, relaciones y tendencias relevantes. Durante esta fase, se examinó minuciosamente las distribuciones de los datos, se identificó posibles valores atípicos y se evaluó la calidad y la integridad de los datos. Además, se exploran las interacciones entre las variables y se generan hipótesis preliminares que pueden orientar el diseño de futuros modelos predictivos o descriptivos, principal objetivo de esta sección que abre lugar para el desarrollo del tercer objetivo de este proyecto, sobre la implementación de técnicas de inteligencia artificial.

El conjunto de datos generado mediante consulta SQL desde la base de datos, consta de 26 campos y un total de 7011 registros. En la tabla 4, se presenta un resumen de los campos, cantidad de registros no nulos y nulos, y tipo de dato, detallando que el conjunto de datos presenta valores nulos en los campos de producción de oro por mes, tonelada de mineral por mes, calificación parcial, y fecha de actividad. Así mismo, se cuenta con la siguiente distribución de tipos de datos: 11 de tipo categórico, 2 de tipo fecha, 7 de tipo decimal (float64), 3 de tipo entero (int64), y 3 campos compuestos por cadenas de texto.

Como se mencionó previamente en la sección 5.1, que se enfoca en la descripción de las fuentes de datos, se identificaron ciertos datos faltantes en las auditorías, tal como la producción de toneladas de oro por mes, la cual solo aplica a los grupos mineros que se dedican a la extracción de dicho mineral. De la misma manera, se observó que el campo de calificación parcial contenía criterios sin calificación debido a que no aplicaban para el caso estudio, por ejemplo, en aquellos sobre el uso de mercurio, puesto que no todos los grupos mineros hacen uso de éste. Estos registros donde la calificación parcial no aplicaba fueron eliminados del análisis de datos. Teniendo en cuenta lo anterior, el conjunto de datos que se empleó para el desarrollo de esta fase cuenta con un total de 6791 registros únicos.

Tabla 4 - Descripción del conjunto de datos previo al análisis exploratorio.

#	Campo	Registros No Nulo	Registros Nulo	Tipo de dato
0	pais	7011	0	Categórico
1	region	7011	0	Categórico
2	mina	7011	0	Categórico
3	clasificacionmina	7011	0	Categórico
4	trabajadores	7011	0	Entero - int64
5	oro_mes	5805	1206	Decimal - float64
6	tipo	7011	0	Categórico
7	tonelada	5270	1741	Decimal - float64
8	mercurio	7011	0	Categórico
9	etapa	7011	0	Categórico
10	certificada	7011	0	Categórico
11	estandar	7011	0	Categórico
12	modulo	7011	0	Categórico
13	categoria	7011	0	Categórico
14	idcriterio	7011	0	Entero - int64
15	criterio	7011	0	Cadena de caracteres
16	calificacionparcial	6791	220	Decimal - float64
17	observacion	7011	0	Cadena de caracteres
18	fechacalificacionparcial	7011	0	Fecha - datetime64[ns]
19	actividad	7011	0	Cadena de caracteres
20	fechaactividad	3451	3560	Fecha - datetime64[ns]
21	PresupuestoARM_$MM	7011	0	Decimal - float64
22	TiempoDedicacionPersonal	7011	0	Decimal - float64
23	TiempoCumplimientoActividad	7011	0	Entero - int64
24	PresupuestoMinaCategoria_$MM	7011	0	Decimal - float64
25	PresupuestoTotal_$MM	7011	0	Decimal - float64

5.3.1. Análisis de dimensiones y descripción estadística

Dimensiones. Cada dimensión representa una característica específica de los datos que se está analizando, para el análisis de datos en proceso se revisaron las diferentes dimensiones del conjunto de datos, generando las siguientes observaciones, sustentadas por las figuras 5 y 6:

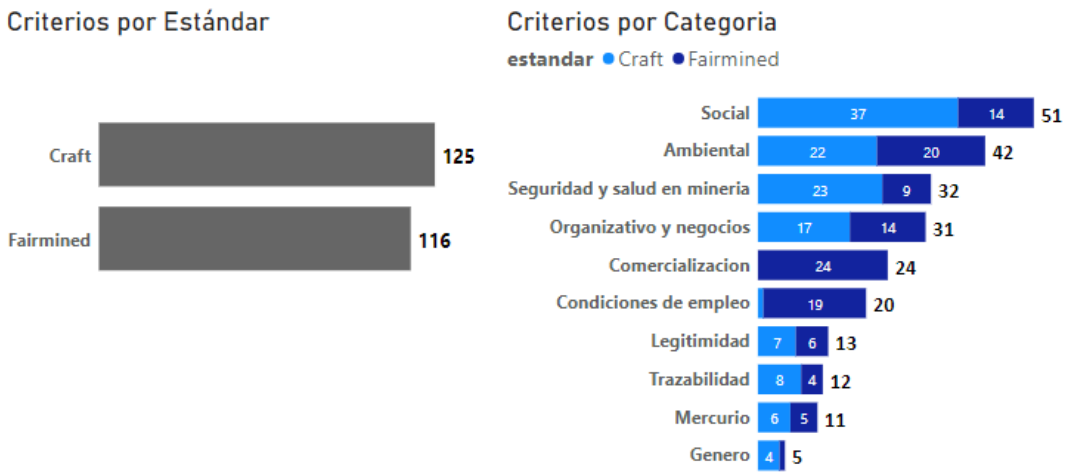


Figura 5 - Conteo de criterios por tipo de Estándar y Categoría.

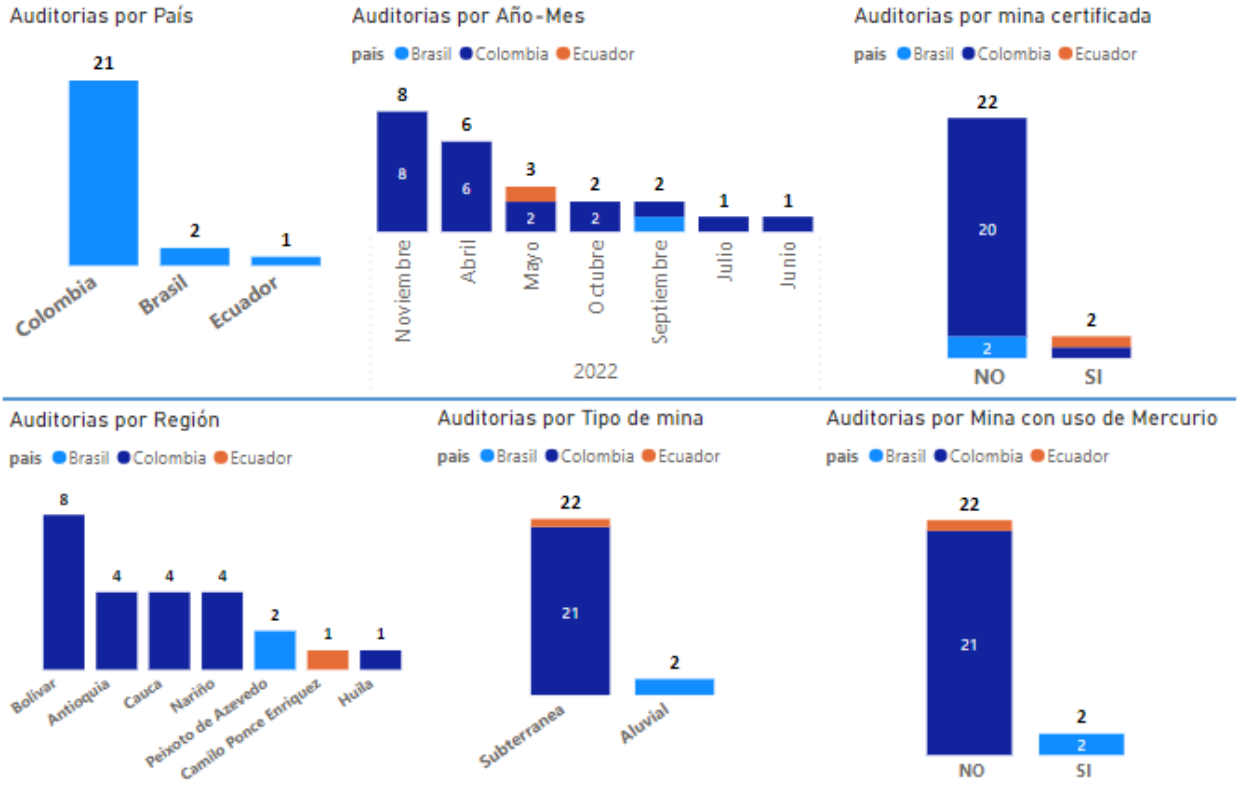


Figura 6 - Conteo de auditorías realizadas por las dimensiones país, año-mes (fecha de auditoria), mina certificada, región, tipo de mina y de acuerdo al uso de mercurio.

- El conjunto de datos contiene una única auditoria por mina y todas las auditorias fueron llevadas a cabo durante el año 2022, principalmente durante los meses de noviembre y abril.
- Las auditorias se componen en general de 241 criterios o preguntas únicas, donde la distribución por tipo de estándar, CRAFT y FAIRMINED, está representada respectivamente por 52% y 48%.
- Las auditorias se componen de 10 categorías. Las categorías más representativas son Social, Ambiental y Seguridad y Salud en minería, las cuales representan más del 50% de los criterios evaluados. La categoría Genero es la menos representativa, con solo 5 criterios.
- El conjunto de datos es una muestra representativa de auditorías llevadas a cabo en 7 regiones diferentes, cada una localizada en uno de los 3 países a continuación: 5 regiones de Colombia, 1 región de Brasil, y 1 región de Ecuador.
- La mayoría de las auditorías fueron realizadas en Colombia, representando más del 87% del conjunto de datos. El porcentaje de datos representativos resultado de las auditorias en Brasil y Ecuador fue respectivamente de 8.3% y 4.2%. Así mismo, la región con mayor porcentaje representativo (33%) de los datos es la región Bolívar, en Colombia, con 8 auditorias.
- El 92% de las minas auditadas no se encuentran actualmente certificadas por ningún órgano controlador nacional ni internacional.
- El 92% de las minas auditadas son de tipo subterránea. Únicamente 2 minas son de tipo Aluvial, ambas localizadas en Brasil.
- El 92% de las minas auditadas reportan no hacer uso de mercurio en sus procesos extractivos. Únicamente 2 minas confirman el uso de mercurio, ambas localizadas en Brasil.

Estadísticas Descriptivas Univariable. Se revisaron las medidas de tendencia central y de dispersión de los campos numéricos del conjunto de datos, las cuales se resumen en las tablas 5 y 6, las cuales se resumen en las siguientes observaciones:

Tabla 5 - Descripción estadística

Descripción Estadística	trabajador	oro_mes	tonelada	calificacion parcial	PresupuestoARM_$MM	TiempoDedicacion Personal	Tiempo Cumplimiento Actividad
Conteo	6791	5629	5094	6791	6791	6791	6791
Media	48.1	4.8	569.4	1.74	0.741	4.1	132.2
Desviación Estándar	70.1	6.0	831.0	0.99	0.396	3.1	95.4
Mínimo	3	0	12	0.0	0.163	0.7	30.0
Percentil 25%	11	1	124	1.0	0.368	1.1	35.0
Percentil 50%	33	2	300	2.0	0.672	3.2	105.0
Percentil 75%	63	6	551	2.8	1.091	6.2	205.0
Máximo	363	25	3200	3.0	1.653	13.3	360.0

Tabla 6 - Descripción estadística en campos de Presupuesto

Descripción Estadística	PresupuestoARM_$MM	PresupuestoMinaCategoria_$MM	PresupuestoTotal_$MM
Conteo	6791	6791	6791
Media	0.741	5.047	246.36
Desviación Estándar	0.396	3.526	62.99
Mínimo	0.163	0.321	95.18
Percentil 25%	0.368	2.253	203.94
Percentil 50%	0.672	4.198	258.39
Percentil 75%	1.091	6.751	313.87
Máximo	1.653	17.344	330.97

- En general, la media (promedio aritmético) y el percentil-50 (mediana) presentan diferencias entre sí, indicando que pueden existir valores extremos o una distribución sesgada de los datos. Para comprender mejor esta observación, se ha construido un diagrama de cajas para los mismos campos, presentado en la figura 7, en la cual se puede percibir la presencia de valores atípicos en los campos de cantidad de trabajadores, oro por mes, tonelada de mineral, y presupuesto total por Categoría. Así mismo, se observa que en general, otras variables presentan valores extremos (mínimo y máximo).

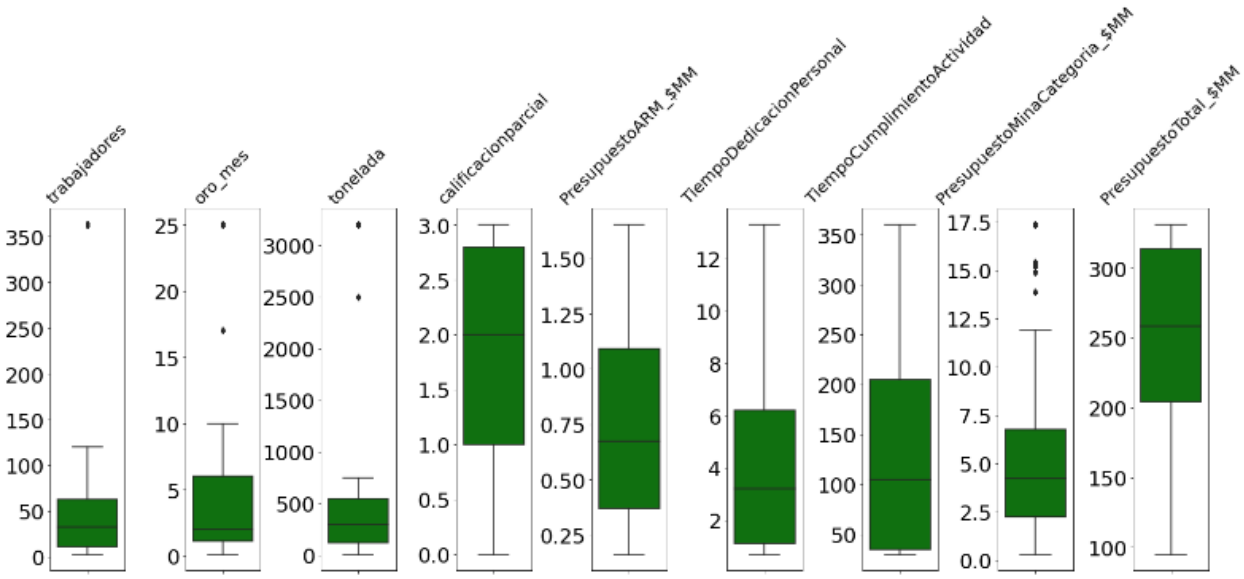


Figura 7 - Análisis Exploratorio de datos. Diagrama de cajas.

- La calificación parcial es el campo en el cual se registran las clasificaciones obtenidas de acuerdo con los criterios evaluados. Dicha calificación permite determinar preliminarmente actividades de mejora para asegurar la sostenibilidad minera. La media y mediana globales de la calificación parcial se sitúan cerca de 1.8, lo que equivale a un 60% de cumplimiento de cada criterio (1.8/3), con una desviación estándar equivalente a 1 punto, dentro de la escala de calificación de 0 a 3. En las figuras 8 y 9, se muestra a mayor detalle el comportamiento de este importante campo, lo que permite concluir de forma inicial:
    - El tipo de estándar no determina o afecta la calificación parcial.
    - A pesar de observarse diferencias significativas entre los países y por regiones (mayor o igual a la desviación estándar), debido a la cantidad de minas y registros por país y región, no es recomendable concluir la existencia de un posible patrón. Sin embargo, es posible afirmar debido a la cantidad de minas y su calificación promedio, que la región Bolívar de Colombia presenta el mejor desempeño general en las auditorías.
    - Las categorías de Trazabilidad, Ambiental, y Genero, presentan los promedios más bajos, siendo la categoría ambiental la más crítica debido a la alta cantidad de criterios evaluados en comparación con las demás categorías.

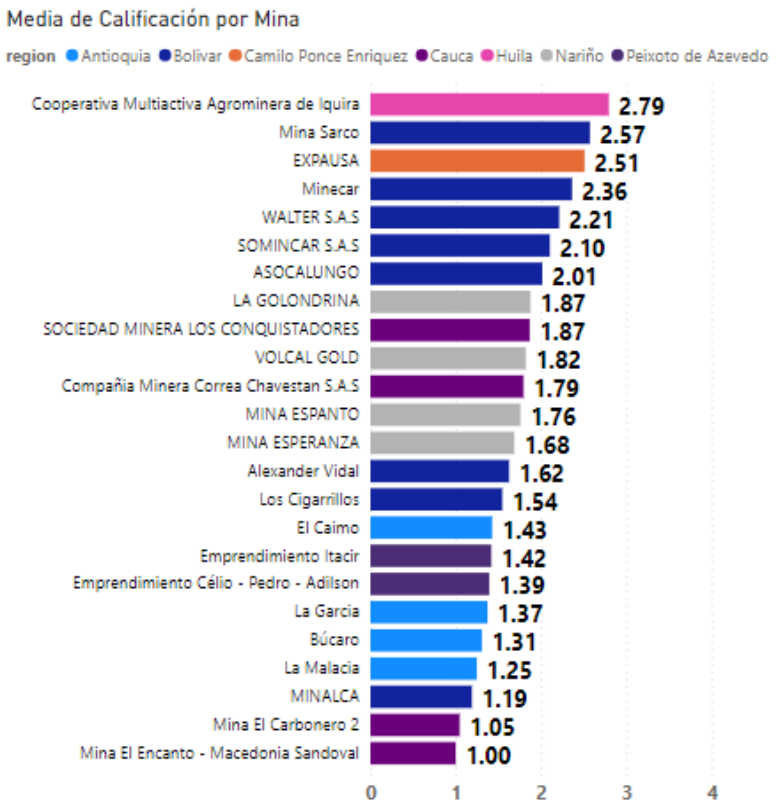


Figura 8 - Media de la calificación por Mina.

- Los campos de presupuesto total y por categoría son valores totales por mina y categoría evaluada, respectivamente. Se observa que el valor máximo de inversión requerida para el cumplimiento de los objetivos establecidos tras los resultados de una auditoria puede llegar a 331 millones de pesos colombianos. A diferencia de otras variables y a pesar de los valores extremos, este campo presenta una baja diferencia relativa entre la media y la mediana.
- El campo de presupuesto ARM, es un campo que define la ARM en función de múltiples factores y equivale a un porcentaje variable del presupuesto total en millones de pesos colombianos - $MM. Este campo presenta una baja diferencia entre su media y mediana, sin embargo, como se muestra en la figura 10, histograma de frecuencia, este campo presenta una distribución asimétrica hacia la izquierda, pero a su vez una distribución bimodal, encontrando el primer grupo en el rango de $0.2 a $0.4 MM y un segundo grupo entre $1.0 a $1.2 MM.

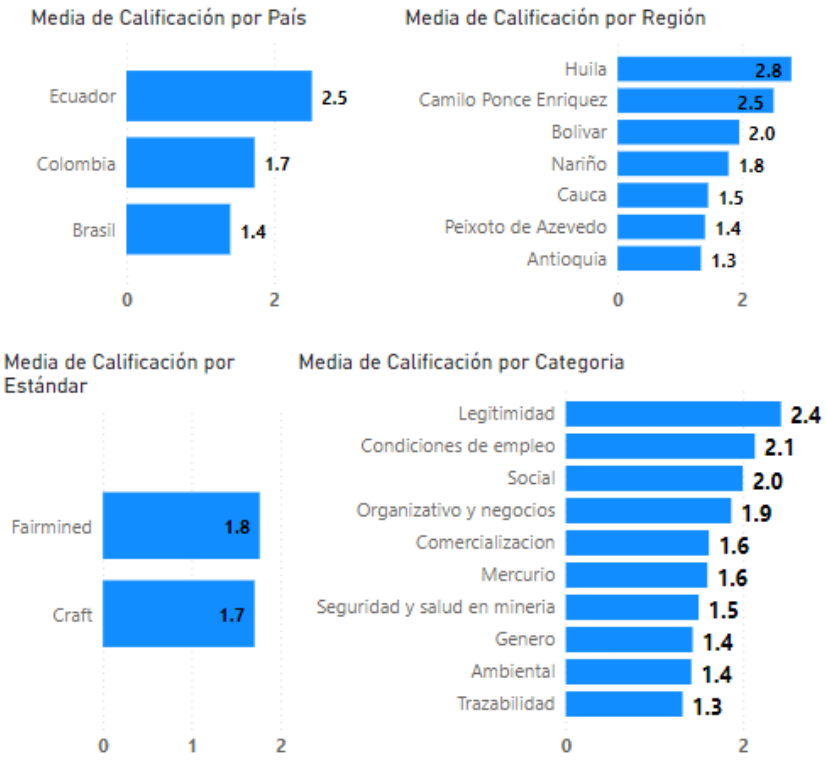


Figura 9 - Media de la calificación por dimensiones País, Región, Estándar y Categoría.

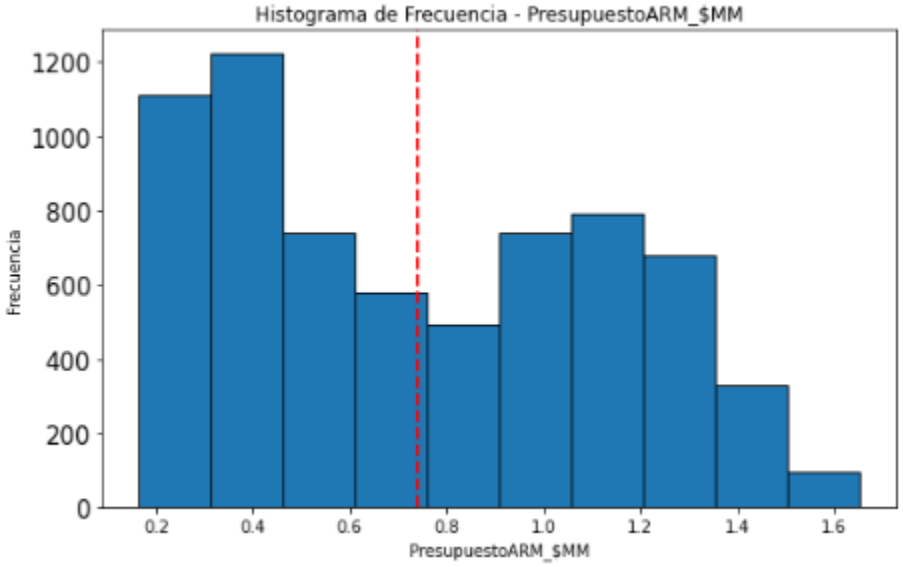


Figura 10 - Histograma de frecuencia - Presupuesto ARM.

- El tiempo estimado de cumplimiento de actividades de mejora propuesta por criterio puede llegar a alcanzar 1 año (365 días) de gestión, sin embargo, se observa que, en promedio, las actividades propuestas tienen un tiempo esperado de cumplimiento entre 4 y 5 meses (132 días). El tiempo de dedicación de personal, es definido dependiendo del tiempo estimado por actividad y depende de múltiples factores, tal como la disponibilidad de personal. El histograma de frecuencia presentado en la figura 11, muestra una clara distribución asimétrica orientada hacia la izquierda, concluyendo que la mayoría del personal de soporte brinda menos de 4 días al cumplimiento de la actividad.

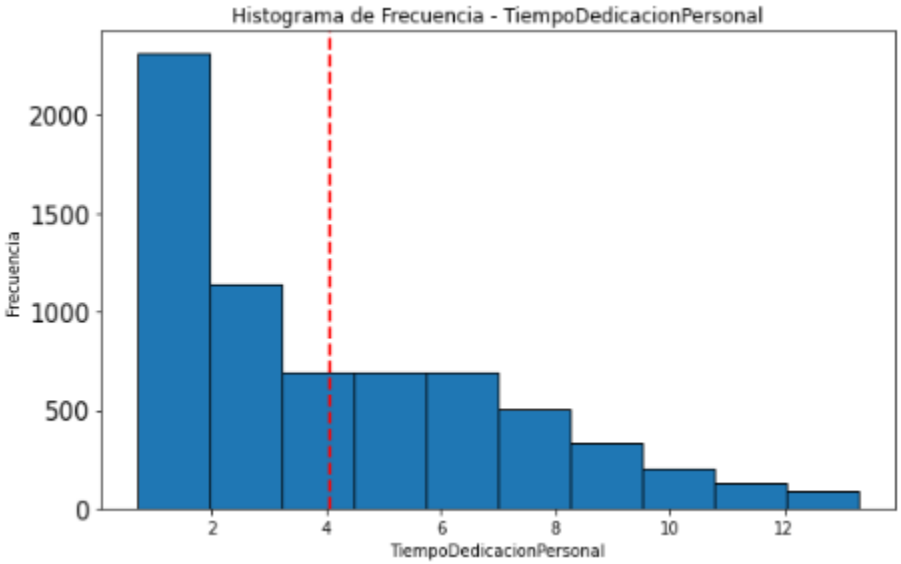


Figura 11 - Histograma de frecuencia - Tiempo Dedicación Personal.

5.3.2. Análisis de relaciones entre variables

Matriz de correlación. A partir de las observaciones de la sección anterior, para esta fase del análisis exploratorio de datos se construyó inicialmente la matriz de correlación Pearson, y mediante la ayuda de un mapa de calor se construyó su visualización, presentada en la figura 12.

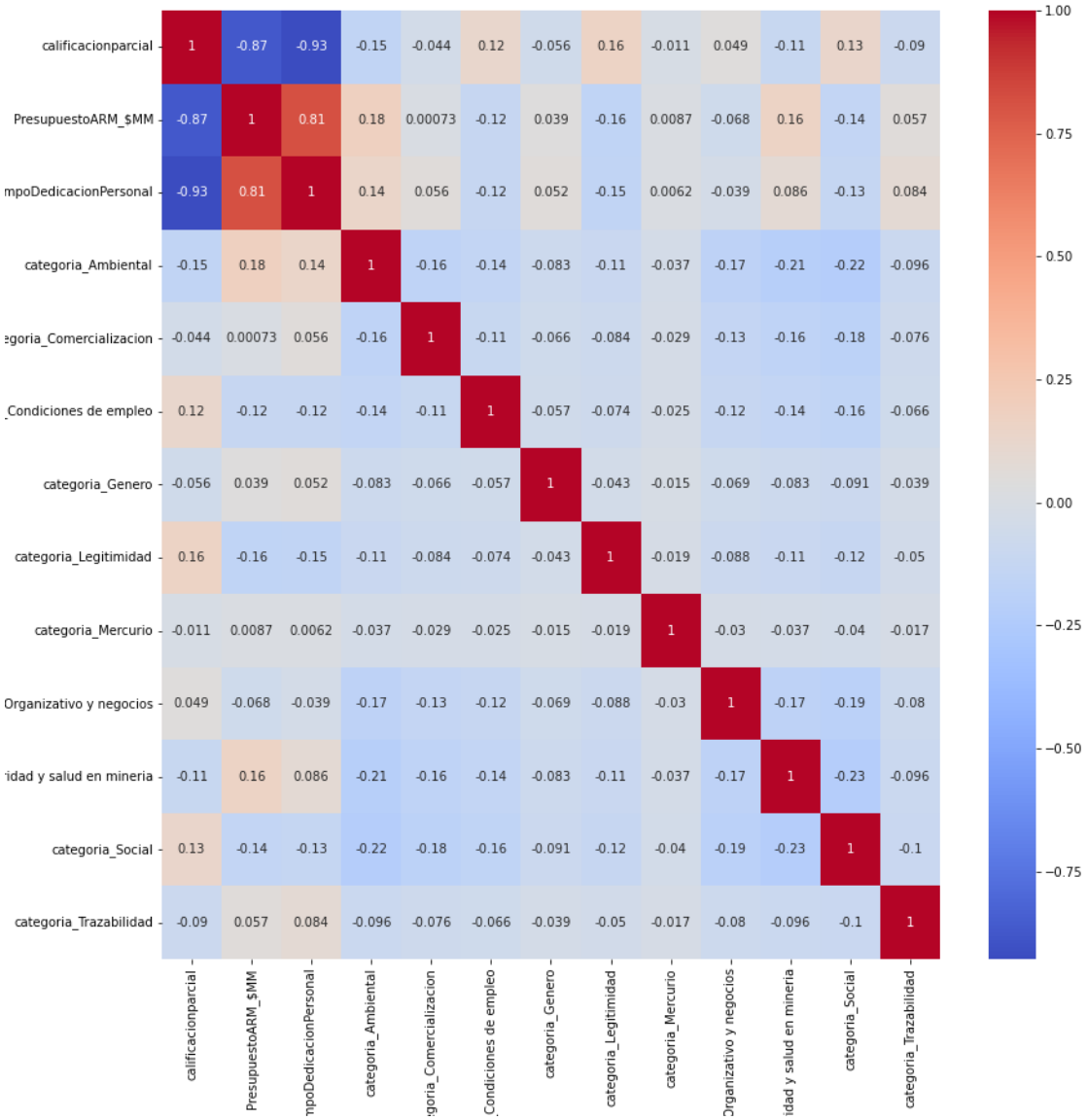


Figura 12 - Mapa de calor de la matriz de correlación.

A pesar de las potenciales relaciones o patrones identificados en la sección de análisis Univariable, la matriz de correlación ayudó a confirmar que las relaciones existentes son únicamente representativas entre las variables de calificación parcial, presupuesto ARM $MM, y tiempo de dedicación de personal, estas dos últimas presentando una fuerte relación inversa (coeficiente de correlación Pearson negativo) con la calificación parcial, obteniendo un coeficiente de correlación de -0.87 y -0.93, respectivamente. Estos dos campos a su vez son directamente correlacionables, con coeficiente de 0.81.

Dado lo anterior, y revisando los intereses de la Alianza por la Minería Responsable, es de interés evaluar la implementación de técnicas de inteligencia artificial que permitan la predicción del presupuesto de aporte de la organización al cumplimiento de los objetivos de los grupos mineros. Por lo cual, se propuso el análisis de relaciones de esta variable, así como, el análisis de colinealidad entre los campos de calificación parcial y tiempo de dedicación de personal.

Presupuesto ARM. Para entender mejor el comportamiento de esta variable, se revisó su relación con otras dimensiones dentro del conjunto de datos. Las principales observaciones se listan a continuación:

- El presupuesto de la ARM representa entre el 80-90% del presupuesto total requerido para el financiamiento del cumplimiento de objetivos de sostenibilidad minera, figura 13.

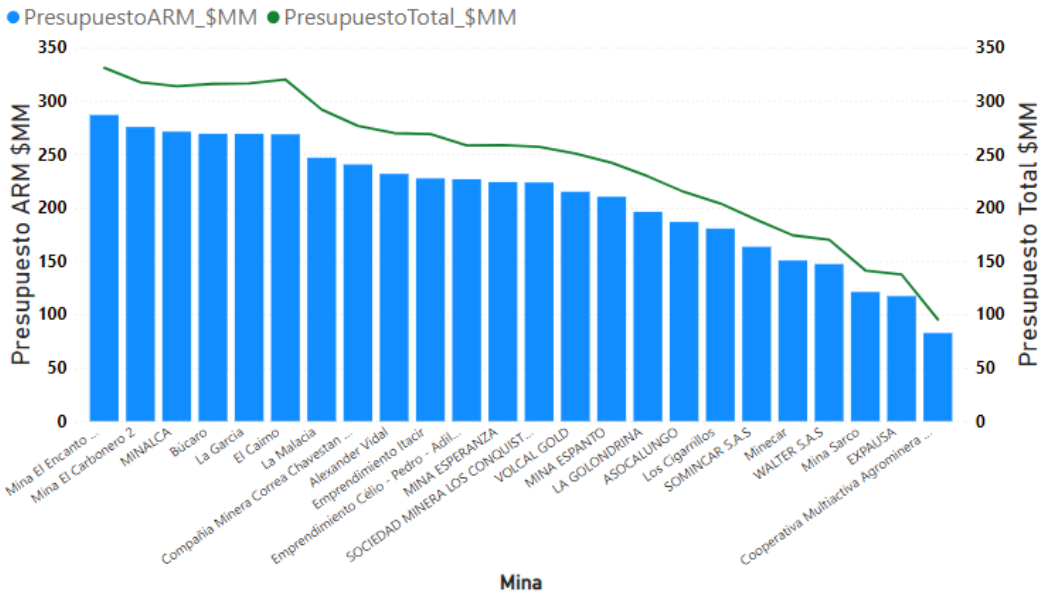


Figura 13 - Presupuesto ARM y total estimado por Mina.

- En general, como se confirmó con la matriz de correlación, no se observó una tendencia o patrón que relacione el presupuesto con cualquier distribución por estándar o categoría, sin embargo, como se muestra en la figura 14, es posible afirmar que categorías con mayor cantidad de criterios evaluados requerirán un mayor volumen de financiamiento, esto

como consecuencia de que habrá lugar para la propuesta de más actividades de mejora, por lo tanto, requerirán de mayor tiempo de dedicación del personal e inversión.

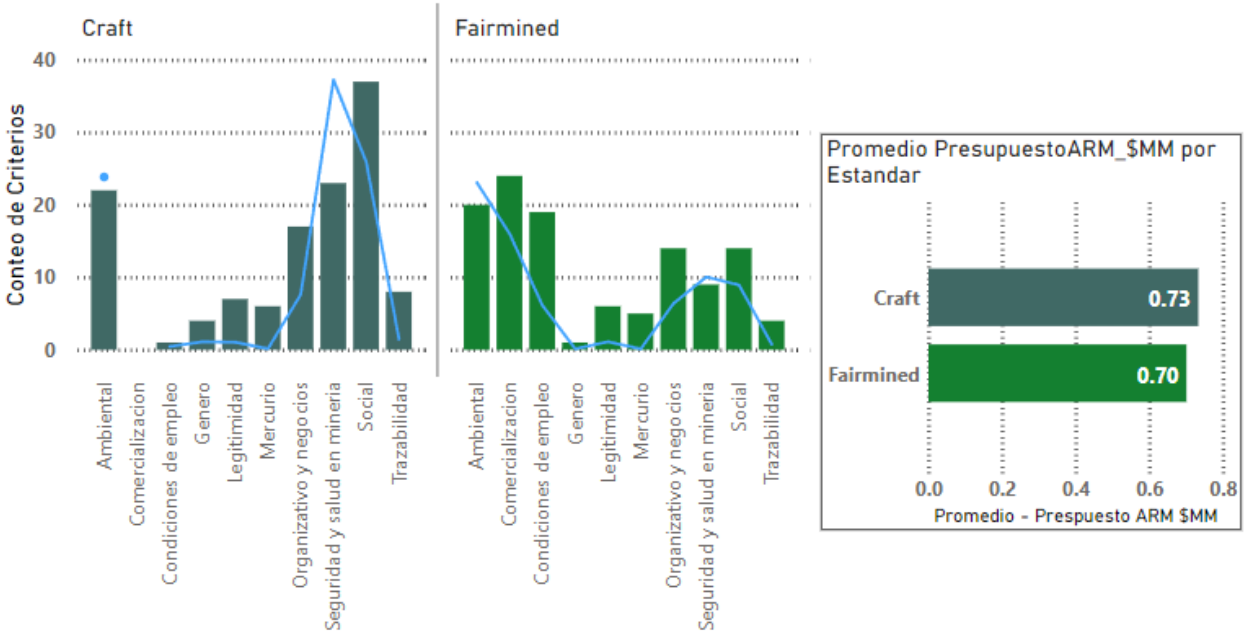


Figura 14 - Distribución de criterios evaluados por categoría y estándar, comparados contra el promedio de presupuesto total $MM por categoría (línea azul claro).

- Los diagramas de dispersión, figura 15, evidencian la relación lineal entre las variables, sin embargo, a su vez sugieren la potencial existencia de colinealidad entre las variables predictoras, calificación parcial y tiempo de dedicación de personal. Por lo anterior, se llevó a cabo un análisis del factor de inflación de varianza (VIF) y un análisis de significancia mediante valor p, los resultados obtenidos se muestran en las tablas 7 y 8, respectivamente.

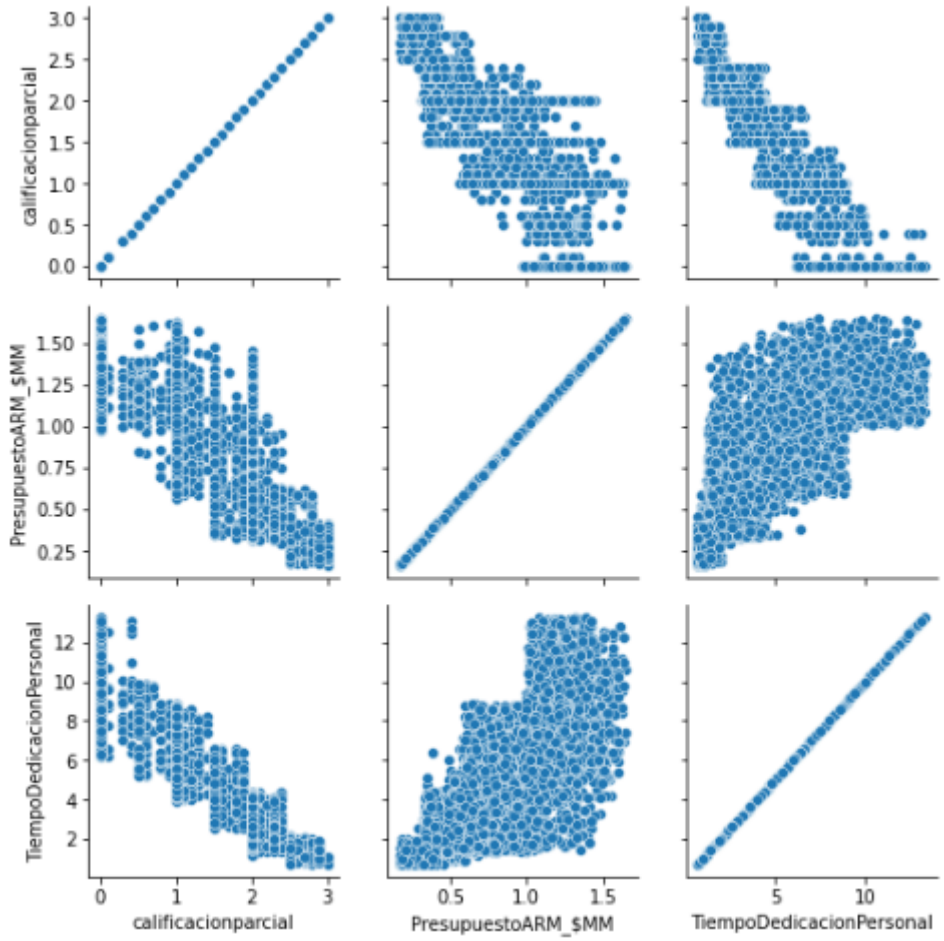


Figura 15 - Diagrama de dispersión entre variable objetivo, Presupuesto ARM, y variables predictoras, Calificación parcial y Tiempo de dedicación de personal.

Tabla 7 - Factor de inflación de varianza VIF

#	Variable	Valor VIF
1	calificacionparcial	7.258132
2	TiempoDedicacionPersonal	7.258132

Tabla 8 – Coeficientes de correlación Pearson y Análisis de significancia valor-p.

Variable	calificacionparcial	TiempoDedicacionPersonal	PresupuestoARM_$MM
calificacionparcial	1	-0.93	-0.87
TiempoDedicacionPersonal	-0.93	1	0.81
PresupuestoARM_$MM	-0.87	0.81	1
P-valor (calificacionparcial vs TiempoDedicacionPersonal): <0.05 P-valor (calificacionparcial vs PresupuestoARM_$MM): <0.05 P-valor (TiempoDedicacionPersonal vs PresupuestoARM_$MM): <0.05			

Colinealidad. La colinealidad en un modelo de predicción puede llevar a varios problemas como coeficientes estimados inestables, sesgo en los coeficientes y amplios intervalos de confianza. Con el fin de evaluar la presencia de colinealidad entre las variables predictoras "calificación parcial" y "tiempo de dedicación de personal", se llevó a cabo un análisis utilizando el factor de inflación de la varianza (VIF) y pruebas de significancia estadística mediante los valores p.

En el análisis VIF, en caso de que ningún factor sea correlacionable, el valor VIF será igual a 1 y, caso opuesto, valores superiores a 5 sugieren un impacto negativo al modelo de predicción [19]. Con base en los resultados obtenidos, tabla 7, se recomienda eliminar una de las variables.

Para la selección de la variable predictora a mantener se revisó el valor de significancia, considerando que valores-p por debajo del umbral 0.05 sugieren una relación o correlación significativa y, por lo tanto, pueden ser relevantes en el modelo de predicción. Los resultados obtenidos en la tabla 8 no proporcionan suficiente evidencia para determinar la eliminación de una de las variables debido a su alta significancia. En la literatura, se han propuesto otros métodos para abordar situaciones similares de alta significancia y multicolinealidad, como el análisis de relevancia teórica, el análisis de sensibilidad del modelo de predicción y otras técnicas más complejas de selección de variables [19]

En este caso particular, considerando el análisis exploratorio previo de los datos y la información proporcionada por el histograma de frecuencia en la figura 16, donde se observa que las calificaciones parciales se concentran principalmente en el rango de 0 a 3 con valores incrementales de 0.5 (0, 0.5, 1, 1.5, 2, 2.5 y 3), se opta por realizar un análisis de relevancia teórico, concluyendo que la variable predictora más representativa es el tiempo de dedicación de personal. No obstante, se llevaron a cabo pruebas de sensibilidad para confirmar la relevancia de esta variable frente a la segunda variable predictora (calificación parcial).

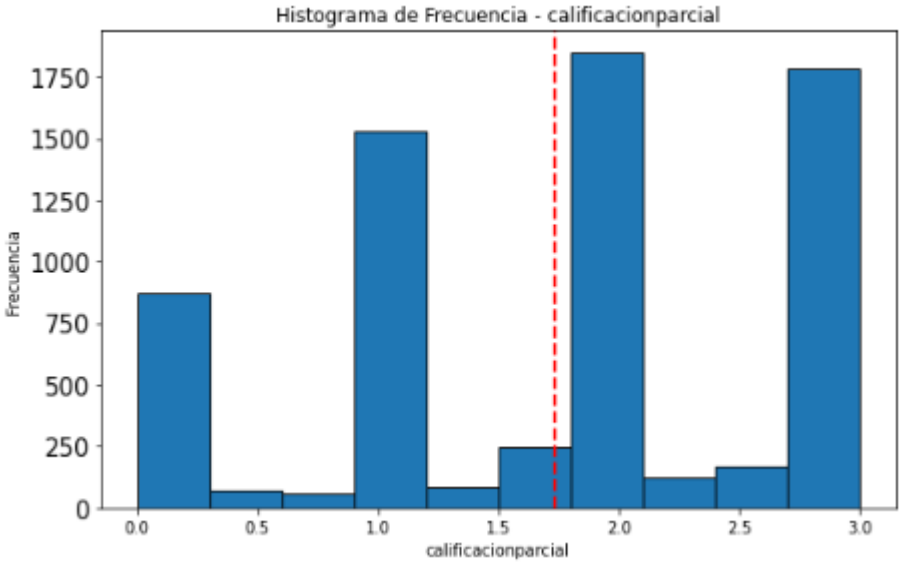


Figura 16 - Histograma de frecuencia - Calificación Parcial.

5.4. APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL

Como se explica en la sección anterior, se identificó una relación entre las variables de tiempo de dedicación de personal y presupuesto ARM, esta última de particular interés para la Alianza por la Minería Responsable, puesto que, una predicción precisa de este campo permitiría simplificar tiempos de análisis, soportaría la orientación de planes acción y la toma de decisiones sobre destino de recursos en el corto a mediano plazo.

5.4.1. Descripción de conjunto de datos

A partir del análisis exploratorio, se simplifica el conjunto de datos manteniendo únicamente las variables necesarias para la división en conjuntos de entrenamiento y prueba de los algoritmos aplicados. En la tabla 9, se presenta en resumen las características del conjunto de datos que se empleó en esta fase del proyecto.

Tabla 9 – Descripción del conjunto de datos para implementación de técnicas de IA

Campo	Registros No Nulo	Registros Nulo	Tipo de dato
idcriterio	6791	0	Entero - int64
calificacionparcial	6791	0	Decimal - float64
PresupuestoARM_$MM	6791	0	Decimal - float64
TiempoDedicacionPersonal	6791	0	Decimal - float64

5.4.2. Revisión de técnicas implementadas y resultados

El análisis exploratorio de los datos permitió identificar una fuerte relación lineal entre las variables predictora (Tiempo de Dedicación de Personal) y objetivo (Presupuesto ARM $MM), por lo cual se implementaron en un primer acercamiento las siguientes técnicas de inteligencia artificial dada su amplia flexibilidad con modelos lineales [20]:

- Modelo de regresión lineal.
- Árbol de decisión.
- Máquina de Soporte Vectorial, regresor. SVR.
- Red Neuronal Artificial.
- Adicionalmente, se implementó el modelo de regresión lineal con regularización Ridge, sugerido en escenarios con multicolinealidad. Esto para evaluar un escenario incluyendo la variable predictora de calificación parcial.

Todos los algoritmos de los modelos mencionados fueron analizados con un histograma de frecuencia de valores reales contra predichos, soportado además por un gráfico QQ (Cuantil-Cuantil) y con un gráfico de residuos. Adicionalmente, se compararon los modelos utilizando 3 medidas de rendimiento (tabla 10): coeficiente de determinación (R^2), error cuadrático medio (MSE), y error absoluto medio (MAE).

Tabla 10 - Medidas de rendimiento.

Coeficiente de determinación (R2)	
Regresión lineal	0.6633
Ridge *(incluyendo variable "calificacionparcial")*	0.7520
Árbol de decisión	0.7267
Red neuronal	0.7196
SVR	0.7216
Error cuadrático medio (MSE)	
Regresión lineal	0.0538
Ridge *(incluyendo variable "calificacionparcial")*	0.0396
Árbol de decisión	0.0437
Red neuronal	0.0448
SVR	0.0445
Error absoluto medio (MAE)	
Regresión lineal	0.1832
Ridge *(incluyendo variable "calificacionparcial")*	0.1534
Árbol de decisión	0.1599
Red neuronal	0.1619
SVR	0.1593

Concluido el primer acercamiento con los modelos anteriores, se obtuvieron las siguientes observaciones:

- Con base en los resultados obtenidos de la comparación entre las medidas de rendimiento, tabla 10, se observa que en general, el modelo de regresión lineal con regularización de Ridge presenta el mejor desempeño, seguido por los modelos de árbol de decisión y máquina de soporte vectorial (SVR). Sin embargo, como se explica en el siguiente punto, la inclusión de la variable con colinealidad, calificación parcial, altera y sesga los resultados obtenidos por medio de los modelos, por lo que, se considera que los modelos de árbol de decisión y SVR evaluados únicamente con la variable predictora de tiempo de dedicación de personal, presentan el mejor rendimiento al ajustarse mejor a la variabilidad de los datos (mayor coeficiente de determinación) y al presentar los errores (MSE, MAE) más bajos en el ajuste con los valores reales.
- A pesar de observarse mejores resultados en las medidas de rendimiento cuando se incluye la segunda variable predictora, calificación parcial, como en el ejemplo de la

regresión Ridge, la adición de esta variable genera sesgos que pueden ser observados a través del histograma de frecuencias entre valores reales y predichos, y en el diagrama de dispersión de la misma comparación. En la figura 17, se presentan los gráficos correspondientes del modelo de regresión Ridge, en los cuales puede observarse la agrupación de los resultados, generando sesgos en la predicción.

- Concluido hasta esta instancia que los modelos de árbol de decisión y SVR presentaron el mejor rendimiento y ajuste con los valores reales, se analizaron los respectivos gráficos de dispersión e histograma de valores reales contra predichos, grafico QQ y de residuos, los cuales se presentan en las figuras 18 y 19, para cada modelo respectivamente. A partir de la interpretación de las gráficas, y a pesar de que ambos modelos presentaron los mejores resultados en las medidas de rendimiento, se pudo concluir que:
    - El modelo de árbol de decisión puede presentar sesgos significativos al no poder representar el espectro completo del rango de distribución de la variable objetivo (presupuesto ARM), efecto que no ocurre con el modelo SVR, el cual presenta mínimo sesgo en la distribución de resultados, pero el cual, a su vez, presenta dificultades para predecir valores extremos (mínimo y máximo).
    - Complementando la observación anterior, y como se confirma con el gráfico QQ, ambos modelos presentan un problema de subestimación y sobreestimación sistemática de los valores predichos hacia los extremos en comparación con los valores reales. Sin embargo, lo anterior no es un fenómeno resultante de los parámetros de ajuste de los modelos, sino potencialmente un producto de la variabilidad no constante de los errores o residuos de los modelos a lo largo del rango de valores predichos, lo cual se confirma con el grafico de residuos en cada modelo. Este fenómeno es conocido como Heterocedasticidad [21], lo cual implica que la dispersión de los residuos cambia a medida que los valores predichos aumentan o disminuyen. Dado lo anterior, se concluye que ninguno de los modelos implementados en el primer acercamiento es recomendado para la predicción del prepuesto ARM debido a la naturaleza propia del conjunto de datos; en su lugar, se recomienda el uso de modelos que tengan en cuenta soluciones para el fenómeno de la heterocedasticidad.

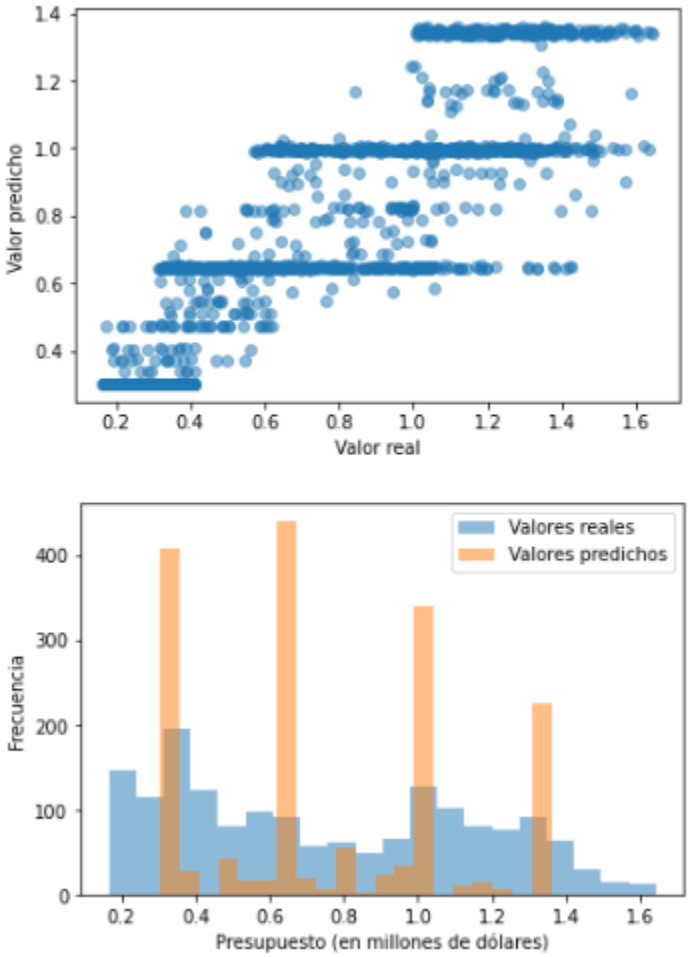


Figura 17 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos mediante el modelo de regresión lineal con regularización Ridge.

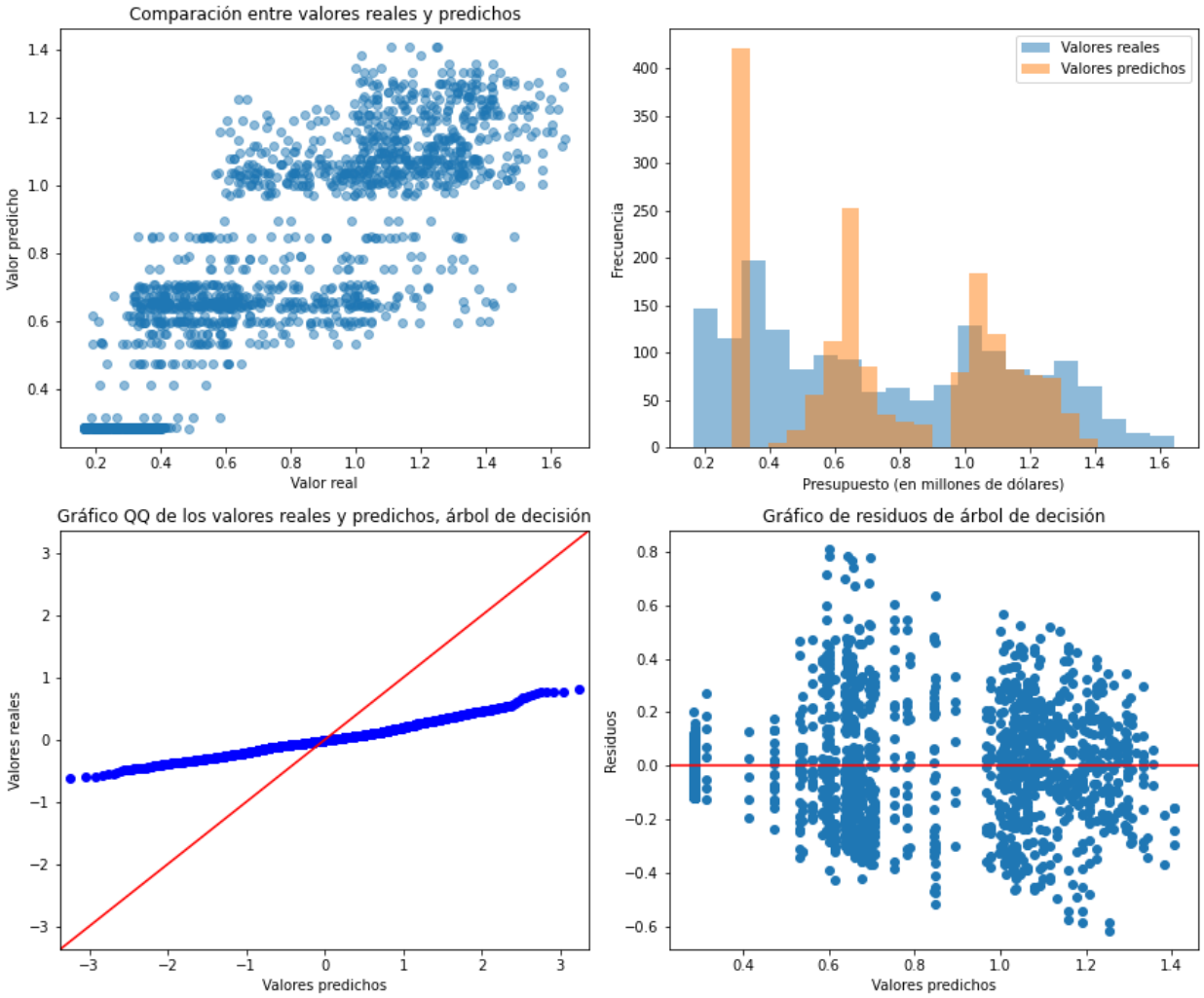


Figura 18 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de árbol de decisión.

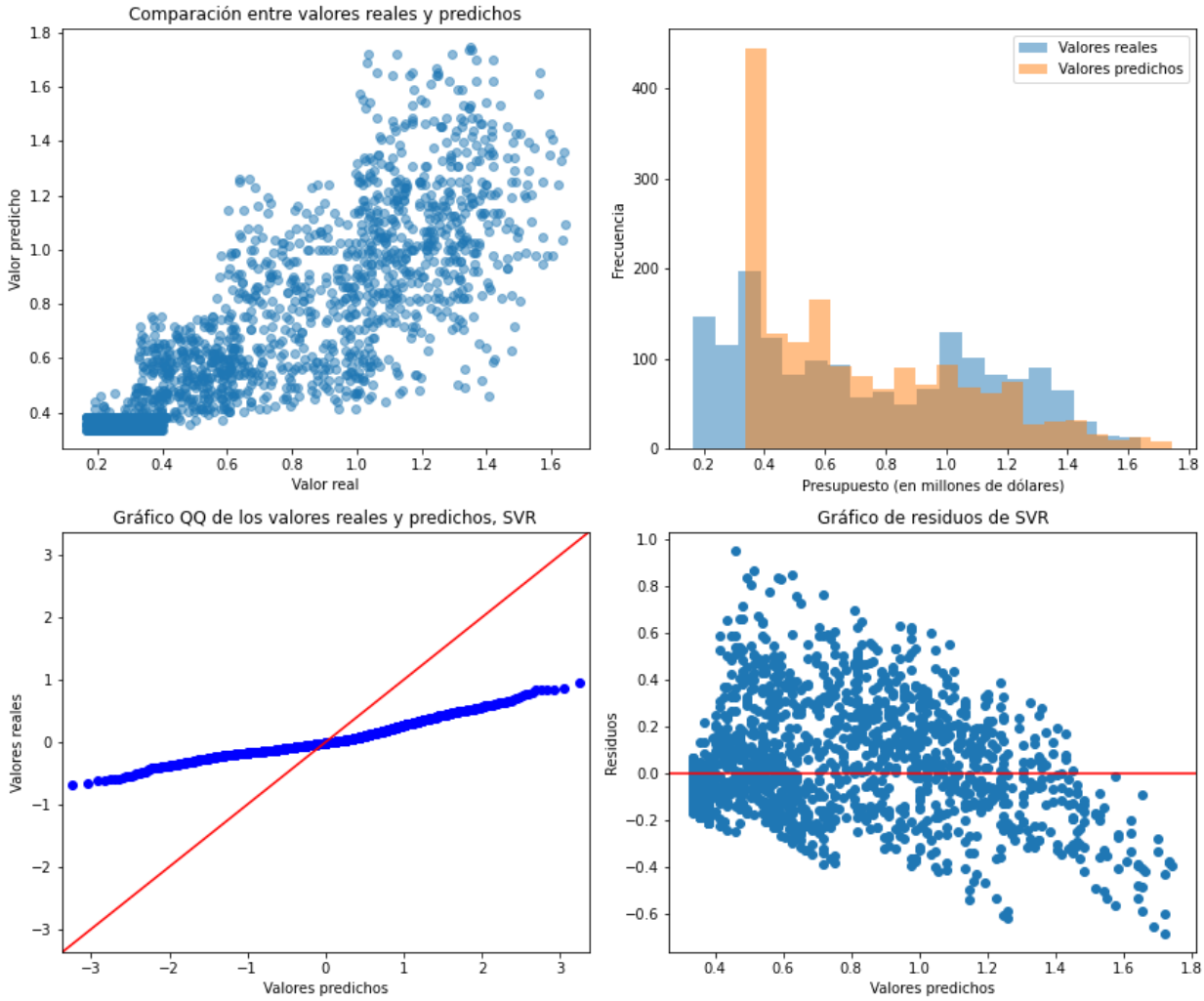


Figura 19 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de máquina de soporte vectorial - SVR.

5.4.3. Resultados del modelo de regresión ponderada

Se ha abordado el problema de la heterocedasticidad utilizando la regresión ponderada [21]. La heterocedasticidad puede afectar la precisión y confiabilidad de los resultados de la regresión, pero mediante la asignación de pesos a las observaciones en función de su varianza, se ha logrado reducir el impacto de este fenómeno. Esto se evidencia en el gráfico QQ (Figura 20), donde se observa una mejor alineación de los valores predichos con la línea diagonal, lo que indica una mejor satisfacción de los supuestos de normalidad y homocedasticidad.

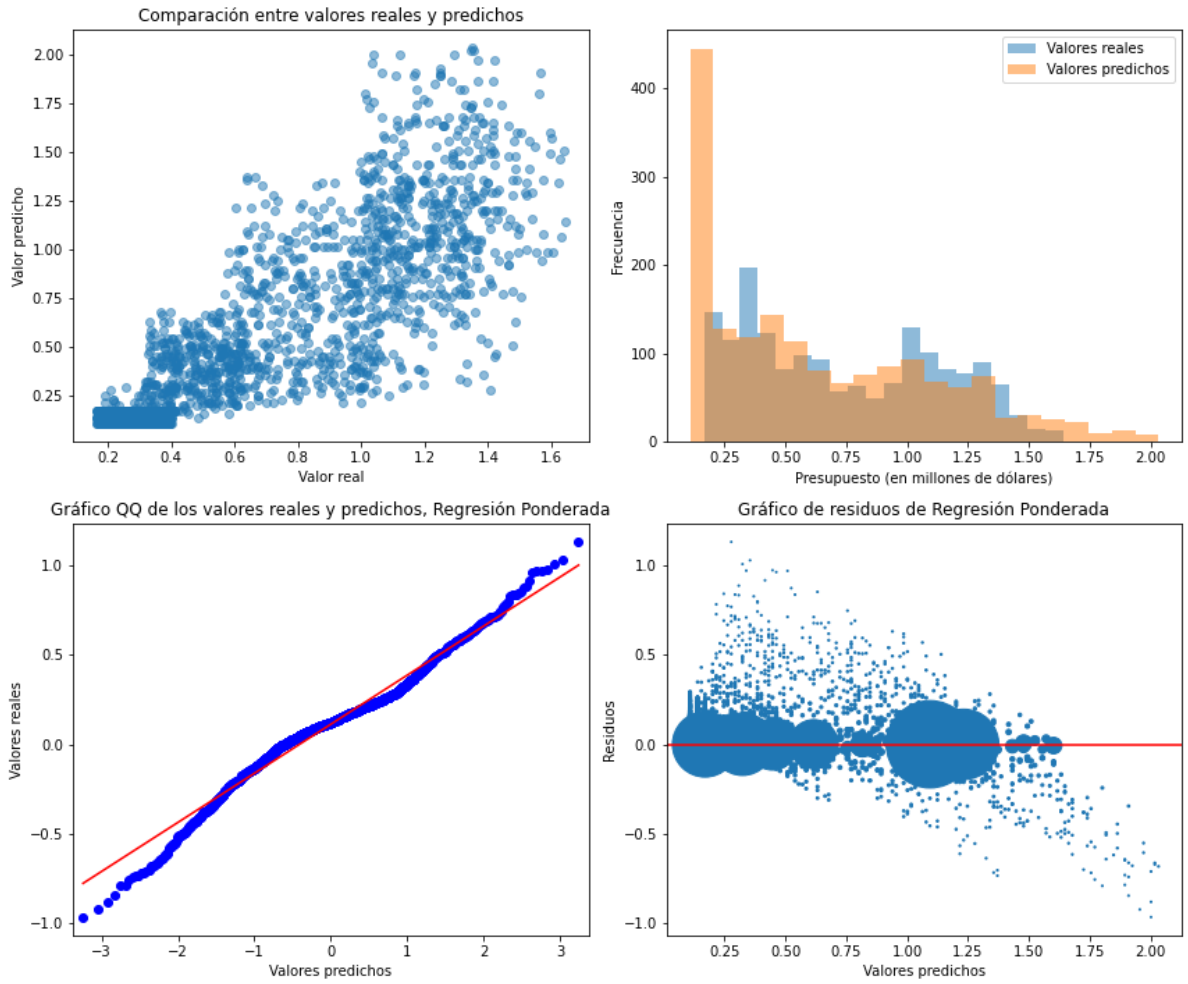


Figura 20 - Graficas de dispersión e histograma de frecuencia entre valores reales y predichos, y graficas QQ y de residuos, modelo de regresión ponderada.

Basado en los resultados del modelo de regresión ponderada, presentados en la tabla 11, se puede concluir que la variable predictora "TiempoDedicacionPersonal" tiene un efecto significativo en el presupuesto de ARM (PresupuestoARM_$MM). Las observaciones se detallan a continuación:

- El modelo de regresión ponderada (WLS) utilizado en este análisis muestra un coeficiente de determinación R cuadrado de 0.861, lo que indica que aproximadamente el 86.1% de la variabilidad en el presupuesto se explica por la variable predictor.

- La estadística F, con un valor de 3.167e+04, muestra la significancia global del modelo, y el valor de probabilidad asociado es prácticamente cero, lo que indica que el modelo tiene una capacidad predictiva altamente significativa.
- El valor t asociado a este coeficiente es de 177.963, que junto con el valor p, confirma la significancia estadística de la relación entre la variable predictor y la variable dependiente
- El coeficiente para la variable predictor "TiempoDedicacionPersonal" es de 0.1540, con un error estándar de 0.001. El coeficiente de 0.1540 indica que, en promedio, por cada unidad adicional de tiempo de dedicación requerido por el personal (días), el presupuesto aumenta en 0.154 millones de pesos colombianos.
- El modelo también muestra un valor de probabilidad prácticamente cero en la prueba de normalidad de Jarque-Bera y un valor de probabilidad cero en la prueba de Omnibus, lo que sugiere que los residuos del modelo se distribuyen de manera normal. El valor de Durbin-Watson es de 1.679, lo que indica que no hay autocorrelación serial en los residuos [22].

Tabla 11 - Resultados modelo de regresión ponderada.

Resultados del modelo de regresión ponderada	
Variable dependiente	PresupuestoARM_$MM
Modelo	WLS
Coeficiente de determinación	0.861
Estadistico-F	3.17E+04
Probabilidad (Estadistico-F)	0.00
Variable Independiente	TiempoDedicaciónPersonal
Coeficiente	0.154
Error Estándar	0.001
t	177.963
P>|t|	0.00
Omnibus	254.76
Probabilidad (Omnibus)	0
Jarque-Bera (JB)	488.783
Probabilidad (JB)	7.28E-107
Durbin-Watson	1.679

En resumen, el modelo de regresión ponderada utilizado muestra un buen ajuste a los datos, con un coeficiente de determinación alto y significativo. La variable "TiempoDedicacionPersonal" tiene una influencia positiva y significativa en la predicción del presupuesto. Estos resultados respaldan la utilización de este modelo para predecir el presupuesto en función del tiempo de dedicación de personal, y proporcionan una base sólida para la toma de decisiones y análisis relacionados con el presupuesto en el contexto de estudio.

Parte del código en Python utilizado para la construcción y entrenamiento del modelo se presentan en la figura 21, así mismo en el anexo 6 se presenta el código completo junto a los otros algoritmos probados en este trabajo.

```
# Dividir los datos en variables predictoras y objetivo
X = datos[["TiempoDedicacionPersonal"]]  # Variables predictoras
y = datos["PresupuestoARM_$MM"]  # Variable objetivo

# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Calcular los pesos para las observaciones
var = np.var(y_train)
weights = np.divide(np.ones(len(y_train)), np.multiply(y_train - np.mean(y_train), y_train - np.mean(y_train)) + var)

# Crear un objeto de regresión ponderado y ajustar el modelo con el conjunto de entrenamiento
modelo = sm.WLS(y_train, X_train, weights=weights)
resultados = modelo.fit()
```

Figura 21 - Ejemplo algoritmo de modelo de regresión ponderada.

5.5. APLICATIVO WEB PARA LA GESTIÓN DE DATOS

5.5.1. Equipo Scrum

Para el desarrollo del aplicativo web, se aplicó la metodología ágil Scrum. La cual, mediante una serie de historias de usuario (anexo 4), permitió un correcto desarrollo del sistema de información mediante el uso de los sprint para cada historia. Teniendo en cuenta lo anterior, se creó el equipo scrum en donde a cada miembro se le asigna un rol para el desarrollo del producto, tabla 12.

Tabla 12 - Roles dentro del desarrollo del producto.

Rol	Asignación	Tareas
Scrum Máster	Bryan Fernandez	Gestionar el proyecto. Eliminar los impedimentos. Organizar las reuniones y planificación de sprints. Realizar los análisis retrospectivos.
Gestor del producto	Andrés Ortega	Definir los objetivos del producto. Determinar las características del producto. Crear las historias de usuario. Priorizar la pila del producto.
Desarrollador	Bryan Fernandez	Entender el negocio y los objetivos que se persiguen. Realizar las practicas agiles en el desarrollo de software. Responsable del mantenimiento del software.
Verificador	Jhon Ramírez	Identificar las necesidades del negocio. Llevar a cabo pruebas de funcionalidades y rendimiento. Escribir fallas presentadas en el software.

5.5.2. Componentes BACKEND

El componente Backend está formado por un api desarrollado con el lenguaje de programación Pyhton y el framework Flask con un patrón de arquitectura similar al Modelo Vista Controlador (MVC); teniendo como referente, que ésta estructura divide en tres capas el software, lo cual permite generar una aplicación que tiene como ventajas reutilizar el código, facilidad de detectar errores, es escalable y tiene buen grado de seguridad [23]. El api es la aplicación del lado del servidor, es decir, el software que se conecta directamente con la base de datos permitiendo realizar funciones como insertar, consultar, actualizar y eliminar registros de las tablas mediante peticiones hacia el servidor. Cuando el servidor responde, envía los resultados en un formato JSON siendo éste una forma de representar los datos entre servicios que se ejecutan en diferentes plataformas y pueden ser consumidos por diferentes lenguajes de programación [24].

Dentro de la funcionalidad del api en el aplicativo de la organización minera, se tienen los servicios de los siguientes módulos

Usuario: Rutas para la gestión de creación, consulta y edición de usuarios. También se vincula rutas relacionadas a los perfiles que tienen asignados los usuarios.
Mina: Rutas para la gestión de creación, consultas y edición de las minas. También se vincula las rutas para la consulta de países y regiones.

Criterio a mina: Rutas para la gestión de consultas y asignación de los criterios a las minas. También vincula las rutas para las consultas de los criterios creados.
Actividad: Rutas para la gestión de creación, consulta y edición de actividades las cuales están asignadas a minas que tienen registrados criterios. También se vincula las rutas para la asignación de los seguimientos de las actividades creadas a la mina.
Predicción: Rutas para la gestión de la predicción mediante el algoritmo de machine learning.

Para realizar pruebas del Backend se utilizó la herramienta Insomnia, que es una aplicación de escritorio de código abierto que simplifica la interacción con el diseño, la depuración y la prueba del api [25]. Para lo anterior, se toma como ejemplo realizar el llamado a una ruta del api para consultar minas registradas en la base de datos. En la figura 22, se muestra el resultado de aplicar el llamado a la ruta "consultar_minas", generando como respuesta, una consulta SQL en la que los datos están presentados en formato JSON para ser consumidos por otra aplicación, en este caso, la aplicación del frontend.

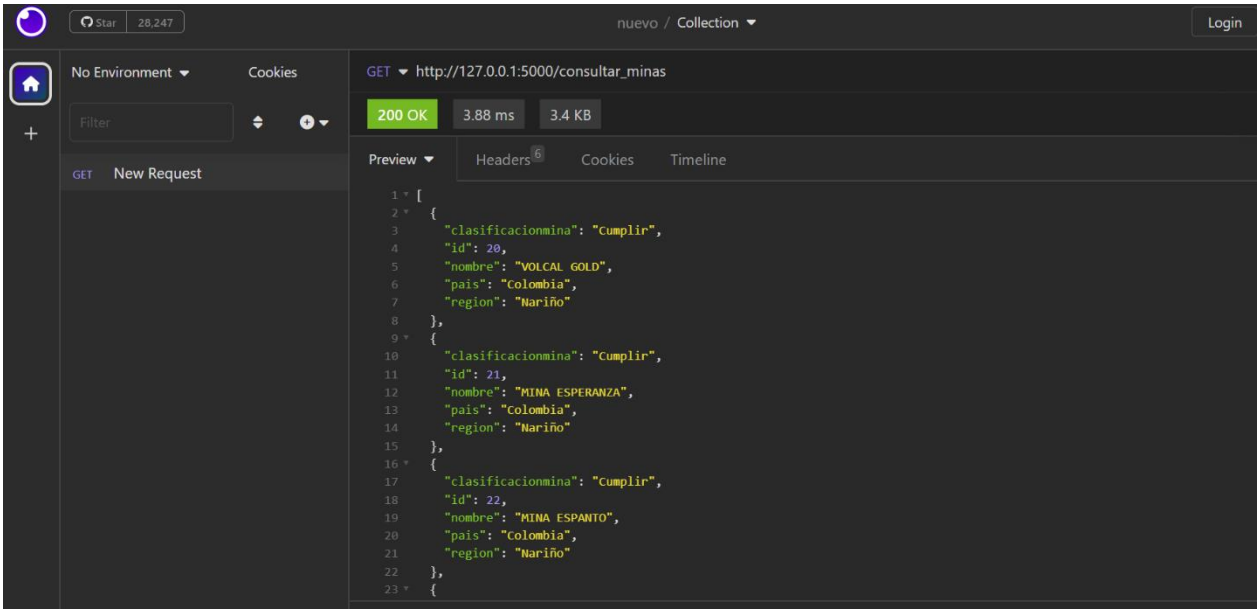


Figura 22 – Prueba Backend en Insomnia.

5.5.3. Componentes FRONTEND

Para el desarrollo FRONTEND del proyecto, se optó por usar HTML como lenguaje de maquetación ya que permite crear páginas web proporcionando semántica y estructura al contenido siendo HTML5 la versión más reciente por sus potentes funcionalidades [26]. Por otro lado, para el estilo del aplicativo web se tomó como herramienta el framework Bootstrap ya que es un potente

conjunto de herramientas de frontend con beneficios entre las que se destacan aplicar estilos de manera rápida, profesional y adaptables a diferentes ordenadores como computadores, tabletas y equipos celulares [27]. Finalmente, como lenguaje de programación del lado del cliente, se implementó Javascript, esto debido a que, es uno de los lenguajes de programación con que se desarrollan las aplicaciones web modernas, gracias a su versatilidad, eficiencia y documentación disponible, permitiendo realizar acciones sin recargar la página y proporcionado diversas formas de interactividad e ingenio[28].

Dentro de la funcionalidad del aplicativo del lado de Frontend se tienen los siguientes módulos

Módulo Login. El módulo Login o inicio de sesión, mostrado en la figura 23, está formado por un formulario en el cual solicita los datos o credenciales de acceso (usuario y contraseña). Una vez ingresado con el usuario y contraseña de manera correcta, el sistema le indica un mensaje de bienvenida y lo direcciona a el sistema de información, específicamente al módulo tablero.

Figura 23 - Modulo Login.

Módulo tablero. El módulo tablero, mostrado en la figura 24, está conformado por un grupo de gráficas que permiten conocer indicadores necesarios en el proceso de la organización minera como la calificación parcial por categoría, calificación parcial por modulo, numero de criterios cumplidos e incumplidos por categoría y numero de criterios cumplidos e incumplidos por modulo.

Módulo usuario. El módulo usuario, mostrado en la figura 25, permite realizar la gestión del personal que usará el aplicativo web. El módulo permite crear un usuario, editar un usuario, dar de baja a un usuario registrado y mostrar todos los usuarios creados en el sistema.

Módulo mina. El módulo mina, mostrado en la figura 26, permite realizar la gestión de registro de una nueva mina al sistema. Adicionalmente, brinda una pestaña con estadísticas básicas de las minas registradas, otra pestaña con el listado de las minas creadas con la opción de editar y finalmente, una pestaña para la creación de una nueva mina.

Módulo criterio mina. El módulo criterio mina, mostrado en la figura 27, permite realizar la asignación de los diferentes criterios de evaluación a una mina. Para lo anterior, la mina debe estar registrada en el sistema ya que al momento se asignar un criterio debemos seleccionar la mina registrada. Adicionalmente el módulo cuenta con una pestaña para consultar todas las minas con criterios asignados y una pestaña con información de las minas con su cantidad de criterios asignados.

Figura 24 - Módulo del tablero de control.

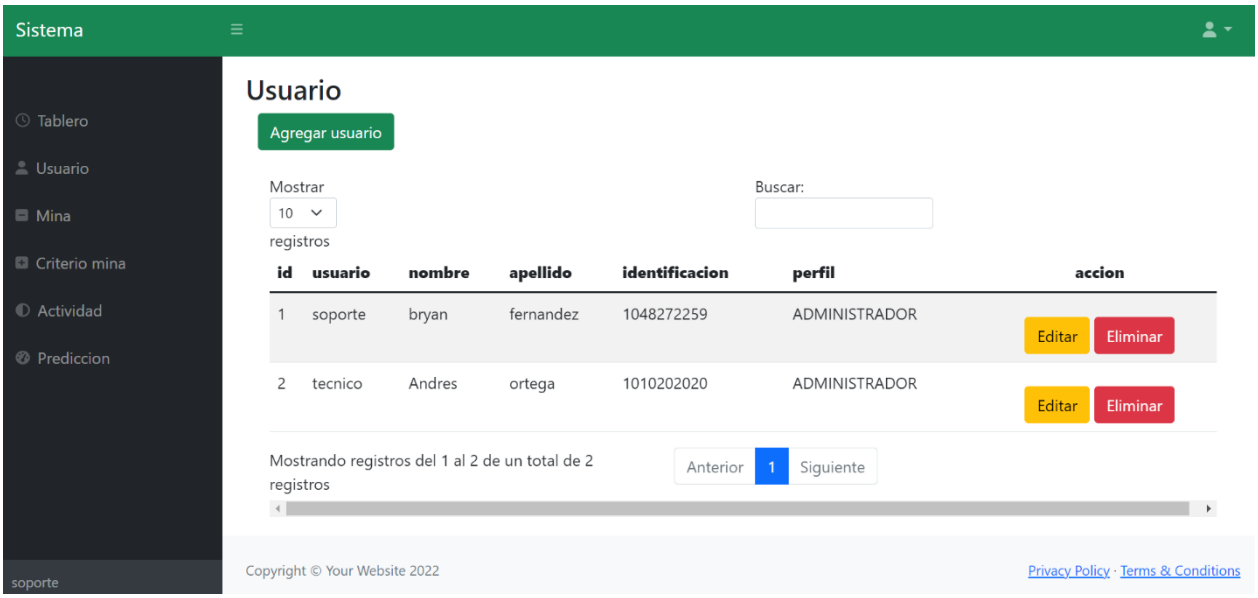


Figura 25 - Módulo de usuario.

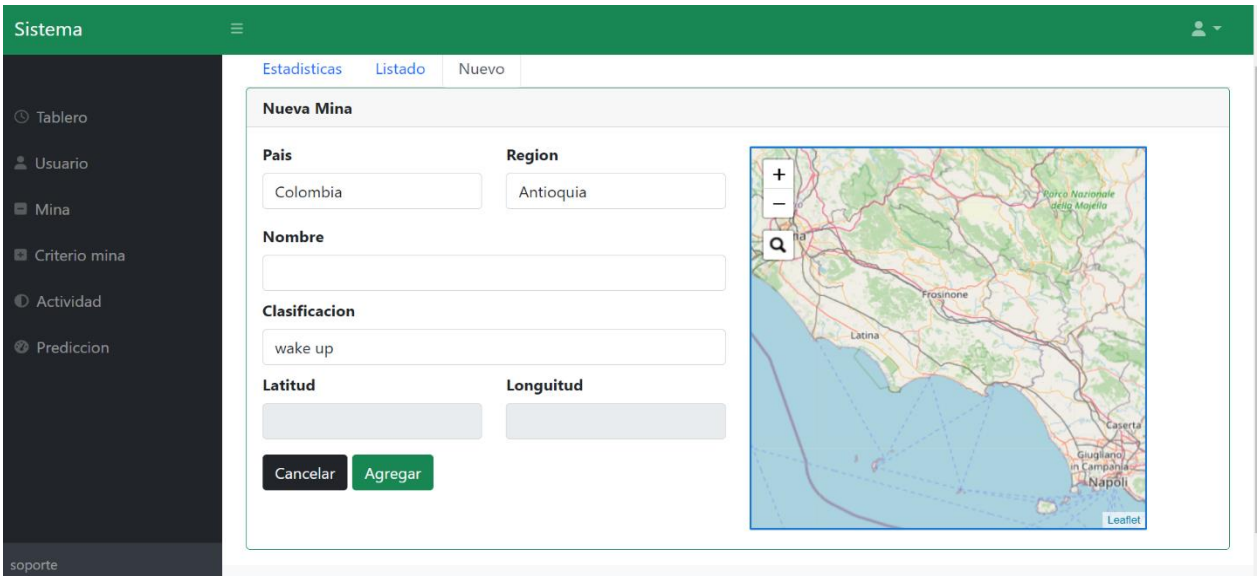


Figura 26 - Módulo de Mina.

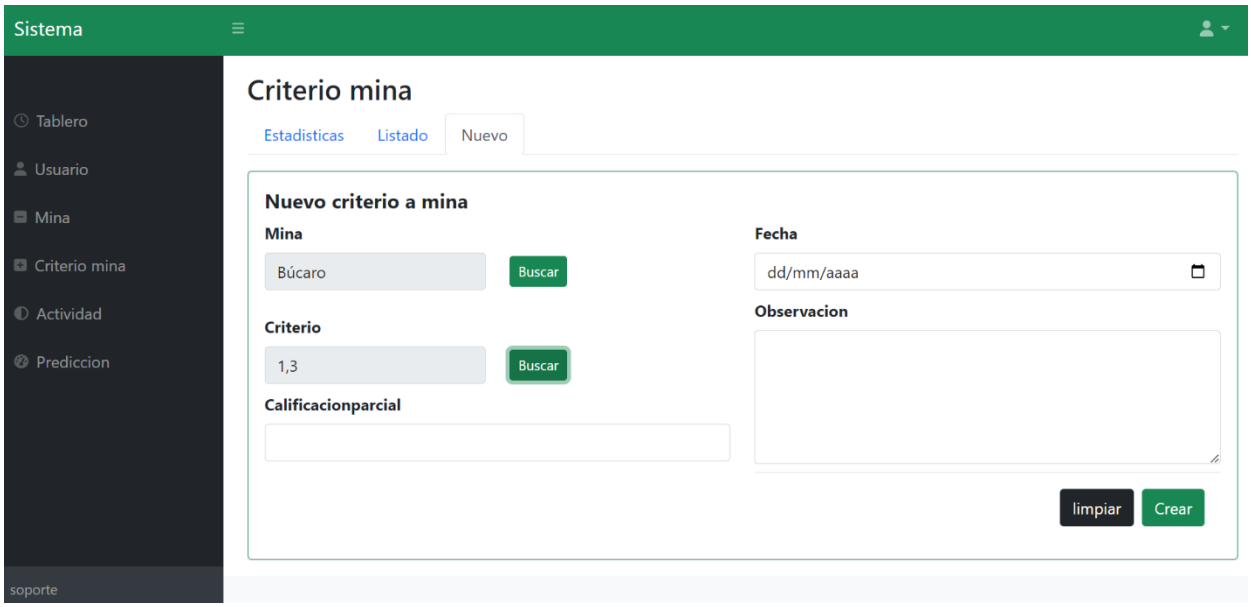


Figura 27 - Módulo criterio mina.

Módulo actividad. El módulo actividad, mostrado en la figura 28, permite realizar la creación de actividades a minas que previamente tienen asignados criterios. Por lo anterior, es necesario indicar la mina y a que criterio se desea crear una actividad. Adicionalmente, el aplicativo permite realizar seguimiento a actividades creadas y esto consiste en asignar los avances y las observaciones a la actividad.

Figura 28 - Módulo actividad.

Módulo predicción. El módulo de predicción, mostrado en la figura 29, permite aplicar un algoritmo de aprendizaje automático para estimar o predecir el presupuesto ARM de la mina teniendo en cuenta una plantilla de carga. El sistema genera los resultados mediante una tabla con la información del presupuesto y gráficos en relación con el estándar y categoría.

pais	region	mina	calificacion parcial	Tiempo Dedicacion Personal	idcriterio	Presupuesto ARM_MM
Colombia	Region	Mina	3	0.7	1	0.40623402126132985
Colombia	Region	Mina	2.5	1.4	2	0.44688580388660787
Colombia	Region	Mina	3	1.1	3	0.4593070593848399

Figura 29 - Módulo de predicción presupuesto ARM.

5.5.4. Tablero de control

Se utilizó Microsoft Power BI como herramienta para crear un tablero de control, debido a las potentes características que esta herramienta brinda; entre las que se destacan la conexión con diferentes orígenes de datos, el análisis de la información, la generación de informes y la información compartida [29]. El tablero está compuesto por una ventana en la cual se muestran los gráficos relacionados con los indicadores de los procesos realizados en la organización minera. Para esto, se estableció una conexión entre Power BI y el motor de base de datos PostgreSQL que almacena la información de gestión de las minas. Utilizando una conexión de tipo importación, se extrajo la información de la base de datos y finalmente se generó un modelo dentro de Power BI para su análisis.

Una vez obtenido el modelo y los datos de las diferentes tablas, se procedió a analizar qué campos eran necesarios para los distintos gráficos a generar. Se tuvieron en cuenta ciertos indicadores clave para el análisis de la gestión de minas. Esto permitió validar que los campos "calificacionparcial", "mina", "categorías", "módulos", "estándares" y "ods" eran datos sumamente importantes para generar gráficos que brindaran información útil a través del tablero de control para el usuario final de la ARM.

El uso de Microsoft Power BI junto con la conexión a la base de datos PostgreSQL proporcionó una poderosa herramienta para visualizar y analizar los indicadores de los procesos de la ARM. La

integración de estos datos clave en los gráficos del tablero permitió obtener una visión clara y precisa de la gestión de las minas, facilitando la toma de decisiones informadas y el seguimiento de los resultados en tiempo real.

Finalizado el conjunto de visualizaciones, se procedió a publicar el tablero en la nube permitiendo ser alojada de manera que, mediante técnicas y etiquetas HTML, se puede embeber dentro del aplicativo web usando las etiquetas de <iframe></iframe>.

6. CONCLUSIONES

La gestión de la base de datos por medio de un modelo relacional se ajustó perfectamente con la manera de levantar y almacenar los datos por parte de Alianza, lo cual era realizado anteriormente por medio de tablas estructuradas en archivos de formato Excel. Esta nueva gestión realizada en PostgreSQL permitió determinar exitosamente y de manera práctica las correspondencias entre los diferentes indicadores de cada entidad, proceso que anteriormente era desarrollado de manera manual con la integración de varios archivos independientes en formato Excel por los profesionales de la Alianza, ahorrando tiempo y mejorando la calidad de los datos, además de abrir la puerta para llevar un registro histórico de los mismos.

La creación de la base de datos relacional permitió la manipulación del conjunto de datos para la realización de análisis de estos y para la aplicación de diferentes técnicas de inteligencia artificial. Los resultados obtenidos respaldaron la utilización del modelo de regresión ponderada para predecir el presupuesto en función del tiempo de dedicación de personal. La influencia positiva y significativa de la variable independiente "Tiempo Dedicación Personal" en la predicción de la variable objetivo, “Presupuesto ARM”, proporcionó una base sólida para la toma de decisiones y análisis relacionados con el presupuesto en el contexto de estudio para la Alianza por la Minería Responsable. Sin embargo, es importante tener en cuenta la presencia de heterocedasticidad identificada durante el análisis exploratorio. Este fenómeno plantea desafíos adicionales, ya que la variabilidad no constante de los errores puede afectar la confiabilidad de las estimaciones y los intervalos de confianza. Por lo tanto, fue recomendado en este caso de aplicación, abordar la heterocedasticidad mediante técnicas apropiadas, como la aplicación de modelos de regresión heterocedásticos o la consideración de pesos adicionales en el análisis.

En cuanto al aplicativo web, se ha conseguido crear un sitio web ligero, cómodo e intuitivo para los usuarios de la organización minera que tienen que acceder a él para realizar la gestión de las minas. El prototipo web cuenta con las opciones básicas como la creación de minas, asignación de criterios, actividades, seguimientos, predicción y un tablero con indicadores importantes en la organización. Adicionalmente, se tiene un buen diseño de la Api como Back-end que permite el tratamiento de los datos en un modelo relacional permitiendo que su información sea representada en formato JSON.

7. RECOMENDACIONES

- Se recomienda la adición de nuevos campos de datos en la base de datos a partir del levantamiento de datos relacionados con características de los grupos mineros, entre otros, que permitan la identificación de otros potenciales campos de aplicación de la ciencia de datos para la Alianza por la Minería Responsable.
- El prototipo de aplicativo web queda abierto a mejoras y actualizaciones con la finalidad de generar nuevas funcionalidades que permitan optimizar los procesos en las organizaciones mineras, así como la escalabilidad del sistema.
- Adicionalmente, la propuesta de investigación queda como base para futuras investigaciones que tengan relación con organizaciones mineras y en conjunto con el aplicativo web, pueda ser instrumento de medición y exploración de nuevos datos que conlleve a nuevos descubrimientos o patrones.

8. BIBLIOGRAFÍA

[1] D. Lemus and R. Pérez, "Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos," 2020. https://revistas.uniandes.edu.co/doi/epub/10.7440/colombiaint102.2020.03

[2] Oecd, "OECD Due Diligence Guidance for Responsible supply Chains of Minerals from Conflict-Affected and High-Risk Areas," 2016, https://www.oecd.org/daf/inv/mne/OECD-Due-Diligence-Guidance-Minerals-Edition3.pdf

[3] P. Europeo, "REGLAMENTO (UE) 2017/ 821 DEL PARLAMENTO EUROPEO Y DEL CONSEJO - de 17 de mayo de 2017por el que se establecen obligaciones en materia de diligencia debida en la cadena de suministro por lo que respecta a los importadores de la Unión de estaño, tantalio y wolframio, sus minerales y oro originarios de zonas de conflicto o de alto riesgo," 2017, https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32017R0821&from=HU

[4] Craft, "Craft Mines," 2020. https://www.craftmines.org/

[5] Fairmined, "¿QUÉ ES FAIRMINED?," 2023. https://fairmined.org/what-is-fairmined/

[6] O. Restrepo Baena and L. Martinez Mendoza, "Sustainability of the Artisanal and Small-Scale Gold Mining in Northeast Antioquia-Colombia," Aug. 2021, doi: 10.3390/SU13169345.

[7] Deloitte, "Responsible Mineral Sourcing & Supply Chain Management," 2018, Accessed: Jul. 12, 2022,https://www2.deloitte.com/content/dam/Deloitte/be/Documents/energy-resources/Resp minerals artical_final.pdf

[8] J. L. Cano, "BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN," 2008, Accessed: Jul. 10, 2022, https://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf

[9] A. Menzinsky, G. López, and J. Palacio, *Scrum Manager*. 2016. http://www.streetsofdublin.com/

[10] K. Schwaber and J. Sutherland, "La Guía de Scrum," 2017, https://scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-Spanish-SouthAmerican.pdf

[11] J. F. Salazar Saldarriaga, "Diseño de un mecanismo de medición basado en Dashboards para la gestión de proyectos en la Empresa Energizando S.A.S," 2019. https://repository.upb.edu.co/handle/20.500.11912/5225

[12] S. Murcia Lemus and S. K. Torres Herrera, "DISEÑO DE UN TABLERO DE CONTROL QUE PERMITA LA IDENTIFICACIÓN TEMPRANA DE LA CORROSIÓN EXTERNA EN LOS DUCTOS DE TRANSPORTE DE GAS NATURAL," 2019. https://repository.uamerica.edu.co/bitstream/20.500.11839/7658/1/5122472-2019-2-IP.pdf

[13] L. Zuñiga Quiñones, "Desarrollo de un modelo predictivo para un sistema de manufactura orientado a la industria 4.0," 2021,

https://repository.javeriana.edu.co/bitstream/handle/10554/55631/123-attachment-1627079269Lina%20Zu%C3%B1iga%20-%20Trabajo%20de%20Grado.pdf?sequence=4

[14] H. Kaikkonen, M. Kivinen, Q. Dehaine, and J. Pokki, "Traceability methods for cobalt, lithium, and graphite production in battery supply chains," 2022. https://www.researchgate.net/publication/360473010_Traceability_methods_for_cobalt_lithium_and_graphite_production_in_battery_supply_chains

[15] F. Tang, H. Mitchell, F. Maubrey, and N. Garrett, "BLOCKCHAIN FOR TRACEABILITY IN MINERALS AND METALS SUPPLY CHAINS: OPPORTUNITIES AND CHALLENGES Blockchain for Traceability in Minerals and Metals Supply Chains: Opportunities and Challenges 2," 2017, https://www.rcsglobal.com/wp-content/uploads/2018/09/ICMM-Blockchain-for-Traceability-in-Minerals-and-Metal-Supply-Chains.pdf

[16] M. Zea, J. Molina, and F. Redobran, "ADMINISTRACIÓN DE BASES DE DATOS CON POSTGRESQL," 2017, doi: 10.17993/IngyTec.2017.18.

[17] O. Merchan Manzano, "Diseño de base de datos," 2016. https://publicaciones.uazuay.edu.ec/index.php/ceuazuay/catalog/view/5/5/1146

[18] M. Marques, "Bases de datos," 2011. https://bdigital.uvhm.edu.mx/wp-content/uploads/2020/05/Bases-de-Datos.pdf

[19] M. Olusegun Akiwande, H. Garba Dikko, and A. Samson, "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis," Dec. 2015, doi: 10.4236/OJS.2015.57075.

[20] S. Raschka and V. Mirjalili, *Python Machine Learning*, vol. 2. 2019.

[21] C. Manchini, R. Ospina, V. Leiva, and C. Martin-Barreiro, "A new approach to data differential privacy based on regression models under heteroscedasticity with applications to machine learning repository data," *Inf Sci (N Y)*, vol. 627, pp. 280–300, May 2023, doi: 10.1016/J.INS.2022.10.076.

[22] M. Taylor, "Learn Statistical Tests For Linear Regression," 2022. https://www.vexpower.com/brief/statistical-tests

[23] D. F. Rosero Guerrero, "SEGURIDAD INFORMÁTICA DEL MODELO VISTA CONTROLADOR (MVC) EN APLICACIONES PUNTO DE PAGO-POS," 2019. https://repository.unad.edu.co/bitstream/handle/10596/39026/MonografiaMVCENPOS.pdf?sequence=1&isAllowed=y

[24] M. Villa Escudero, "PROPUESTA DE UN MÉTODO DE MEJORES PRÁCTICAS PARA DE SERVICIOS WEB RESTFul," 2019, http://dspace.espoch.edu.ec/bitstream/123456789/13028/1/20T01267.pdf

[25] Kong Inc, "Introduction to Insomnia | Insomnia Docs," 2021. https://docs.insomnia.rest/insomnia/get-started

[26] Autentia, "HTML y CSS Front," 2023. https://www.autentia.com/wp-content/uploads/libros/Front_HTMLyCSS-Autentia.pdf

[27] Bootstrap team, "Bootstrap," 2023. https://getbootstrap.com/

[28] M. Haverbeke, *Eloquent JavaScript*. 2018,https://eloquentjs-

es.thedojo.mx/Eloquent_JavaScript.pdf
[29] A. Lopez Hernandez, “Presentación de datos ODataen Power BI,” 2017, http://repositorio.ual.es/bitstream/handle/10835/6619/15592_ProyectoFinal.pdf?sequence=1

9. ANEXOS

ANEXO 1. Diccionario de datos.

Nombre de la tabla		País			
Descripción		Tabla que contiene los registros de los países			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre del país
3	Código	50	Texto		Código del país

Nombre de la tabla		Región			
Descripción		Tabla que contiene los registros de las regiones			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre de la región
3	Código	50	Texto		Código de la región
4	Idpais	50	Texto	Fk	Campo numérico que se relaciona con el id de país

Nombre de la tabla		Mina			
Descripción		Tabla que contiene la información de las minas			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre de la mina
3	Idregion	50	Numero		Campo numérico que se relaciona con el id de región
4	Clasificacionmina	50	Texto		Campo que almacena la clasificación de la mina
5	Latitud	50	Texto		Ubicación(latitud) de la mina
6	Longitud	50	Texto		Ubicación(longitud) de la mina

Nombre de la tabla		Mina			
Descripción		Tabla que contiene la información de las minas			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre de la mina
3	Idregion	50	Numero		Campo numérico que se relaciona con el id de región
4	Clasificacionmina	50	Texto		Campo que almacena la clasificación de la mina
5	Latitud	50	Texto		Ubicación(latitud) de la mina
6	Longitud	50	Texto		Ubicación(longitud) de la mina
7	Trabajadores		Numero		Número de trabajadores
8	oro_mes		Numero		Cantidad de oro mes (tonelada)
9	mercurio		Texto		Tiene o no mercurio
10	certificada		Texto		Esta o no certificada
11	etapa		Texto		Esta o no en explotación

Nombre de la tabla		Categoría			
Descripción		Tabla que contiene la información de las categorías			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	PK	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre de la categoría
3	Código	50	Texto		Código de la categoría
4	Descripción	50	Texto		Descripción de las categorías

Nombre de la tabla		Criterio			
Descripción		Tabla que contiene los criterios para las valoraciones			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre del criterio
3	Código	50	Texto		Código del criterio
4	Idcategoria		Numero	Fk	Campo numérico que se relaciona con el id de categoría
5	Idmodulo		Numero	Fk	Campo numérico que se relaciona con el id de modulo

Nombre de la tabla		Estándar			
Descripción		Tabla que contiene los estándares			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre del estándar
3	Descripción	50	Texto		Descripción del estándar

Nombre de la tabla		Modulo			
Descripción		Tabla que contiene los criterios para las valoraciones			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre del modulo
3	Idestandar		Numero	Fk	Campo numérico que se relaciona con el id estándar

Nombre de la tabla		Perfil			
Descripción		Tabla que contiene los perfiles del sistema web			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Nombre	50	Texto	Único	Nombre del modulo
3	Descripción	50	Texto		Descripción del perfil

Nombre de la tabla		Usuario			
Descripción		Tabla que contiene la información de los usuarios			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	PK	Campo autonumérico incremental
2	Usuario	50	Texto	Único	Nombre del usuario
3	Contrasena	100	Texto		Contraseña del usuario
4	Nombre	50	Texto		Descripción de las categorías
5	Apellido	50	Texto		Apellido del usuario
6	Identificación	20	Texto	Único	Identificación del usuario
7	Correo	50	Texto		Correo del usuario
8	Teléfono	20	Texto		Teléfono del usuario
9	Idperfil		Numero	Fk	Campo numérico que se relaciona con el id perfil

Nombre de la tabla		Seguimiento			
Descripción		Tabla con el seguimiento del plan			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Idactividad		Numero	Fk	Campo numérico que se relaciona con el id de actividad
3	Avance	50	Numero		Avance de la actividad
4	Observación	50	Texto		Comentario de la actividad
5	Fecha		Fecha		Fecha del seguimiento

Nombre de la tabla		Actividad			
Descripción		Actividades para el plan de mejoramiento			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Iddetallemina		Numero	Fk	Campo numérico que se relaciona con el id detallemina
3	Descripción	50	Texto		Descripción de la actividad solicitada
4	Fecha	50	Fecha		Fecha de asignación de actividad
5	Producto	50	Texto		Producto de la actividad
6	Responsable	50	Texto		Responsable de la actividad

Nombre de la tabla		Detallemina			
Descripción		Tabla detalle de las minas y su criterio			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Idmina		Numero	Fk	Campo numérico que se relaciona con el id de mina
3	Idcriterio		Numero	Fk	Campo numérico que se relaciona con el id de criterio
4	Calificacionparcial		Numero		Calificación final después del seguimiento
5	Fecha		Fecha		Fecha de asignación
6	Estado	50	Texto		Estado del criterio en relación a la mina
7	Observación	50	Texto		Observaciones a las minas con el criterio asignado
8	Idusuario		Numero	Fk	Campo numérico que se relaciona con el id del usuario

Nombre de la tabla		Seguimientocriterio			
Descripción		Tabla con el seguimiento al criterio			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Idmina		Numero	Fk	Campo numérico que se relaciona con el id de mina
3	Idcriterio		Numero	Fk	Campo numérico que se relaciona con el id de criterio
4	Avance	50	Numero		Avance de la actividad
5	Fecha		Fecha		Fecha del seguimiento

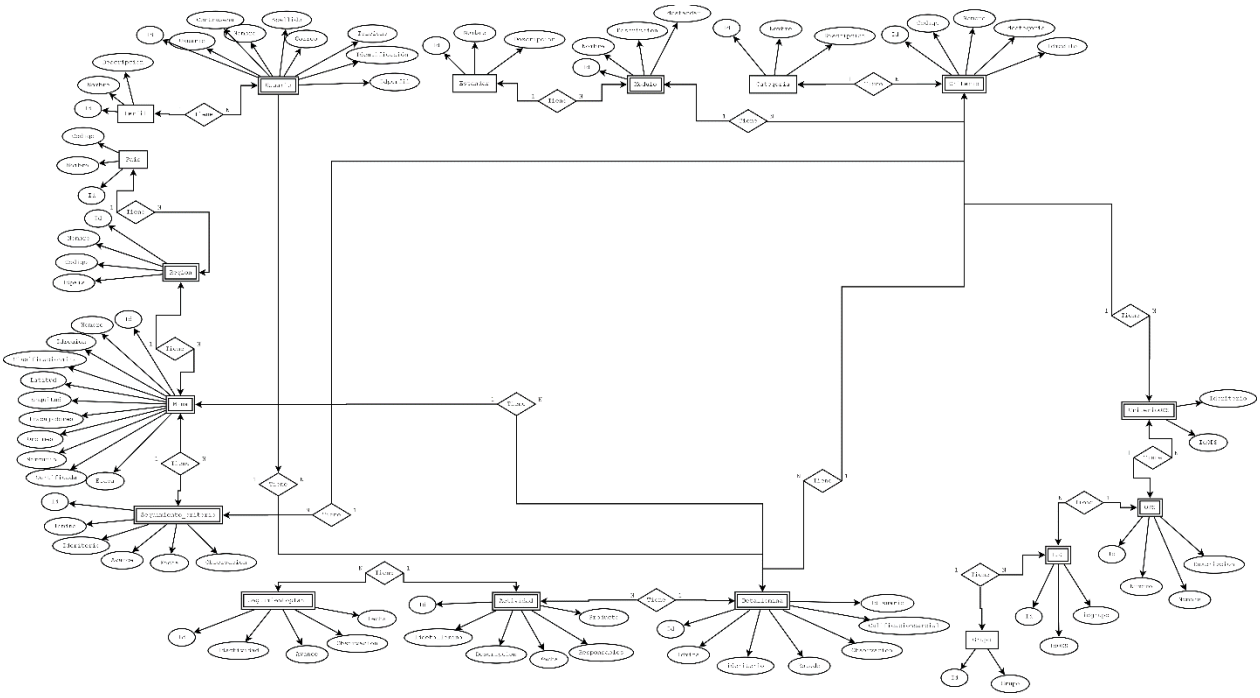
Nombre de la tabla		Grupo			
Descripción		Tabla con información del grupo			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Grupo	10	Texto	Único	Campo que almacena el grupo

Nombre de la tabla		ODS			
Descripción		Tabla con información del ODS			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Numero		Numero	Único	campo que almacena el número del grupo
3	Nombre	50	Texto		Nombre del ODS
4	Descripción	300	Texto		Descripción del ODS

Nombre de la tabla		ESG			
Descripción		Tabla con información del ODS			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	IdODS		Numero		campo que almacena el id del criterio
3	Idcriterio		Numero		campo que almacena el id del ODS

Nombre de la tabla		CRITERIO ODS			
Descripción		Tabla con información del criterio ODS			
	Campo	**Tamaño**	**Tipo de dato**	**Campo clave**	**Descripción**
1	Id		Numero	Pk	Campo autonumérico incremental
2	Idcriterio		Numero		campo que almacena el id del OS
3	IdODS		Numero		campo que almacena el id del criterio

ANEXO 2. Modelo entidad relación.



ANEXO 3. Script de la base de datos.

```
--creación de la tabla país--
create table PAIS(
        id serial primary key,
        nombre character varying(50) not null unique,
        codigo character varying(50)
        );

--creación de la tabla estándar--
create table ESTANDAR(
        id serial primary key,
        nombre character varying(50) not null unique,
        descripcion character varying(50)
        );

--creación de la tabla categoría --
create table CATEGORIA(
        id serial primary key,
        nombre character varying(50) not null unique,
        codigo character varying(50),
        descripcion character varying(50)
        );

--creación de la tabla perfil --
create table PERFIL(
        id serial primary key,
        nombre character varying(50) not null unique,
        descripcion character varying(50)
        );

--creación de la tabla región --
create table REGION(
        id serial primary key,
        nombre character varying(50) not null unique,
        codigo character varying(50) not null unique,
        idpais integer not null,
        foreign key(idpais) references pais(id)
        );
```

```
--creación de la tabla modulo --
create table MODULO(
        id serial primary key,
        nombre character varying(50) not null unique,
        descripcion character varying(50) not null unique,
        idestandar integer not null,
        foreign key(idestandar) references estandar(id)
        );

--creación de la tabla criterio --
create table CRITERIO(
        id serial primary key,
        codigo character varying(50) not null unique,
        descripcion text not null unique,
        idcategoria integer not null,
        idmodulo integer not null,
        foreign key(idcategoria) references categoria(id),
        foreign key(idmodulo) references modulo(id)
        );

--creación de la tabla mina --
create table MINA(
        id serial primary key,
        nombre character varying(50) not null unique,
        idregion integer not null,
        estado integer default 1,
        foreign key(idregion) references region(id)
        );

--creación de la tabla usuario --
create table USUARIO(
        id serial primary key,
        usuario character varying(50) not null unique,
        contrasena character varying(50) not null,
        nombre character varying(50) not null,
        apellido character varying(50) not null,
        correo character varying(50) not null,
        telefono character varying(50) not null,
        identificacion character varying(50) not null unique,
        idperfil integer not null not null,
```

```
        estado integer default 1,
        foreign key(idperfil) references perfil(id)
        );

--creación de la tabla seguimiento_criterio --
create table SEGUIMIENTO_CRITERIO(
        id serial primary key,
        idmina integer not null,
        idcriterio integer not null,
        avance character varying(50),
        fechahora timestamp without time zone,
        observacion character varying(100),
        foreign key(idmina) references mina(id),
        foreign key(idcriterio) references criterio(id)
        );

--creación de la tabla detalle_mina --
create table DETALLE_MINA(
        id serial primary key,
        idmina integer,
        idcriterio integer,
        calificacionparcial real,
        fecha timestamp without time zone,
        estado character varying(300),
        observacion text,
        idusuario integer default 1,
        foreign key(idmina) references mina(id),
        foreign key(idcriterio) references criterio(id),
        foreign key(idusuario) references usuario(id)
        );

--creación de la tabla actividad --
create table ACTIVIDAD(
        id serial primary key,
        iddetallemina integer,
        descripcion text,
        fecha timestamp without time zone,
        producto text,
        responsable character varying(50),
        foreign key(iddetallemina) references detalle_mina(id)
```

```
        );

--creación de la tabla seguimiento_plan --
create table SEGUIMIENTO_PLAN(
        id serial primary key,
        idactividad integer,
        avance character varying(20),
        observacion text,
        fecha timestamp without time zone,
        foreign key(idactividad) references actividad(id)
        );


--creación de la tabla seguimiento_criterio --
create table SEGUIMIENTO_CRITERIO(
        id serial primary key,
        idmina integer not null,
        idcriterio integer not null,
        avance character varying(50),
        fechahora timestamp without time zone,
        foreign key(idmina) references mina(id),
        foreign key(idcriterio) references criterio(id)
        );

--creación de la tabla grupo --
create table GRUPO(
        id serial primary key,
        grupo character varying(10),
        descripcion character varying(20)
        );

--creación de la tabla ods --
create table ODS(
        id serial primary key,
        numero integer,
        nombre character varying(100),
        descripcion text
        );
```

```
--creación de la tabla esg --
create table ESG(
        id serial primary key,
        idods integer,
        idgrupo integer,
        foreign key(idods) references ods(id),
        foreign key(idgrupo) references grupo(id)
        );

--creación de la tabla criterio_ods --
create table CRITERIO_ODS(
        id serial primary key,
        idcriterio integer,
        idods integer,
        foreign key(idcriterio) references criterio(id),
        foreign key(idods) references ods(id)
        );
```

ANEXO 4. Metodología Ágil.

Pila del producto.

Historia de Usuario	Tamaño estimado(horas)	Prioridad
HU1	10	Media
HU2	12	Alta
HU3	12	Alta
HU4	12	Alta
HU5	12	Alta
HU6	12	Media
HU7	10	Baja
HU8	12	Media

Sprint 1.

SPRINT 1	
Horas disponibles	24
Horas estimadas	22
Horas restantes	2
Historia	**Horas estimadas**
HU1	12
HU7	10

Historia de usuario 1.

Historia de usuario 1	
ID:HU1	**Rol**: propietario
Nombre de la historia: Iniciar sesión	
Prioridad: Alta	**Riesgo en desarrollo**: Medio
Tamaño estimado: 10	**Sprint asignado**: 1
Descripción: Como propietario quiero iniciar sesión en la aplicación para visualizar los contenidos de la aplicación.	
Criterios de aceptación Los datos deben ser validados por evitar errores de sintaxis. Si el inicio de sesión es exitoso lo lleva a la página a la que quería ingresar.	

Sprint 2.

SPRINT 2	
Horas disponibles	26
Horas estimadas	24
Horas restantes	2
Historia	**Horas estimadas**
HU2	12
HU3	12

Historia de usuario 2.

Historia de usuario 2	
ID:HU2	**Rol**: propietario
Nombre de la historia: Gestionar las minas	
Prioridad: Alta	**Riesgo en desarrollo**: Medio
Tamaño estimado: 12	**Sprint asignado**: 2
Descripción: Como propietario quiero sistematizar la información de la mina en el aplicativo	

web.
Criterios de aceptación Los datos deben ser validados por evitar errores o campos vacíos. Los datos de la mina se deben registrar en el aplicativo. Los datos de la mina se deben eliminar en el aplicativo. Los datos de la mina se deben actualizar en el aplicativo. Los datos de la mina se deben visualizar en el aplicativo.

Historia de usuario 3.

Historia de usuario 3	
ID:HU3	**Rol**: propietario
Nombre de la historia: Gestionar los criterios a las minas	
Prioridad: Alta	**Riesgo en desarrollo**: Medio
Tamaño estimado: 12	**Sprint asignado**: 2
Descripción: Como propietario quiero sistematizar la información relacionada a la asignación de cada criterio a cada una de las minas creadas.	
Criterios de aceptación Los datos deben ser validados por evitar errores o campos vacíos. Las minas registradas se deben visualizar para su selección. Los criterios registrados se deben visualizar para su selección. Cada criterio se debe asignar a cada mina. Los criterios asignados a la mina se deben visualizar en el aplicativo.	

Sprint 3.

SPRINT 3	
Horas disponibles	26
Horas estimadas	24
Horas restantes	2
Historia	**Horas estimadas**
HU4	12
HU5	12

Historia de usuario 4.

Historia de usuario 4	
ID:HU4	**Rol**: propietario
Nombre de la historia: Gestionar las actividades de las minas	
Prioridad: Alta	**Riesgo en desarrollo**: Medio
Tamaño estimado: 12	**Sprint asignado**: 3
Descripción: Como propietario quiero sistematizar la información relacionada a la asignación de las actividades de las minas en relación a sus criterios.	
Criterios de aceptación Los datos deben ser validados por evitar errores o campos vacíos. Las minas registradas se deben visualizar para su selección. Los criterios registrados en cada una de las minas se deben visualizar para su selección. Los datos de las actividades de la mina se deben registrar en el aplicativo.	

Historia de usuario 5.

Historia de usuario 5	
ID:HU5	**Rol**: propietario
Nombre de la historia: Gestionar los seguimientos de las actividades	
Prioridad: Alta	**Riesgo en desarrollo**: Medio

Tamaño estimado: 12	**Sprint asignado**: 3
Descripción: Como propietario quiero sistematizar la información relacionada a los seguimientos de las actividades creadas.	
Criterios de aceptación Los datos deben ser validados por evitar errores o campos vacíos. Las actividades registradas deben tener la opción de gestionar el seguimiento. Los datos de los seguimientos de las actividades se deben editar en el aplicativo. Los datos de los seguimientos se deben visualizar en el aplicativo.	

Sprint 4.

SPRINT 4	
Horas disponibles	26
Horas estimadas	24
Horas restantes	2
Historia	**Horas estimadas**
HU6	12
HU8	12

Historia de usuario 6.

Historia de usuario 6	
ID:HU6	**Rol**: propietario
Nombre de la historia: Generar un tablero o tablero de control con información relacionada a la gestión de las minas.	
Prioridad: Media	**Riesgo en desarrollo**: Medio
Tamaño estimado: 12	**Sprint asignado**: 4
Descripción: Como propietario visualizar un tablero o tablero de control con gráficos que permitan monitorear y analizar indicadores relacionados con la gestión de las minas.	
Criterios de aceptación	

El tablero 1 debe tener una serie de gráficos con información de las minas y su relación con la calificación parcial en base a las categorías y módulos. Adicionalmente, la cantidad de calificaciones parciales completas en tres (3) y menos de tres (3) El tablero El tablero 2 debe tener una serie de gráficos con información de los ODS en relación a las calificaciones parciales de las minas. El tablero 3 debe tener una serie de gráficos con información de la cantidad de actividades de las minas en relación a categorías, módulos y responsables.

Historia de usuario 7.

Historia de usuario 7	
ID:HU7	**Rol**: propietario
Nombre de la historia: Gestionar usuarios	
Prioridad: Media	**Riesgo en desarrollo**: Medio
Tamaño estimado: 10	**Sprint asignado**: 1
Descripción: Como propietario gestionar los usuarios aplicando operaciones como insertar, consultar, actualizar y eliminar usuarios en el aplicativo.	
Criterios de aceptación Los datos deben ser validados por evitar errores o campos vacíos. Los datos del usuario se deben registrar en el aplicativo. Los datos del usuario se deben eliminar en el aplicativo. Los datos del usuario se deben actualizar en el aplicativo. Los datos del usuario se deben visualizar en el aplicativo.	

Historia de usuario 8.

Historia de usuario 8	
ID:HU8	**Rol**: propietario
Nombre de la historia: Encontrar modelos de predicción	
Prioridad: Alta	**Riesgo en desarrollo**: Medio
Tamaño estimado: 12	**Sprint asignado**: 4
Descripción: Como propietario conocer algún tipo de predicción teniendo en cuenta la calificación parcial y el tiempo de dedicación del personal.	
Criterios de aceptación Encontrar algún tipo de correlación entre los datos de la calificación parcial y la dedicación del personal. Aplicar un algoritmo de aprendizaje automático para predecir el presupuesto del ARM.	

ANEXO 5. Consulta SQL conjunto de datos para análisis exploratorio.

```
select
pa.nombre as pais,
re.nombre as region ,
mi.nombre as mina,
mi. clasificacionmina,
mi. trabajadores,
mi. oro_mes,
mi. tipo,
mi. ton_mes,
mi. mercurio,
mi. etapa,
mi. certificada,
es.nombre as estandar,
mo.nombre as modulo,
cat.nombre as categoria,
od.numero as no_ods,
gru.grupo,
cri.id as idcriterio,
cri.descripcion,
dm.calificacionparcial,
dm.observacion,
dm.fecha as fechacalificacionparcial,
act.descripcion as actividad,
act.fecha as fechaactividad,
act.responsable,
sp.avance as avance_plan,
sp.fecha as fechaavanceplan,
sc.avance as avanc_criterio,
sc.fechahora as fechaavancecriterio
from
detalle_mina dm
full join mina mi on dm.idmina = mi.id
full join region re on mi.idregion = re.id
full join pais pa on re.idpais = pa.id
full join criterio cri on dm.idcriterio = cri.id
full join modulo mo on cri.idmodulo = mo.id
full join estandar es on mo.idestandar = es.id
full join categoria cat on cri.idcategoria =
cat.id
full join criterio_ods cods on cri.id =
cods.idcriterio
full join ods od on cods.idods = od.id
full join esg esga on od.id = esga.idods
full join grupo gru on esga.idgrupo = gru.id

full join actividad act on
act.iddetallemina=dm.id
full join seguimiento_plan sp on
sp.idactividad = act.id
full join seguimiento_criterio sc on mi.id =
sc.idmina
and cri.id = sc.idcriterio
```

Pseudo-1: Implementação Passo-a-Passo do Algoritmo de Hilbert

```
Entrada: Ordem n, Coordenadas (x, y)
Saída: Distância ao longo da curva de Hilbert
Função xy2d(n, x, y):
    d ← 0
    Para s ← n/2; s > 0; s ← s/2 faça:
        rx ← (x & s) > 0
        ry ← (y & s) > 0
        d ← d + s * s * ((3 * rx) ^ ry)
        Se ry = 0 então:
            Se rx = 1 então:
                x ← n - 1 - x
                y ← n - 1 - y
            Fim Se
            x, y ← y, x
        Fim Se
    Fim Para
    Retorne d
```

Algoritmo de Engenharia Social

Pseudo-2: Algoritmo de Engenharia Social para Manipulação de Confiança

```
Entrada: Alvo, Objetivo, Contexto
Saída: Sequência de Ações para Obter Acesso
Função engenharia_social(alvo, objetivo, contexto):
    perfil ← coletar_informacoes(alvo)
    pretexto ← criar_pretexto(perfil, contexto)
    Se pretexto = nulo então:
        Retorne "Falha: Não foi possível criar um pretexto convincente"
    Fim Se
    abordagem ← escolher_abordagem(perfil, pretexto)
    sequencia_acoes ← []
    Enquanto não objetivo_alcancado(sequencia_acoes) faça:
        proxima_acao ← gerar_proxima_acao(abordagem, sequencia_acoes)
        Se proxima_acao = nulo então:
            Retorne "Falha: Não foi possível progredir"
        Fim Se
        sequencia_acoes.adicionar(proxima_acao)
        Se detectar_suspeita(alvo, proxima_acao) então:
            Retorne "Falha: Suspeita detectada"
        Fim Se
    Fim Enquanto
    Retorne sequencia_acoes
```

Pseudo-3: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-4: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-5: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-6: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-7: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-8: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-9: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-10: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-11: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-12: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-13: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-14: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-15: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-16: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-17: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-18: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-19: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-20: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-21: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-22: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-23: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-24: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-25: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-26: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-27: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-28: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-29: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-30: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-31: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-32: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-33: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-34: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-35: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-36: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-37: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-38: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-39: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-40: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-41: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-42: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-43: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-44: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-45: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-46: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-47: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-48: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-49: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-50: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-51: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-52: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-53: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-54: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-55: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-56: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-57: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-58: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-59: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-60: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-61: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo Real

```
Entrada: Fluxo de Comunicação, Modelo de Detecção
Saída: Alertas de Ataques de Engenharia Social
Função detectar_engenharia_social(fluxo, modelo):
    Enquanto fluxo.ativo() faça:
        mensagem ← fluxo.proxima()
        características ← extrair_características(mensagem)
        pontuação ← modelo.prever(características)
        Se pontuação > limiar_detecção então:
            alerta ← criar_alerta(mensagem, pontuação)
            enviar_alerta(alerta)
        Fim Se
    Fim Enquanto
```

Pseudo-62: Algoritmo para Detecção de Ataques de Engenharia Social em Tempo