

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA GEOLÓGICA



T E S I S

**Aplicación del Machine Learning y el pronóstico de mineralización en la
concesión minera Gus del Aire, Huánuco 2023**

Para optar el título de profesional de:

Ingeniero Geólogo

Autor:

Bach. Gianmarco Antony CASTRO MELENDEZ

Asesor:

Mg. Eder Guido ROBLES MORALES

Cerro de Pasco – Perú – 2025

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA GEOLÓGICA



T E S I S

**Aplicación del Machine Learning y el pronóstico de mineralización en la
concesión minera Gus del Aire, Huánuco 2023**

Sustentada y aprobada ante los miembros del jurado:

Dr. Favio Máximo MENA OSORIO
PRESIDENTE

Mg. Saturnino Eleuterio FLORES COAGUILA
MIEMBRO

Mg. Luis Arturo LAZO PAGAN
MIEMBRO



Universidad Nacional Daniel Alcides Carrión

Facultad de Ingeniería

Unidad de Investigación

INFORME DE ORIGINALIDAD N° 308-2025-UNDAC/UIFI

La Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión en mérito al artículo 23° del Reglamento General de Grados Académicos y Títulos Profesionales aprobado en Consejo Universitario del 21 de abril del 2022, La Tesis ha sido evaluado por el software antiplagio Turnitin Similarity, que a continuación se detalla:

Tesis:

**Aplicación del Machine Learning y el pronóstico de mineralización en la
concesión minera Gus del Aire, Huánuco 2023**

Apellidos y nombres del tesista

Bach. Gianmarco Antony CASTRO MELENDEZ

Apellidos y nombres del Asesor:

Mg. Eder Guido ROBLES MORALES

Escuela de Formación Profesional

Ingeniería Geológica

Índice de Similitud

11 %

APROBADO

Se informa el Reporte de evaluación del software similitud para los fines pertinentes.

Cerro de Pasco, 13 de agosto del 2025



Firmado digitalmente por BALOMBO
GIDRO Ruben Edgar FAU
20154800046 soft
Motivo: Soy el autor del documento
Fecha: 13.08.2025 09:57:08 -05:00

DEDICATORIA

Esta tesis va dedicada a Dios por guiar mis pasos y permitirme no perder las esperanzas y la fe, para poder culminar una etapa más de mi vida.

A mis padres Raúl y Dina por creer en mí, por estar en los momentos más difíciles aconsejándome y motivándome. Enseñándome en cada escalón de mi vida el ser perseverante y respetuoso, sin su apoyo y amor no hubiera conseguido llegar a ser el profesional que ahora soy.

A mis hermanos Alex, Maxcarlos y Cyntia por su apoyo incondicional, como a mis sobrinas Amy y Airis que me hacen continuar como ejemplo en este proceso llamado vida.

A mi asesor que me instruyo a mejorar a través de los consejos y la inspiración durante el transcurso de mi tesis, quien sacrifico su tiempo al corregirme y animarme para o tirar la toalla.

A mis docentes y jefes, por su pasión de enseñar y por intuir la disciplina para guiarme por un buen camino.

En memoria de mis abuelos Petronila, Damaso y Juvencio que de donde estén siempre fueron el motivo de salir adelante.

AGRADECIMIENTO

El estudio titulado “Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023”, no habría sido posible sin el apoyo del Ing. Edgar Yoni Aire puesto que se dio acceso a diferente información geológica local de la concesión minera Gus del Aire. Por ello mi mayor agradecimiento.

De igual manera mi agradecimiento a los ingenieros Luis Alberto Pizano y Ricardo Alberto Valls Alvares pertenecientes a un grupo de investigadores dentro del área de Machine Learning aplicado en la geología y/o geociencias quienes a través de sus conocimientos fueron una inspiración para adentrarme en el aprendizaje automatizado.

Al equipo que participo en el estudio de la concesión minera Gus del aire, que en cada momento del estudio estuvieron apoyando en el proyecto.

RESUMEN

El estudio “Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023” se llevó a cabo en el distrito de Panao, provincia de Pachitea, departamento de Huánuco, en la Cordillera Oriental, a altitudes entre 3500 y 4000 m.s.n.m. Su objetivo fue predecir depósitos minerales mediante modelos de Machine Learning (ML), integrando datos geológicos, geoquímicos, estructurales y análisis espectral.

La metodología comenzó con una revisión bibliográfica, seguida de la recopilación y selección de datos para conformar una base de datos. Luego, se aplicaron técnicas de preprocesamiento y transformación, permitiendo la implementación y validación de modelos de ML, específicamente Bosques Aleatorios (RF) y Redes Neuronales Artificiales (ANN). La evaluación del rendimiento, a través de la matriz de confusión y la curva ROC, mostró una precisión del 99.91% para RF y 99.81% para ANN.

El mapa predictivo resultante identificó zonas prioritarias para la exploración minera, optimizando recursos y reduciendo riesgos económicos. Estos resultados coinciden con estudios previos, como el de Rodríguez, Coba y Mamani (2021), quienes evidenciaron que el uso del ML mejora la eficiencia en exploración geológica.

En conclusión, la aplicación de ML permitió identificar zonas con potencial mineralógico de manera eficiente y precisa, facilitando la exploración y reduciendo costos. Además, la metodología utilizada demostró ser replicable en otros proyectos geológicos, consolidándose como una herramienta clave en la minería moderna.

Palabras clave: Bosques aleatorios, geología, geoquímica, machine learning, redes neuronales artificiales.

ABSTRACT

The study "Application of Machine Learning and Mineralization Forecasting in the Gus del Aire Mining Concession, Huánuco 2023" was conducted in the district of Panao, province of Pachitea, department of Huánuco, in the Eastern Cordillera, at altitudes between 3,500 and 4,000 meters above sea level. Its objective was to predict mineral deposits using Machine Learning (ML) models, integrating geological, geochemical, structural, and spectral analysis data.

The methodology began with a literature review, followed by data collection and selection to build a database. Then, preprocessing and transformation techniques were applied, enabling the implementation and validation of ML models, specifically Random Forest (RF) and Artificial Neural Networks (ANN). Performance evaluation, using the confusion matrix and ROC curve, showed an accuracy of 99.91% for RF and 99.81% for ANN.

The resulting predictive map identified priority areas for mineral exploration, optimizing resources and reducing economic risks. These results align with previous studies, such as that of Rodríguez, Coba, and Mamani (2021), who demonstrated that ML improves efficiency in geological exploration.

In conclusion, the application of ML allowed for the efficient and accurate identification of areas with mineral potential, facilitating exploration and reducing costs. Moreover, the methodology proved to be replicable in other geological projects, establishing itself as a key tool in modern mining.

Keywords: Random forests, geology, geochemistry, machine learning, artificial neural networks.

INTRODUCCIÓN

El presente trabajo de investigación aborda la aplicación de técnicas de Machine Learning en el pronóstico de mineralización dentro de la concesión minera Gus del Aire, ubicada en el distrito de Panao, provincia de Pachitea, departamento de Huánuco. Este estudio surge como respuesta a la necesidad de optimizar los procesos de exploración minera mediante el uso de tecnologías avanzadas, integrando información geológica, geoquímica, estructural y datos espectrales para identificar zonas potencialmente mineralizadas.

La investigación tiene un enfoque aplicado, no experimental y tecnológico, y está orientada a resolver un problema práctico en el ámbito geológico mediante la implementación de herramientas digitales de análisis. Con un marco metodológico descriptivo-correlacional y predictivo, se busca caracterizar las variables geológicas del área de estudio y establecer su relación con la ocurrencia de mineralización, empleando algoritmos de aprendizaje supervisado, como Bosques Aleatorios y Redes Neuronales Artificiales.

El documento se estructura en cuatro capítulos principales: el Capítulo I describe el problema de investigación, los objetivos y la justificación del estudio; el Capítulo II presenta el marco teórico y los antecedentes relevantes; el Capítulo III detalla la metodología utilizada, incluyendo el diseño de los modelos predictivos y las técnicas de procesamiento de datos; finalmente, el Capítulo IV aborda la presentación y discusión de los resultados obtenidos.

Mediante este trabajo, se busca contribuir al desarrollo de herramientas replicables y eficaces en la exploración minera, promoviendo la aplicación de tecnologías

innovadoras que mejoren la toma de decisiones y reduzcan los riesgos asociados a la evaluación de recursos minerales.

INDICE

DEDICATORIA

AGRADECIMIENTO

RESUMEN

ABSTRACT

INTRODUCCIÓN

INDICE

INDICE DE TABLAS

INDICE DE FIGURAS

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1.	Identificación y determinación del problema	1
1.2.	Delimitación de la investigación	3
1.3.	Formulación del problema.....	3
1.3.1.	Problema general	3
1.3.2.	Problemas específicos.....	3
1.4.	Formulación de objetivos	4
1.4.1.	Objetivo general.....	4
1.4.2.	Objetivos específicos	4
1.5.	Justificación de la investigación	5
1.6.	Limitaciones de la investigación	5

CAPÍTULO II

MARCO TEÓRICO

2.1.	Antecedentes de estudio	6
2.2.	Bases teóricas – científicas.	18

2.3.	Definición de términos básicos.....	37
2.4.	Formulación de hipótesis	40
2.4.1.	Hipótesis general	40
2.4.2.	Hipótesis específicas.....	40
2.5.	Identificación de variables	41
2.6.	Definición operacional de variables e indicadores	41

CAPITULO III

METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN

3.1.	Tipo de investigación	43
3.2.	Nivel de investigación	43
3.3.	Métodos de investigación	44
3.4.	Diseño de la investigación.....	45
3.5.	Población y muestra	46
3.6.	Técnicas e instrumentos de recolección de datos	47
3.7.	Técnicas de procesamiento y análisis de datos.....	49
3.8.	Tratamiento estadístico.....	51
3.9.	Orientación ética, filosófica y epistémica	52

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1.	Descripción del área de trabajo	53
4.2.	Presentación, análisis e interpretación de resultados.....	68
4.3.	Prueba hipótesis.....	86
4.4.	Discusión de resultados	89

CONCLUSIÓN

RECOMENDACIONES

REFERENCIA BIBLIOGRAFICA

ANEXOS

INDICE DE TABLAS

Tabla 1. Abundancia de elementos de la corteza continental superior.....	36
Tabla 2 Clasificación de los elementos	37
Tabla 3 Operacionalización de variables.....	42
Tabla 4. Coordenadas WGS84 de la concesión minera Gus del Aire	53
Tabla 5. Vías de acceso al proyecto de investigación “Gus del Aire”	54
Tabla 6 Estadística descriptiva	62
Tabla 7 Cálculo de background y Threshold.....	63
Tabla 8 Capas de entrada usadas para generar mapas de prospectividad mineral (MPM) de depósitos tipo vetiforme usando Machine Learning.....	70
Tabla 9 Capas de entrada para los algoritmos RF y ANN.....	71
Tabla 10 Clasificación de las variables de entrada para los algoritmos de Machine Learning.....	73
Tabla 11 Interpretación de los valores de coeficientes de la matriz de correlación	75
Tabla 12 Hiperparámetros definidos del modelo RF.....	76
Tabla 13 Hiperparámetros definidos del modelo ANN.....	80

INDICE DE FIGURAS

Figura 1. Diferencia entre el estudio supervisado vs el estudio no supervisado	19
Figura 2. Gráfico de underfitting, overfitting y un modelo	22
Figura 3. Aplicación de la precisión y recall.....	23
Figura 4. Gráfica explicativa de las diferentes posibilidades de la curva ROC	25
Figura 5. Gráfica detallada de un árbol de decisiones.....	26
Figura 6. Gráfica de los bosques aleatorio	27
Figura 7. Estructura y elementos que conforman un modelo de las redes neuronales artificiales	28
Figura 8. Abundancia relativa de los elementos químicos en el sistema solar	35
Figura 9. Esquema de metodología de trabajo	45
Figura 10. Tratamiento estadístico del estudio.....	52
Figura 11. Columna estratigráfica de la geología local en la concesión minera “Gus del Aire”	57
Figura 12. Relación entre la distribución no normalizada y normalizada de Au	59
Figura 13. Relación entre la distribución no normalizada y normalizada de Ag	59
Figura 14. Relación entre la distribución no normalizada y normalizada de Zn.....	60
Figura 15. Relación entre la distribución no normalizada y normalizada de Pb.....	60
Figura 16. Relación entre la distribución no normalizada y normalizada Cu	61
Figura 17. Aplicación de la ventana deslizante	65
Figura 18. Características de las capas de salida.....	71
Figura 19. Matriz de correlación para las variables predictorias	74
Figura 20. Matriz de confusión para el modelo RF	77
Figura 21. Características importantes de las variables para el modelo RF	78
Figura 22. Curva ROC de RF	79

Figura 23. Matriz de confusión para le modelo ANN	81
Figura 24 Curva ROC de ANN	82
Figura 25 Mapa de predicción con RF	83
Figura 26. Mapa de predicción con ANN	84
Figura 27 Curva ROC de comparación entre RF y ANN.....	86

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema

La exploración minera es una de las actividades importantes que se realiza en el mundo minero el cual a través del avance de la tecnología y de la ciencia mediante la aplicación de la inteligencia artificial al mundo moderno esto viene afectando a diferentes tipos de profesionales, como el caso de los ingenieros geólogos a través del Machine Learning ya que es la aplicación de algoritmos que nos sirve para identificar y/o elaborar predicciones y/o pronósticos. (Brandmeier et al., 2020)

Si bien el Machine Learning está teniendo mucho impacto en el siglo XXI, actualmente son pocos geólogos quienes aplican esta herramienta en la búsqueda y/o predicción de targets de exploración favorables en el Perú, el cual lleva a los estudios geológicos a una optimización de análisis, interpretación y de costos. (Rodríguez et al., 2014).

Luego de los diferentes trabajos de prospección geológica llevados a cabo en la concesión minera Gus del Aire como: geología, geoquímica, tectónica,

imágenes satelitales de alta resolución, el principal problema radica en lograr una predicción confiable ya que se tendrá en cuenta la precisión y exactitud, capacidad de manejar gran cantidad de data, validación y técnicas de evaluación, reducción del error humano y el análisis de error de áreas con mineralización de interés económico a partir de la data disponible, para ello desarrollaremos algoritmos de aprendizaje automático o machine learning que es un campo de la inteligencia artificial que nos ayudara a generar un modelado predictivo.

Entre los principales factores que dificultan el desarrollo de las actividades de exploración se encuentran:

Escasez de información local confiable: La falta de datos históricos detallados y precisos sobre la geología local limita el desarrollo de modelos predictivos robustos, obligando a depender de información secundaria o incompleta.

Manejo de grandes volúmenes de datos: La exploración moderna genera enormes cantidades de datos provenientes de estudios geoquímicos, estructurales y espectrales. Procesar y analizar esta información requiere herramientas computacionales avanzadas que minimicen el error humano.

Limitaciones en la implementación de algoritmos de Machine Learning: Aunque estas herramientas han demostrado ser eficaces, su aplicación en el Perú sigue siendo limitada debido a la falta de experiencia técnica y a la necesidad de adaptar los algoritmos a las características específicas de los datos geológicos locales.

1.2. Delimitación de la investigación

1.2.1. Delimitación espacial

Esta investigación se desarrollará en la concesión minera Gus del Aire que tiene una extensión de 500 hectáreas. Ubicada en el distrito de Panao, Provincia de Pachitea, departamento de Huánuco ubicándose entre 8880000 Norte y 415000 Este.

1.2.2. Delimitación Temporal

La presente investigación se desarrollará en los meses de agosto del 2023 a octubre del 2024.

1.2.3. Delimitación social

La investigación se realiza para los dueños de la concesión minera y la misma población del departamento de Huánuco, provincia de Pachitea y distrito de Panao.

Con el fin de impulsar la investigación geológica, generar prospectos mineros y futuros targets de exploración en la búsqueda de mineralización favorable.

1.3. Formulación del problema

1.3.1. Problema general

¿Cuál es el modelo de predicción en la mineralización en la concesión minera Gus del Aire, luego de la aplicación de algoritmos de machine learning?

1.3.2. Problemas específicos

- ¿De qué manera se desarrolla una matriz de confusión para el buen desempeño en la validación del algoritmo en el pronóstico de mineralización?

- ¿Cuál es la importancia del criterio de la curva ROC en la evaluación del pronóstico de mineralización?
- ¿Qué efecto tendría el uso del Machine Learning en la reducción de tiempo de los resultados del análisis?

1.4. Formulación de objetivos

1.4.1. Objetivo general

Predecir el modelo de mineralización en la concesión minera Gus del Aire, luego de la aplicación de algoritmos de machine learning.

1.4.2. Objetivos específicos

- Analizar el desempeño de los algoritmos de Machine Learning aplicados mediante la construcción de matrices de confusión, evaluando la precisión y la capacidad predictiva de los modelos en relación con los datos geológicos, como concentraciones geoquímicas, proximidad a fallas y características litológicas.
- Generar y comparar curvas ROC para la validación de los algoritmos predictivos, identificando la sensibilidad y especificidad de cada modelo y evaluando su capacidad para pronosticar zonas de mineralización favorable, utilizando como base los datos geológicos y espectrales recolectados.
- Evaluar el impacto del uso del Machine Learning en la eficiencia de la exploración minera, determinando cómo la integración de datos geológicos y estructurales en los algoritmos puede reducir significativamente los tiempos de análisis y mejorar la toma de decisiones en el área de estudio.

1.5. Justificación de la investigación

El proyecto propone aplicar técnicas de Machine Learning para el pronóstico y la identificación de zonas mineralizadas en la concesión minera Gus del Aire, aprovechando las características geológicas de la zona. Esto permitirá comprender el comportamiento geológico mediante algoritmos avanzados, optimizando la precisión en la predicción de mineralización.

Asimismo, el estudio contribuirá a futuros trabajos de investigación, demostrando la utilidad del Machine Learning en proyectos de distintas escalas y áreas de la geología, fortaleciendo su aplicación en la exploración minera.

1.6. Limitaciones de la investigación

- La investigación está limitada a la concesión minera Gus del aire.
- La existencia de data histórica es escasa el cual se recolecto información local de la zona, por lo tanto, no solo nos limitaremos al uso de data obtenida por Geocatmin sino también a los mapeos y anotaciones obtenidas en campo.
- Siendo Python un programa abierto los algoritmos, los códigos se actualizan y se incrementan trayendo a esto la modificación de los cogidos del lenguaje de programación en Python, por lo tanto, se debe estar actualizado con los cambios y mejoras.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. Antecedentes internacionales

Carrasco Pardo, D. (2021). La inteligencia artificial en la investigación minera.

La Inteligencia Artificial (IA) es una tecnología disruptiva con enorme potencial, que está siendo incorporada a pasos agigantados en todos y cada uno de los sectores de la economía. La adopción por parte de la industria minera, y más específicamente en las labores prospectivas, sin embargo, está siendo mucho más lenta. Aún son pocas las empresas mineras que utilizan IA en labores prospectivas. En este proyecto se trabajó en este aspecto, investigando la aplicabilidad de este tipo de técnicas en la investigación minera. Se construyeron cuatro modelos distintos de IA en este proyecto. Una vez entrenados, estos resultaron capaces de predecir con éxito mineralizaciones en la zona de estudio. El conjunto de datos utilizado en este proceso fue enteramente preparado y digitalizado en este trabajo, a partir de datos históricos de las mineralizaciones de

mercurio del distrito minero de Almadén. Este proceso de investigación y digitalización de los datos fue un proceso largo, que ha posibilitado la creación de un gran conjunto de datos con información minera del Sinclinal de Almadén. Dentro de este conjunto, están recogidos todos los indicios, sondeos y las mineralizaciones conocidas de mercurio del sinclinal, debidamente georreferenciados. En este proyecto, se realizaron dos experimentos distintos con cada modelo. El primero de ellos consistía en la identificación exacta del indicio o mineralización correspondiente a una determinada localización geográfica. En este primer experimento se obtuvieron precisiones muy elevadas con algunos de los modelos, dejando de manifiesto la gran capacidad de predicción de estos. El segundo experimento consistió en un proceso de reconstrucción del mapa geológico de la zona de estudio, por cada uno de los modelos. Esto fue muy interesante, ya que además de realizar un análisis estadístico, fue posible realizar una representación muy visual de las predicciones obtenidas. Este trabajo ha incluido una comparación del rendimiento obtenido con cada uno de los cuatro modelos en cada experimento. Esto posibilitó el estudio en detalle del comportamiento de éstos en cada situación, haciendo posible la extracción de conclusiones acerca de cuál de ellos funciona mejor en cada escenario. Se piensa que los modelos generados en este proyecto podrán ser utilizados en el futuro para la predicción de datos de índole geológica y minera de mineralizaciones de elementos metálicos, independientemente del metal a prospectar.

Aldás Núñez, R. J., Tuz Chamorro, K. V., Vega Ocaña, J. A., Velasco Haro, M. S., & Mejía Escobar, C. I. (2022). Delimitación automática de ceniza volcánica en imágenes satelitales mediante Deep Learning.

La Inteligencia Artificial ha tenido un gran impacto en los últimos años, cada vez este campo de la Informática es más utilizado en el tratamiento de problemas en el área geológica. Una de las principales aplicaciones es la detección y segmentación de ceniza en imágenes satelitales. Para tal fin, proponemos un modelo de Deep Learning basado en una red neuronal convolucional (CNN) y entrenado con un dataset de imágenes satelitales que tienen aplicado el filtro “ash”, que proporciona una coloración rosada rojiza a la ceniza, facilitando el proceso de segmentación. Los resultados obtenidos indican una precisión del 99%, conveniente para su aplicación práctica para la segmentación de la ceniza emitida por el Volcán Sangay, el cual ha presentado periodos de actividad volcánica en los últimos años. Las imágenes segmentadas generadas por nuestro modelo son congruentes con los estudios publicados por el IG-EPN.

Mantilla Dulcey, A. G. (2023). Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático.

Los mapas de prospectividad mineral son un insumo esencial para delimitar zonas potenciales fuentes de materia prima en el nuevo paradigma de transición energética. Las técnicas usadas generalmente consideran la regresión logística, los pesos de evidencia, los métodos heurísticos y la estadística bivariada. No obstante, en los métodos heurísticos se deben asignar pesos basados en el criterio experto, en los bivariados se deben clasificar las capas de valores continuos a valores categóricos o nominales, y finalmente la regresión logística considera una clasificación lineal de los datos. Por otro lado, los métodos de aprendizaje automático proporcionan un enfoque flexible y preciso en el uso de variables continuas y permiten encontrar modelos que aproximan de manera

adecuada un conjunto de datos o factores condicionantes con valores de probabilidad relacionados con la ocurrencia mineral. En este trabajo se presenta una aplicación detallada y sistemática de los métodos de aprendizaje automático redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial.

Adicionalmente, se realizó un flujo preprocesamiento (estandarización, análisis de componentes principales, análisis de distribución) de las variables de entrada, con el objetivo de representar el problema de predicción mineral con un problema de clasificación con superficies de decisión. Para comprobar la efectividad y el flujo de trabajo propuesto se analizó una zona del territorio de Yukón (Canadá), donde se inició a partir de una revisión exhaustiva de las variables o criterios mapeables de exploración relacionados con ocurrencia minerales. Los mapas de probabilidad calculados tuvieron una precisión promedio de 98.96% y la evaluación del rendimiento de los modelos de aprendizaje automático se realizaron de manera cuantitativa usando matrices de confusión, curvas ROC y precisión. Finalmente, para mejorar los modelos basados en información o criterio geológico, se propuso la técnica de aumento de datos y la aritmética de bandas entre los mapas de probabilidad, lo que permitió delimitar un total de 4 zonas potenciales. Por esta razón, la incorporación del aprendizaje automático en el flujo de trabajo de exploración mineral supone una mejora considerable en la optimización de recursos y el grado de confiabilidad en los objetivos de exploración.

Morales Madariaga, O. A. (2023). Aplicación de métodos de machine learning para la exploración geoquímica de tierras raras en yacimientos tipo IOCG, distritos Candelaria–Punta del cobre y Mantoverde, región de Atacama, Chile.

En esta tesis se utilizaron algoritmos de Machine Learning no supervisados, específicamente K-means y Agrupamiento Jerárquico, en conjunto con Análisis de Componentes Principales, para buscar patrones ocultos en la distribución de elementos de Tierras Raras (REE) en los depósitos tipo IOCG de los distritos mineros Candelaria-Punta del Cobre y Mantoverde. La base de datos analizada (81 muestras) se obtuvo de la literatura, y se procesó utilizando técnicas estadísticas y de minería de datos para identificar grupos similares de elementos. Los resultados obtenidos se analizaron para establecer su relación con la geología local. Se evaluó también la capacidad de ambos algoritmos para detectar patrones de agrupamiento en los datos, y la calidad de estos agrupamientos, empleando distintas métricas. Ambos algoritmos consistentemente agruparon a las REE en dos conjuntos: un grupo que reúne a algunas REE ligeras (La, Ce y Nd) junto con los óxidos Fe_2O_3 y MgO ; y un grupo que reúne a todas las REE pesadas (Gd, Dy, Ho, Er, Tm, Yb, y Lu) con algunas REE ligeras (Pr, Sm, Eu) y Na_2O . Los resultados obtenidos demostraron que tanto K-means como Agrupamiento Jerárquico son eficaces para identificar agrupamientos de elementos que son geoquímicamente coherentes, tanto por su afinidad química como por su relación con asociaciones de alteración y mineralización descritos en los distritos estudiados y otros depósitos tipo IOCG. Las asociaciones de elementos obtenidos no son concluyentes en cuanto a los minerales que hospedarían la mineralización de REE, por lo cual se interpreta la presencia de diferentes minerales portadores de REE. En base a lo obtenido, se concluye que los algoritmos utilizados son una herramienta útil para identificar asociaciones de elementos relacionados a la presencia de REE, por lo cual pueden ser de gran ayuda para identificar guías de exploración de estos elementos.

Hidalgo Gajardo, A. E. (2023). Métodos de machine learning aplicados a la química mineral de cloritas para su uso en exploración de pórfidos cupríferos.

La química de la clorita es considerada una herramienta de vectorización a pórfidos cupríferos. En El teniente, este mineral fue estudiado por Wilkinson et al. (2020) y Schulz (2020). Wilkinson et al. (2020) realizaron una regresión simple de la distancia al centro de El Teniente a partir de la razón $\ln(\text{Ti}/\text{Sr})$ en la clorita, con un error de ± 640 m (1σ). Además de la química mineral, Schulz (2020) investigó las cloritas del Adit 71 (el antiguo túnel de acceso) mediante microscopía óptica y definió tres colores de interferencia (verde-grisáceo, marrón-anaranjado y azul-violeta), cuya distribución se relaciona con los dominios hidrotermales y con la distancia a la Brecha Braden, aunque el control químico sobre esta propiedad óptica aún no está del todo comprendido. Las muestras del Adit 71 registran la zonación de sulfuros del sistema porfídico, caracterizada por un halo de mena ($> 1\%$ calcopirita y $< 3\%$ pirita) proximal, seguido de un halo pirítico ($> 2\%$ pirita and $\leq 1\%$ calcopirita) y un halo periférico ($< 1\%$ calcopirita y pirita) hacia sectores más distales.

Considerando estos antecedentes y el auge de la inteligencia artificial, los análisis de las cloritas de El teniente han sido refinados utilizando algoritmos de machine learning (ML), para mejorar su uso como vector. Dada la alta incidencia de contaminación mineral reportada en microanálisis de cloritas, y a que los modelos de ML dependen de la calidad de los datos, es necesario estudiar el problema de contaminación y establecer un criterio de control de calidad eficiente. En la primera parte de esta Memoria, se realizó un análisis detallado de la contaminación mineral y de los métodos de control de calidad actuales. Se utilizaron herramientas estadísticas clásicas y se evaluaron aspectos texturales,

mineralógicos, analíticos, y mineraloquímicos. La búsqueda de vectores de contaminación y de cambios en los patrones de distribución de elementos químicos, observados en diagramas de dispersión, permitieron reconocer y caracterizar la contaminación por titanita, rutilo, circón, entre otros. Además, permitió discutir la variabilidad química de Ti, Zr, Ca, K y Na en microanálisis de cloritas. Los hallazgos encontrados fueron incorporados en la etapa de depuración de los análisis de cloritas de El teniente, recopilados para su modelamiento con ML. En la segunda parte, se realizó un análisis exploratorio y se construyeron los siguientes modelos, utilizando algoritmos de ML: dos modelos para clasificar el color de interferencia de la clorita, uno a partir de su composición mayoritaria y variables cristaloquímicas, y otro a partir de los análisis LA-ICP-MS; un modelo para clasificar la zona de sulfuros en la que se encuentra una clorita, a partir de los análisis LA-ICP-MS; un modelo regresivo para predecir la distancia al centro de El Teniente a una muestra de clorita, a partir de su composición química. Todos los modelos fueron contruidos utilizando variables (cristalo-)químicas de la clorita como variables predictoras. Para los modelos de clasificación se utilizó el algoritmo CART y solo se incluyeron las cloritas del Adit 71. El modelo regresivo utilizó los análisis LA-ICP-MS de cloritas de El teniente de ambos trabajos, y se abordó mediante el algoritmo Random Forest (RF). El color de interferencia de la clorita se relacionó con la proporción de Fe en el sitio M1, el cual distingue el verde-grisáceo del azul-violeta. El color marrón-anaranjado fue reconocido por enriquecimientos sutiles en Ni+Zn a razones Fe/Mg intermedias; el comportamiento de Ni y Zn en las cloritas de este color es similar al de otros metales de transición divalentes (ej. Mn), y coincide con un empobrecimiento de cationes trivalentes. Dado que estos

elementos varían con la distancia al centro de pórfidos de cobre, el color de interferencia de la clorita es una potencial herramienta de vectorización de bajo costo. Los halos sulfurados fueron clasificados mediante la razón Ti/Na, V y Co: el halo de mena se distingue por altos contenidos de V y bajas razones Ti/Na, mientras que mayores razones Ti/Na y bajos contenidos de Co son característicos de las cloritas del halo pirítico. Un análisis posterior sugiere que la clorita monitorea los procesos de mineralización metálica, al hospedar Co (y posiblemente otros metales) en ausencia de sulfuros, deprimiéndose en este elemento cuando co-cristaliza con pirita. El modelo de regresión permitió disminuir cerca de tres veces el error de la regresión simple de Wilkinson et al. (2020), siendo las variables más importantes el Ti y Sr; este modelo considera tanto el gradiente termal del depósito como los procesos de mineralización metálica, haciéndolo robusto frente a sistemas infértiles. Al reentrenar el modelo regresivo con datos de cloritas y epidotas, se obtuvieron mejores rendimientos, reconociéndose el potencial de una aproximación multi-mineral en la exploración. En conclusión, los algoritmos de ML permiten expandir el conocimiento de los factores que controlan la composición química de la clorita, así como mejorar el uso de este mineral como herramienta de vectorización.

2.1.2. Antecedentes nacionales

Castro Andrade, J. A., & Calderón Celis, J. M. (2021). Modelo geometalúrgico de la caracterización de arcillas para la eficiencia de la flotación usando la metodología de machine learning.

El modelo geometalúrgico es el resultado de la integración de las disciplinas de geología, minería y metalurgia con la finalidad de agregar valor y disminuiría el riesgo en un proceso como por ejemplo la eficiencia de flotación

de calcopirita. El objetivo de este estudio fue calibrar modelos para cuantificar el contenido de arcillas y gangas de un yacimiento hidrotermal del tipo pórfido cobre-oro por espectroscopía NIR. La metodología seleccionada fue diseñar un modelo quimiométrico basado en 173 compositos de taladros diamantinos a los que se le realizó ensayos de difracción de rayos x y espectroscopia de infrarrojo cercano, validándose los resultados con una validación cruzada a través de una metodología machine learning. La creación de los modelos se realizó mediante una regresión no lineal regularizada por el método de Ridge. Se obtuvo modelos con baja linealidad para minerales de calcita y plagioclasa, con valores de R^2 (0.51 y 0.78, respectivamente). El modelo de regresión presenta una linealidad para minerales de esmectita, cuarzo, ortoclasa y muscovita mostró un alto R^2 (0.95, 0.93, 0.64 y 0.59, respectivamente). Los resultados encontrados para el contenido de arcillas y gangas indican que los análisis de difracción de rayos x se pueden sustituir, en gran parte, por los modelos espectrales. En el caso de la calcita y plagioclasa, sería conveniente realizar una campaña de caracterización, con la finalidad de mejorar el modelo y poder sustituir los análisis de difracción de rayos x para estas especies, lo que permitirá generar un modelo geometalúrgico de una forma rápida y eficiente con un método semi cuantitativo.

Bazán Santa Cruz, O. (2021). Innovaciones estadísticas y redes neuronales artificiales aplicadas a la prospección geoquímica en los cuadrángulos de Chaparra y Atico, Caravelí – Arequipa.

La prospección geoquímica siempre ha sido y seguirá siendo una de las herramientas más efectivas para la búsqueda de recursos minerales, que se integra con la cartografía geológica base y otras disciplinas como la teledetección y la geofísica. Esta investigación pretende impulsar una nueva metodología de

prospección geoquímica combinando el índice de enriquecimiento relativo local (LREI) de elementos indicadores de mineralización (Cu, Au, Ag, Pb, Zn y Mo), coeficientes de rango correlacional (RCC) y el análisis de componentes principales (PCA), a través de superposiciones difusas (fuzzy modelling) y el uso de aprendizaje profundo (deep learning) no supervisado, para generar mapas de prospectividad geoquímica en los cuadrángulos de Chaparra (32o) y Atico (33o) en la provincia de Caravelí, Arequipa. El mapa de prospectividad generado mediante superposición difusa y el mapa autoorganizado (SOM) fueron validados plotando la ubicación de los depósitos y ocurrencias minerales conocidos, los cuales en la mayoría de los casos se plotan dentro de áreas con alta o moderada prospectividad. Por lo que se concluye que las innovaciones estadísticas univariantes, multivariantes e innovaciones tecnológicas aplicadas a la cartografía geoquímica son herramientas efectivas para la prospección geoquímica.

Rodríguez Mejía, R., Coba Palomino, L. A., & Mamani Huisa, M. I. (2021). El machine learning de la carta geológica nacional 50k y la exploración minera.

La constante actualización de información que puede ser utilizada para mejorar la Carta Geológica Nacional, obliga a que esta se actualice constantemente y forme la base para el desarrollo socio-económico del país.

Actualmente la implementación de inteligencia artificial está en todos los sectores; en ese sentido, los usuarios de Ingemmet necesitan una Carta Geológica Nacional inteligente que les permita interactuar con los atributos de las bases de datos para obtener mejores resultados. La integración de la Carta Geológica Nacional 50k es el primer paso para crear inteligencia artificial. Por ello,

Ingemmet viene implementado el machine learning de la Carta Geológica Nacional integrada a lo largo de todo el territorio.

Las aplicaciones del machine learning de la Carta Geológica Nacional son tangibles y contribuyen a la reactivación económica en diferentes sectores, especialmente en la exploración de depósitos minerales.

Mamani Sacsi, J. N. (2022). Aplicación de redes neuronales como algoritmos de Machine Learning para la predicción de recuperaciones auríferas mediante cianuración por Agitación en Yauca-Caravelí.

La minería atraviesa grandes desafíos hoy en la actualidad y junto a ella, la tecnología trae nuevos avances como la Inteligencia artificial “IA” que está en pleno crecimiento y desarrollo para ofrecer la resolución de diferentes problemas en la Industria en general. Machine Learning o aprendizaje automático, es un subcampo de la IA, donde las máquinas o sistemas computacionales aprenden en base a información en pruebas de ensayo y error. Este sub campo nos ofrece diferentes algoritmos que realizan el proceso de aprendizaje, uno de los más importantes por su gran acogida, constante innovación y crecimiento son las Redes Neuronales artificiales quienes simulan el proceso de aprendizaje de las redes neuronales biológicas, mediante estímulos, transformación de información y salidas de respuesta.

En el presente trabajo formaremos las redes neuronales con el lenguaje de programación Python en Google Colab utilizando Keras y tensorflow, además de apoyarnos en otras librerías para la Ciencia de datos e inteligencia artificial. Logrando obtener un modelo de predicción de recuperaciones auríferas.

El trabajo se desarrollará de esta manera: se realizará el análisis descriptivo de la data de estudio que serán los datos de las pruebas metalúrgicas

de cianuración por agitación; procesamiento de datos; modelamiento de la red neuronal con Tensorflow y Keras, optimización por hiper parámetros; optimización con datos estandarizados y diferentes topologías de la red; evaluación del modelo con una validación cruzada de Scikit Learn; por último la fase de predicción y el correcto guardado del entrenamiento en formatos H5 y Json.

Castillo Liberato, O. A. (2022). Desarrollo de modelos predictivos de regresión en la industria minera mediante el uso de algoritmos de machine learning.

Una de las características que las empresas mineras siempre buscan minimizar es el alto riesgo que implica operar un proyecto. Estos pueden estar influenciados por factores geológicos, ambientales, sociales, de mercado, operativos, políticos, etc. Para afrontar estos riesgos y aumentar sus beneficios, muchas empresas han iniciado la transformación digital implementando tecnologías basadas en Inteligencia Artificial. Los resultados obtenidos han mejorado significativamente la productividad, el consumo energético, reducido los niveles de contaminación y optimizado la seguridad de sus trabajadores. La naturaleza propia de cada operación minera y el conocimiento limitado de estos temas en muchas operaciones actuales del Perú, nos incentivó a realizar modelos predictivos de regresión en casos de la industria minera mediante el uso de algoritmos de Machine Learning, un sub campo de la inteligencia artificial, utilizando un flujograma más didáctico e intuitivo. El estudio propone soluciones en tres casos de la industria minera a través de modelos predictivos de regresión. Estos modelos permiten estimar con un alto grado de certeza: el precio de oro, el contenido de sílice en el concentrado de hierro y el consumo de combustible en

camiones mineros de gran tonelaje. Este estudio es de enfoque cuantitativo y del tipo exploratorio, lo que permitió desarrollar estos modelos usando diferentes algoritmos de machine learning a partir de datos históricos. Según las métricas de evaluación, los modelos que tuvieron un buen desempeño para predecir el precio de oro y el consumo de combustible se desarrollaron usando los algoritmos SVR y Árbol de regresión. Sin embargo, para predecir el contenido de sílice en el concentrado de hierro fue necesario el uso del algoritmo de ensamble Random Forest, principalmente por la complejidad del caso. Finalmente, es importante mencionar que para llegar a los resultados esperados hace falta un trabajo estructurado que implica conocimiento de ciencia de datos, conocimiento de minería, y un entendimiento claro del problema que se resuelve usando algoritmos de machine learning.

2.2. Bases teóricas – científicas.

2.2.1. Machine learning

El Machine Learning o llamado también aprendizaje automático es una rama de la inteligencia artificial (IA) que implica el desarrollo de algoritmos y modelos que permiten a las computadoras aprender patrones y realizar tareas sin estar programadas explícitamente (Goodfellow et al., 2016). En lugar de seguir instrucciones específicas, los sistemas de aprendizaje automático pueden analizar datos, identificar patrones y aprender de la experiencia para mejorar su desempeño en una tarea específica (Goodfellow et al., 2016). Existen muchos enfoques diferentes para el aprendizaje automático, pero los dos más populares son el aprendizaje supervisado y el no supervisado.

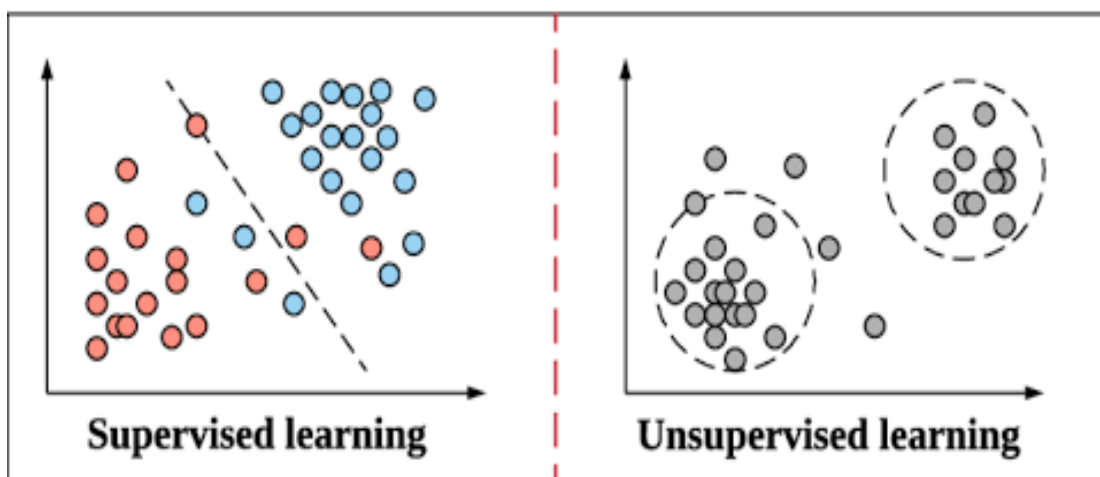
- **Aprendizaje supervisado:** En este enfoque, el algoritmo se entrena utilizando un conjunto de datos etiquetados, donde cada registro del conjunto

de datos está asociado con una etiqueta o resultado deseado. El algoritmo aprende a hacer coincidir la entrada con la salida para poder hacer predicciones precisas a partir de datos sin etiquetar (Kuncheva 2004; Hastie et al. 2009).

- **Aprendizaje no supervisado:** En este caso, el algoritmo trabaja con datos sin etiquetar y debe encontrar patrones o estructuras por sí solo. El objetivo es comprender la estructura interna de los datos, como la agrupación o la reducción de dimensionalidad (Barlow, 1989; Kuncheva 2004; Hastie et al. 2009).

El aprendizaje automático se utiliza en una variedad de aplicaciones, como reconocimiento de voz, reconocimiento facial, diagnóstico médico, recomendación de productos, análisis de datos y muchas más. Un modelo de aprendizaje automático puede recibir grandes cantidades de datos para mejorar su capacidad de predecir y adaptarse a diferentes situaciones (Martín, 2017).

Figura 1 Diferencia entre el estudio supervisado vs el estudio no supervisado



Fuente. Tomada de El blog de demóstenes (2021) Supervised and Unsupervised Learning

2.2.2. Deep Learning

Deep learning o aprendizaje profundo es un sub-campo de la Machine learning y se basa en Redes neuronales. Bien es cierto que las redes neuronales son algoritmos de machine learning, este tomó mucho peso en los últimos años y su campo empezó a ampliarse dando nacimiento al aprendizaje profundo y se encarga de resolver problemas muy complejos con grandes cantidades de información.

La utilización del Deep learning demanda gran capacidad de procesamiento, esto no se lograba años anteriores, pero en la actualidad contamos con mejores procesadores y componentes computacionales siendo posible su uso, alguno de sus aplicaciones son el procesamiento de lenguaje natural, asistencia de respuesta en vehículos inteligentes, visión artificial, entro muchos más.

2.2.3. Underfitting y Overfitting

En este apartado se va a tratar el tema del underfitting y overfitting, dos conceptos importantes en los modelos de ML. Es muy común que los científicos de datos, en especial aquellos que no tienen mucha experiencia, entrenen modelos que, a la hora de predecir datos nuevos, el modelo tiene un mal comportamiento. El underfitting y overfitting son peligros constantes y omnipresentes en el ML (Bashir et al., 2020). El problema que surge al momento de crear un modelo es que se busca que no “memorice” los datos de entrenamiento, si no que aprenda la relación real. En este escenario existe una técnica de la ciencia de datos bien establecida que permite desarrollar o escoger el modelo óptimo: la validación (Bashir et al., 2020).

Underfitting

El underfitting (subajuste) se refiere a un modelo que ha memorizado demasiado el conjunto de datos de entrenamiento y no puede generalizar o encajar bien en un conjunto de datos nuevos, esto ocurre cuando el modelo creado es muy simple (Guyon y Yao, 1999; Andrews, 2018; Gonzales, 2018a). El underfitting ocurre cuando un algoritmo aprende los detalles y el ruido en el conjunto de entrenamiento en la medida en que impacta negativamente el rendimiento del modelo en un nuevo conjunto de datos, por lo que el clasificador resultante tendrá una alta tasa de error tanto en el conjunto de entrenamiento como en nuevos datos, siendo incapaz de encontrar la relación entre las variables de entrada y salida con precisión (Lawrence et al., 1997; IBM, 2021).

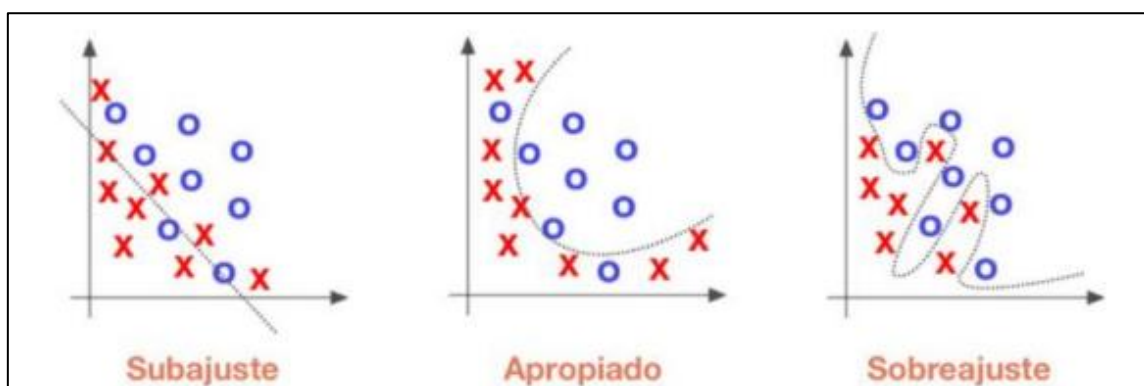
Overfitting

El overfitting (sobreajuste) se produce cuando un algoritmo reduce el error mediante la memorización de ejemplos de entrenamiento, con características ruidosas o irrelevantes, en lugar de aprender la verdadera relación general entre X e Y (Piotrowski y Napiorkowski, 2013; Krueger et al., 2017; Zhang et al., 2017).

En el caso del overfitting, el algoritmo está demasiado adaptado a los datos de entrenamiento que le son enseñados y será incapaz de generalizar nuevos datos de entrada que le son desconocidos. Cuando el modelo está sobreentrenado, el algoritmo solo considerará válidos los datos idénticos a los del conjunto de entrenamiento incluyendo el ruido (Chan et al., 2011; Panchal et al., 2011; Domingos, 2000). Esto significa que el algoritmo, en un momento determinado del período de entrenamiento, ya no mejora su capacidad para resolver el problema, sino que simplemente empieza a aprender alguna regularidad aleatoria

contenida en el conjunto de patrones de entrenamiento (Jabbar y Khan, 2015). En la figura 2 se puede observar un ejemplo gráfico sobre el underfittingy el overfitting.

Figura 2 Gráfico de underfitting, overfitting y un modelo



Fuente. de AprendeIA [Imagen], Gonzales (2018a),

<https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/>.

2.2.4. Matriz de confusión

Una matriz de confusión es una herramienta utilizada en el campo de la clasificación en machine learning para evaluar el rendimiento de un modelo en la predicción de clases. Esta matriz muestra la cantidad de veces que el modelo clasifica correctamente o incorrectamente las instancias en cada una de las clases (Kohavi y Provost, 1998).

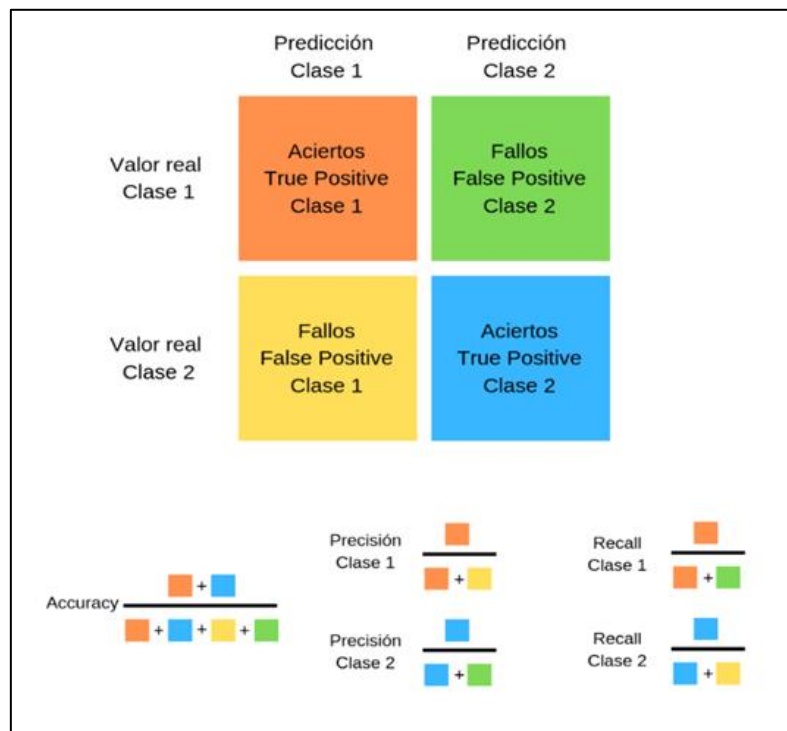
La matriz de confusión tiene cuatro componentes principales:

- **Verdaderos positivos (TP):** Representa el número de casos positivos que el modelo clasificó correctamente (Kohavi y Provost, 1998).
- **Falsos positivos (FP):** Indica el número de casos negativos que el modelo clasificó incorrectamente como positivos (Kohavi y Provost, 1998).
- **Falsos negativos (FN):** Muestra el número de casos positivos que el modelo clasificó incorrectamente como negativos (Kohavi y Provost, 1998).

- **Verdadero Negativo (TN):** Representa el número de casos negativos que el modelo ha clasificado correctamente (Kohavi y Provost, 1998).

La matriz de confusión generalmente se representa como una tabla con dos filas y dos columnas para problemas de clasificación binaria, pero para problemas de clasificación multiclase se puede extender a un mayor número de clases. Aquí está la representación básica de una matriz de confusión binaria:

Figura 3 *Aplicación de la precisión y recall*



Fuente. Aprende Machine Learning. (2019) Clasificación con datos desbalanceados.

A partir de estos valores se pueden calcular diversas métricas de evaluación del modelo, como precisión, sensibilidad, especificidad y tasa de falsos positivos, entre otras cifras. Estas métricas proporcionan más información sobre el rendimiento del modelo que solo la tasa de aciertos, especialmente cuando las clases no están equilibradas.

2.2.5. Curva ROC

La curva ROC (Característica Operativa de Receptor), es una herramienta gráfica que se utiliza para evaluar el rendimiento de un modelo de clasificación binaria en diferentes umbrales de decisión. La tasa de verdaderos positivos (sensibilidad) se compara con la tasa de falsos positivos (1 – especificidad) para diferentes umbrales (Cerde y Cifuentes, 2012). A continuación, se ofrece una explicación más detallada de los conceptos relacionados con la curva ROC:

- **Sensibilidad (tasa de verdaderos positivos):** Mide el porcentaje de casos positivos identificados correctamente por el modelo.

$$\text{Sensibilidad} = \text{TP} / (\text{TP} + \text{FN})$$

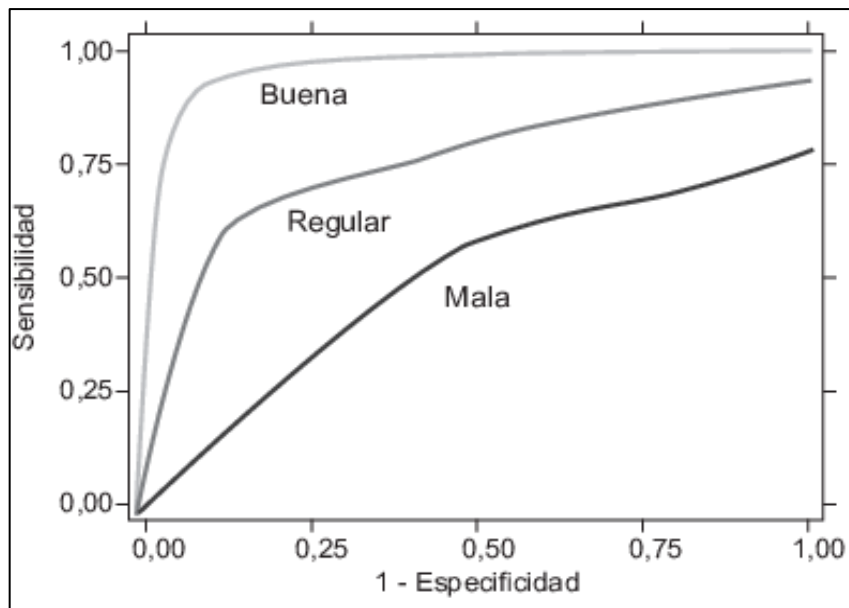
- **Especificidad (tasa de verdaderos negativos):** Mide el porcentaje de casos negativos identificados correctamente por el modelo.

$$\text{Especificidad} = \text{TN} / (\text{TN} + \text{FP})$$

- **Falso positivo (FP) y falso negativo (FN):** Representan errores del modelo en la clasificación de instancias.

Las curvas ROC se generan variando los umbrales de decisión del modelo y calculando la sensibilidad y especificidad para cada umbral. Estos valores se utilizan para trazar puntos en el espacio ROC y la curva resultante muestra cómo cambia el rendimiento del modelo a medida que se ajusta el umbral (Cerde y Cifuentes, 2012).

Figura 4 Gráfica explicativa de las diferentes posibilidades de la curva ROC



Fuente. Manterola, C. (2012) Esquema explicativo de distintas posibilidades de curvas ROC

2.2.6. Árbol de decisiones (Decisions tree)

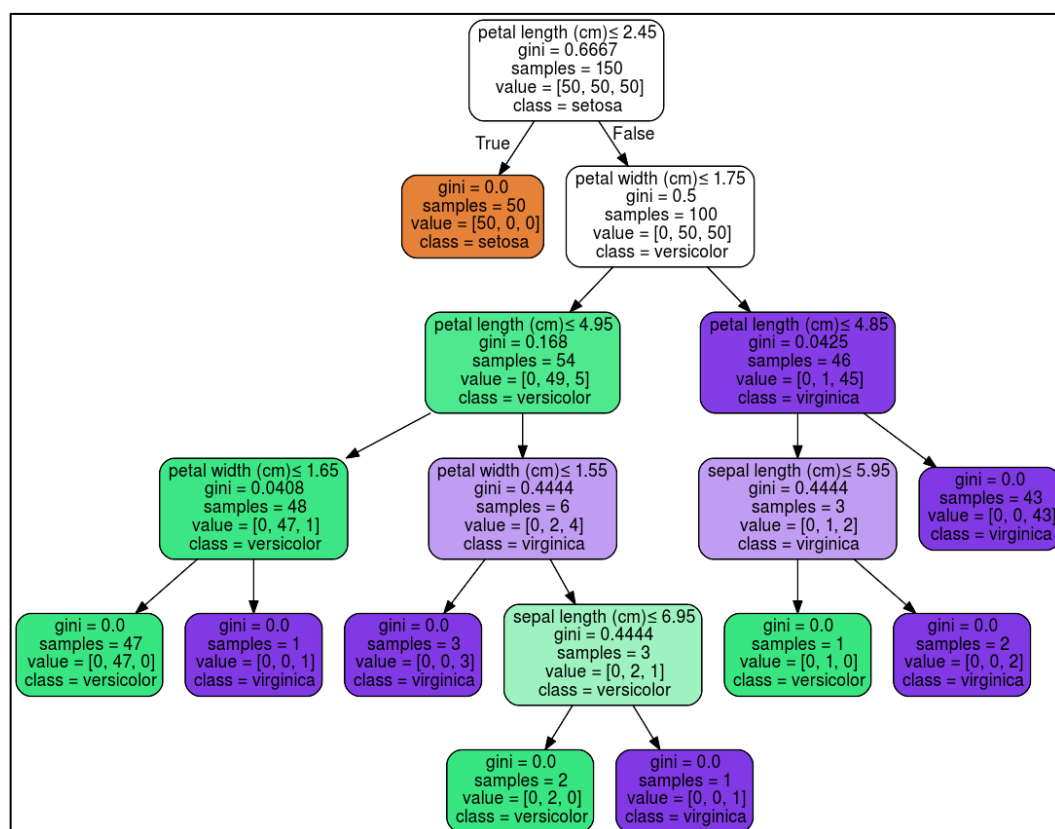
El árbol de decisiones es una herramienta del machine Learning dentro del aprendizaje supervisado. Es una representación visual de un conjunto de reglas de decisión que se aplican para predecir el valor de una variable objetivo en función de varias características. El cual está compuesto por nodos, aristas, nodos de decisión, nodos hoja, división del árbol, entropía y ganancia de información, profundidad del árbol, poda del árbol (Pruning), Reglas de decisión (Breiman et al. 1984).

El cual el árbol de decisiones no se expresa exactamente con una ecuación matemática como lo haría una función lineal, más bien esta expresión se representaría como: Si X es verdadero, entonces seguir por la rama izquierda, ahora si X es falso seguir por la rama derecha.

Esta herramienta nos ayuda a establecer la agrupación de información para clasificar a un ítem, asignando los ítems a cada grupo de una población de

esta manera haciendo predicciones de cada evento. Conllevando a una reducción de datos para la realización de modelos. Identificando las variables de un grupo con otro y así filtrando la información para obtener criterios cualitativos.

Figura 5 Gráfica detallada de un árbol de decisiones



Fuente. Esri. (2024). Diagrama de herramientas de GeoAI en ArcGIS

2.2.7. Bosques aleatorios (Random forest)

Breiman (2001) ideó inicialmente los bosques aleatorios (RF) como un conjunto de árboles de decisión cuya predicción depende de los valores de un vector aleatorio muestreado de forma independiente. Primero, el algoritmo toma muestras aleatoriamente del conjunto de datos original y crea un conjunto de datos "bootstrapped". Este proceso se llama "bagging". Luego, selecciona atributos aleatoriamente y evalúa los árboles de decisión en cada nodo calculando su impureza. Existe una variedad de métodos para medirla, como el índice Gini

o el criterio de entropía. El primero evalúa la disparidad en cada nodo, lo que significa que, a mayor índice de Gini, menor pureza. Está proporcionado por:

$$I_g(f) = \sum_{i=1}^m f_i(1 - f_i)$$

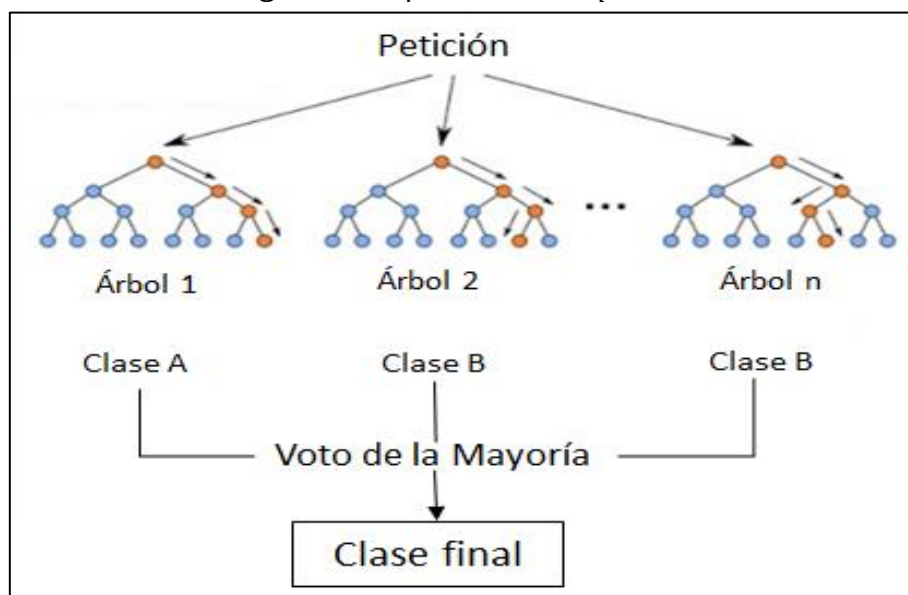
f_i = Probabilidad de la clase i en el nodo m .

Donde se calcula $f_i = n_j/n$

n_j = Numero de muestras que pertenece al grupo j

n =Número total de muestras del nodo.

Figura 6 Gráfica de los bosques aleatorio



Fuente. Medium (2020). Bosques aleatorios

2.2.8. Redes neuronales artificiales (Artificial neural networks)

Las redes neuronales artificiales (ANN) son un componente fundamental en el campo de la inteligencia artificial y el aprendizaje profundo. Están inspiradas en el funcionamiento del cerebro humano y están diseñadas para realizar tareas específicas mediante el procesamiento de información de manera similar a cómo lo hacen las neuronas en el sistema nervioso (Srivastava, 2014)

Es una herramienta de propósito general flexible y no lineal que se compone de las siguientes partes: capas de entrada, neuronas, capas ocultas, función de activación, función de pérdida, optimizador y tasa de aprendizaje. La forma matemática en que la red calcula las salidas de cada capa se describe en la siguiente ecuación (Srivastava, 2014).

$$y_i = fa \left(b_j + \sum_i W_{ji} X_i \right)$$

y_i = Valor de salida de las neuronas en la capa.

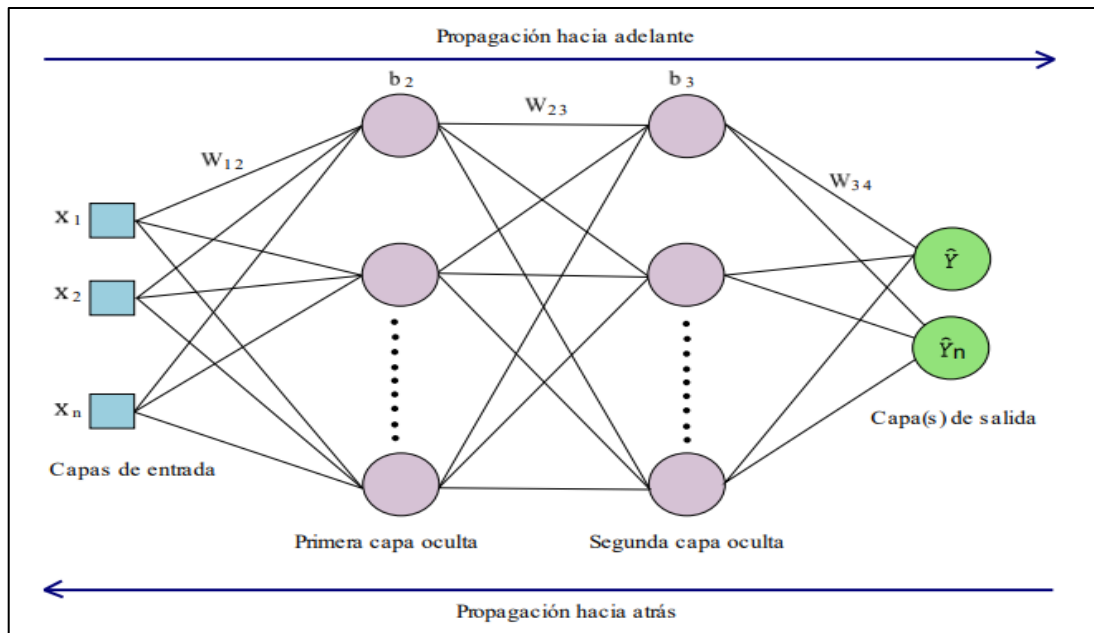
j, fa = Función de activación.

b_j = Parámetro independiente “bias”.

W_{ji} = Pesos de la red de la conexión de capas i y j .

X_i =Valor de entrada de las neuronas en la capa.

Figura 7 Estructura y elementos que conforman un modelo de las redes neuronales artificiales



Fuente. Mantilla Dulcey, A. G. (2023). Estructura y elementos que conforman un modelo de las redes neuronales artificiales.

2.2.9. Integración de datos geológicos

El éxito del pronóstico de mineralización depende de la correcta integración de datos geológicos, geoquímicos, estructurales y espectrales. La cartografía geológica, combinada con técnicas de Machine Learning, permite generar mapas predictivos que destacan zonas con alto potencial mineral (Gilberto Silva Romo et al., 2010). Esta metodología no solo optimiza la exploración, sino que también reduce los riesgos económicos y ambientales asociados.

2.2.10. Data Mining

El Data Mining es el conjunto de técnicas y tecnologías que permiten extraer información y encontrar patrones repetitivos de interés con conocimientos útiles y comprensibles desde grandes bases de datos de manera automática o semiautomática que expliquen el comportamiento de estos datos (Han et al., 2011, 2012; Bramer, 2007). Antes de adentrarse al machine learning, se deben definir algunos conceptos claves de la Data Mining.

Conjunto de datos (dataset)

El conjunto de datos son una parte integral de los sistemas de predicción. Son los datos históricos usados para entrenar un sistema o modelo de detección de patrones. Este compuesto por instancias, e instancias de factores, características o propiedades (Torralba y Efros, 2011; Gónzales, s.f). El conjunto de datos engloba a los conjuntos de entrenamiento, validación y prueba. Es importante que el conjunto de datos tenga un volumen suficiente como para generar resultados estadísticamente significativos, y a la vez, que sea representativo del conjunto de datos global (Santos, 2022).

Datos de entrenamiento

Los datos de entrenamiento son aquellos datos etiquetados usados para entrenar un modelo o algoritmo, con el fin de que aprenda de dichos conjuntos, ajuste los parámetros y aumente la precisión. Suele ser el conjunto más grande (Batista et al., 2004; Zhu et al., 2012). La calidad del modelo de machine learning va a ser directamente proporcional a la calidad de los datos. Es por ello por lo que las labores de preprocesamiento, limpieza y depuración consumen un gran porcentaje del tiempo de trabajo (Santos, 2022; Cracknell, 2014).

Datos de prueba

Los datos de prueba son los datos reservados para evaluar la generalización del modelo. Evalúa que tan bien funciona el modelo generado a partir de los datos de entrenamiento. En otras palabras, si las repuestas son acertadas o no para un caso nuevo (Howden, 1975; Mitchell, 1999; Bhavsar y Ganatra, 2012).

Datos de validación

Los datos de validación ayudan a evaluar o validar los hiperparámetros del modelo, previniendo el sobreajuste o subajuste (Breck et al., 2019).

Función objetivo

La función objetivo es aquella función que será optimizada dadas las limitaciones o restricciones determinadas y con variables que necesitan ser minimizadas o maximizadas mediante programación lineal o no lineal (Saaty y Gass, 1954; Gass y Saaty, 1955; Bezdek et al., 1981).

Hiperparámetros

Los Hiperparámetros son los valores de las configuraciones utilizadas durante el proceso de entrenamiento. Son utilizados para parametrizar el proceso de instanciación del modelo (Rozada, 2019; Wang y Gong, 2018). Estos valores

generalmente no se obtienen de los datos, sino que son indicados por el científico de datos. No se usan para modelar el modelo directamente, influyen en la capacidad y características de aprendizaje del modelo (Mackay, 1996). Ejemplos de hiperparámetros pueden ser el número de vecinos de KNearest Neighbors, la profundidad máxima de un árbol de decisión, etc. El valor óptimo de un hiperparámetro no puede ser conocido a priori para un problema dado. Para ello se utilizan valores genéricos, reglas genéricas, los valores que han funcionado anteriormente en problemas similares. La validación cruzada es una buena opción para buscar los hiperparámetro (Rodríguez, 2021).

Etiquetado de datos

El etiquetado de datos se denomina al proceso de identificación de las propiedades de los datos (documentos, texto, imágenes, números, etc.) y asignándole un nombre a dichas propiedades. Los datos pueden estar etiquetados y no etiquetados (Evans y Eysers, 2012; Cracknell, 2014). Se dice que un conjunto de datos está etiquetado cuando éstos tengan asociados una etiqueta, y que serán datos sin etiquetar cuando no lo tengan. Cabe destacar que un conjunto de datos puede ser “no etiquetado” por qué la etiqueta no existe, pero no lo haya sido aún añadida a los datos o porque no existe (InteractiveChaos, s.f; Bruening y Waterman, 2010).

2.2.11. Ruido

El ruido es cualquier anomalía, error aleatorio de varianza de una variable no medida, no deseada en los datos y, debido al ruido, la clase puede ser más difícil de aprender (Han et al, 2011; Alpaydin, 2020). Según Alpaydin (2020) hay varias interpretaciones del ruido:

- Puede haber imprecisión en el registro de los atributos de entrada, lo que puede desplazar los puntos de datos en el espacio de entrada.

- Puede haber errores en el etiquetado de los puntos de datos, que pueden reetiquetar los casos positivos como negativos y viceversa. Esto se denomina a veces ruido del profesor.
- Puede haber atributos adicionales, que no hemos tenido en cuenta, que afecten a la etiqueta de una instancia. Estos atributos pueden estar ocultos o latentes, es decir, pueden ser inobservables. El efecto de estos atributos descuidados se modela, pues como un componente aleatorio y se incluye en el "ruido".

2.2.12. Sesgo

En machine learning el término sesgo (bias en inglés) fue introducido por Mitchell (1980), el cual refiere al error persistente o sistemático que se espera que cometa el algoritmo de aprendizaje cuando se entrena con conjuntos de entrenamientos de tamaño m (Dietterich y Kong, 1995). En otras palabras, se puede definir como la diferencia entre la predicción esperada del modelo con respecto a los valores verdaderos observados. El sesgo habitualmente presenta una relación inversa con la complejidad de los modelos, a mayor complejidad del modelo utilizado es de esperar un menor sesgo (Rodríguez, 2019). Un sesgo elevado puede hacer que un algoritmo pase por alto las relaciones relevantes entre las características y los resultados objetivo (subajuste o underfitting) (Campesato, 2020). El sesgo en los sistemas de IA proviene de diversas fuentes, desde el algoritmo y los atributos de entrada elegidos hasta las correlaciones ocultas en los datos de entrenamiento (Roselli et al, 2019).

2.2.13. Varianza

Es la cantidad en que la estimación de la función objetivo cambiará si se utiliza diferentes datos de entrenamiento (Gonzales, 2018b). La varianza indica el grado de dispersión de los datos (Han et al, 2003). El error debido a la varianza es

la variabilidad de la predicción de un modelo para un punto de datos determinado (Campesato, 2020). La varianza puede ser baja apuntando pequeños cambios en la estimación de la función objetivo con cambios en el conjunto de datos de entrenamiento, Por el contrario, una alta varianza sugiere grandes cambios en la estimación de la función objetivo con cambios en el conjunto de datos de entrenamiento (Gonzales, 2018b).

2.2.14. Pronóstico de mineralización

El pronóstico de mineralización es una evaluación que se realiza para determinar la probabilidad de que un área específica contenga depósitos minerales económicamente explotables. Este proceso involucra una combinación de estudios geológicos, geoquímicos, geofísicos y, a menudo, de técnicas de modelado y simulación. Melgarejo, J. C., Proenza, J. A., Galí, S., & Llovet, Xavier. (2010). Técnicas de caracterización mineral y su aplicación en exploración y explotación minera.

2.2.15. Cartografía geológica

La cartografía ha sido una de las herramientas fundamentales de estudio e investigación desde los inicios de las ciencias de la tierra. Como el estudio geológico. Esto se debe principalmente a que resulta útil para visualizar gráficamente las relaciones espaciales y temporales de los objetos como rocas y estructuras tectónicas.

A partir de estas representaciones gráficas es posible resolver numerosos aspectos sobre el origen de los diferentes grupos petrológicos que forman la corteza terrestre, la evolución paleogeográfica de los continentes a diferentes escalas y la evolución de las características tectónicas de la Tierra.

Su aplicación a problemas de exploración ha llevado al desarrollo de diversas representaciones relacionadas con características asociadas a acuíferos, minerales y recursos petroleros. (Gilberto Silva Romo; Claudia Cristina Mendoza Rosales; Emiliano Campos Madrigal, 2010),

2.2.16. Geoquímica

La geoquímica es una rama de la ciencia que estudia cómo los elementos químicos se encuentran en minerales, menas, rocas, suelos, aguas y la atmósfera. Además, utilizando las características de sus átomos y iones para examinar cómo circulan estos elementos químicos en la naturaleza (Goldschmidt, 1954). La geoquímica de exploración es una rama de la geoquímica aplicada que utiliza elementos químicos, propiedades y distribución de la naturaleza para localizar y caracterizar depósitos minerales económicos. Para conceptualizar la distribución y circulación de elementos químicos en la naturaleza, es práctico tener un marco del ciclo geoquímico entre los diversos embalses y compartimentos terrestres. (Lasaga y Ohmoto, 2002; Eriksson y otros, 2004; Condie, 2011; Brimblecombe, 2014; Petsch, 2014).

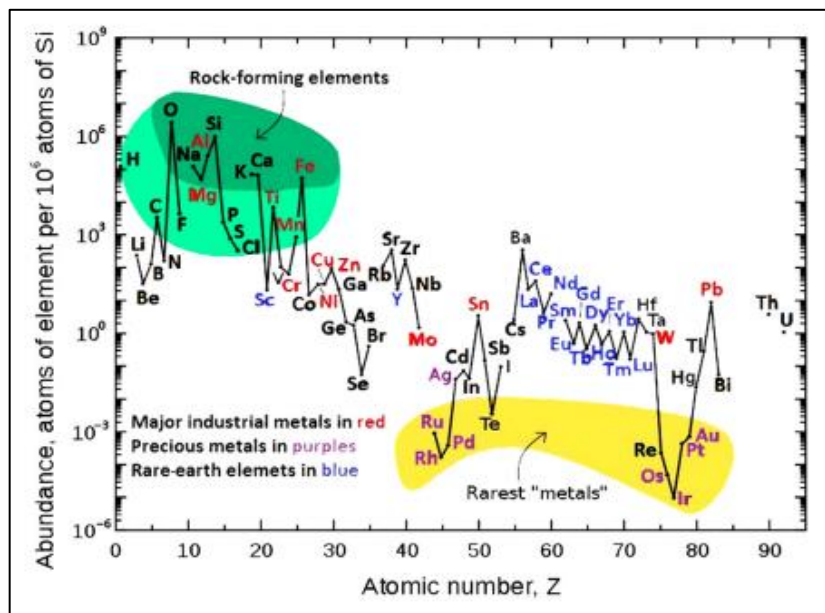
Los procesos internos y externos de la Tierra que involucran carbono, fósforo, azufre, agua y roca se conocen como ciclo geoquímico. Estos suceden en el interior de la Tierra. En el ciclo geoquímico, las formaciones minerales se ven significativamente afectadas por la litosfera (litosfera, corteza, manto y núcleo), la hidrosfera (hidrosfera, biosfera y suelos), la atmósfera (troposfera, estratosfera, mesosfera y termosfera) y el espacio exterior (meteoritos y asteroides). Por lo tanto, el estudio del ciclo geoquímico incluye la química de las rocas, la química del suelo, la química del agua, la química orgánica (o biogeoquímica) y la química de los gases. Podemos apreciar los procesos geoquímicos que conducen a la formación de

minerales teniendo en cuenta los métodos utilizados para la exploración geoquímica. (Lasaga y Ohmoto, 2002).

2.2.17. Clasificación geoquímica

Por el interés de la exploración minera, sólo se consideran las abundancias de elementos; Los elementos principales y traza de la corteza.

Figura 8 Abundancia relativa de los elementos químicos en el sistema solar



Fuente. Anders, E., Ebihara, M. (1982). Abundance of elements in the solar system.

a) Clasificación general

- **Elementos mayores:** Como su nombre indica, los elementos principales son aquellos que constituyen una gran parte de la Tierra o del sistema solar, generalmente más del 98,5% del peso total de los elementos en la Tierra o en el sistema solar. Según (Rudnick y Gao 2003). Los elementos principales se muestran en porcentaje de peso de óxido.
- **Elementos traza:** Los elementos traza representan solo el 0,1% de todos los elementos en la corteza terrestre o en el sistema solar (Turekiano & Wedepohl,

1961), a diferencia de los elementos principales. Ni, Cu, Zn, Pb, W, Ta y otros son los elementos trazas. Los litófilos, siderófilos, calcófilos, atmófilos son algunas de las diversas categorías de elementos trazas que también pertenecen a la Clasificación de Goldschmidt. Los elementos trazan también se pueden dividir en elementos del grupo del platino (PGE), metales y elementos de tierras raras (REE).

Tabla 1 Abundancia de elementos de la corteza continental superior

	UCC AVE	GS MED	WS1 MED	WS2 MED	WS3 MED	USA AVE	USA MED	China AVE	AP MED	TC MED
Majors total (wt.%)										
Al ₂ O ₃	14.5	13.5	15.1	13.4		13.6	9.7	12.3	10.5	8.1
CaO	3.6	1.9	2.0	2.1		3.4	1.4	3.1	1.2	0.5
Fe ₂ O ₃	5.6	5.4	50.	5.7		3.7	2.8	4.4	3.6	3.2
K ₂ O	2.8	1.6	1.7	1.7		1.8	1.8	2.3	1.9	2.0
MgO	2.5	1.0	1.5	0.8		1.5	1.0	1.5	1.0	0.5
MnO	0.10	0.11	0.07	0.13	0.07	0.07	0.05	0.08	0.08	0.04
Na ₂ O	3.27	0.87	1.35	0.67		1.62	1.10	1.52	0.79	0.30
P ₂ O ₅	0.15	0.18	0.17	0.18		0.10		0.15	0.18	0.06
SiO ₂	66.6	70.6	59.9	70.6		66.3			66.8	77.5
TiO ₂	0.64	0.77	0.67	0.83		0.48	0.42	0.68	0.62	0.58
Trace total (mg/kg)										
As	5	5	5	6	5	7	6		7	3
Ba	624	500	500	500	362	580	502		391	315
Ce	63	(50)	65	50	49	75			59	42
Co	17	8	10	8	7	9	7	13	9	8
Cr	92	200	80	70	42	54	50	58	64	48
Ga	18			20	12	17	15		12	10
Nb	12		12	10	12	11			13	9
Ni	47	40	20	50	18	19	15	25	21	15
Pb	17	10	17	35	25	19	17	25	21	13
Rb	84	60	65	150	50	67			75	51
Sr	320	300	240	250	147	240	148	186	102	68
Th	11	6	9	9	8	9			9	8
V	97	100	90	90	60	80	67	78	70	55
Y	21		20	40	12	25	23		28	21
Zn	67	50	70	70	62	60	52	69	62	31
Zr	193	300	230	400	300	230	188		263	304

Fuente. UCC: Upper continental crust, from Rudnick and Gao (2003); GS: Global soil, by Vinogradov (1954); WS1: World Soil, by Bowen (1979); WS2: Soil World, by Koljonen (1992); WS3: World Soil, by Kabata-Pendias (2001); USA: Contiguous soils of the USA, AVE by Shacklette and Boerngen (1984) and MED by Garrett (2009) or Gustavsson et al. (2001); China: Chinese soil, by Li and Wu (1999); and Ap, Tc are from De Caritat and Reimann (2012).

b) Clasificación de Goldschmidt

La clasificación geoquímica de Goldschmidt se basa en la afinidad química y la tendencia de ciertos elementos a comportarse de manera similar en procesos geológicos. Goldschmidt propuso una clasificación que agrupaba los elementos en función de su participación en la formación de minerales y compuestos específicos (Goldschmidt et al., 1937).

Tabla 2 *Clasificación de los elementos*

Lithophile	Siderophiles	Chalcophile	Atmophile
Li, Na, K, Rb, Cs	Fe ^a , Co ^a , Ni ^a	(Cu), Ag ^b	(H), N, (O)
He, Ne, Ar, Kr, Xe	Ru, Rh, Pd, Zn, Cd, Hg ^c ,	Be, Mg, Ca, Sr, Ba	
B, Al, Sc, Y, REE	Os, Ir, Pt	Ga, In, Tl	
Si, Ti, Zr, Hf, Th	Au, Re ^d , Mo ^d	(Ge), (Sn), Pb	
P, V, Nb, Ta	Ge ^a , Sn ^a , W ^e	(As), (Sb), Bi	
O, Cr, U	C ^e , Cu ^a , Ga ^a	S, Se, Te	
H, F, Cl, Br, I	Ge ^a , As ^c , Sb ^d	(Fe), Mo, (Os)	
(Fe), Mn, (Zn), (Ga)		(Ru), (Rh), (Pd)	

Fuente. Hollabaugh, C. L. (2007). Modification of Goldschmidt's geochemical classification of elements to include arsenic, mercury and lead as biophilic elements. Develop. Reign. Science. 931; Makishima, A., 2017. Origin of elements. In: Origins of the Earth, the Moon and Life, pp. 124.

2.3. Definición de términos básicos

2.3.1. Muestra

Una muestra es un subconjunto representativo de una población más grande. Se selecciona para realizar inferencias sobre la población en su conjunto. La calidad de la muestra es crucial para que las conclusiones extraídas sean válidas y aplicables a la población completa (Tamara Otzen & Carlos Manterola, 2017).

2.3.2. Muestreo geológico

El muestreo geológico es una técnica utilizada en geología para recopilar información sobre las características físicas, químicas y mineralógicas de las rocas y su entorno en la Tierra. Este proceso es esencial para comprender la historia geológica, la composición mineral y la estructura de una región específica. Basándose a aspectos como la ubicación, tipo de muestra (Sedimentos, suelos, minerales, y otros materiales geológicos), herramientas, documentación y análisis de laboratorio (Geología viva, 2021).

2.3.3. Cuerpo mineralizado

Un cuerpo mineralizado se refiere a una concentración de minerales de interés económico que se encuentra en la corteza terrestre. Estos minerales pueden incluir metales preciosos como oro y plata, metales base como cobre y zinc, y minerales industriales como fosfatos y carbonatos. La mineralización se produce a través de procesos geológicos como la actividad hidrotermal, la sedimentación y la metamorfosis, entre otros. Un cuerpo mineralizado puede ser una veta, un depósito, una masa o cualquier otra formación geológica que contenga una cantidad suficiente de minerales para justificar su extracción.

2.3.4. Yacimiento mineral

Un yacimiento mineral es una concentración natural de uno o más minerales en la corteza terrestre que tiene valor económico y puede ser explotado con fines industriales. Estos yacimientos pueden contener una variedad de minerales, metales y compuestos químicos que son valiosos para la industria. Está formado por procesos geológicos que puede incluir la actividad magmática, sedimentario o metamórfica (Pedro Hugo, 2003).

2.3.5. Ganga

Se refiere a los minerales o materiales no deseados y sin valor económico que acompañan a los minerales valiosos en un depósito. Estos minerales son indeseados porque no contienen los elementos o compuestos de interés económico y, por lo tanto, no son aprovechables desde el punto de vista industrial. La presencia de ganga puede complicar la extracción y el procesamiento de minerales valiosos, ya que estos deben separarse de la ganga para obtener el producto final deseado. Este proceso puede implicar operaciones físicas o químicas para separar los minerales útiles de los materiales no deseados. (Agencia Nacional de Minería, 2015).

2.3.6. Mena

Se refiere a la porción de un depósito de minerales que tiene un contenido lo suficientemente alto de sustancias valiosas como para ser económicamente rentable su extracción y procesamiento. En otras palabras, la mena es el material mineral del cual se extraen los metales u otros elementos útiles. (Agencia Nacional de Minería, 2015).

2.3.7. Estructuras geológicas

Las estructuras geológicas son características tridimensionales en la corteza terrestre que resultan de procesos geológicos. Estas estructuras proporcionan información valiosa sobre la historia geológica y las fuerzas que han actuado en una región. Por ejemplo, pliegues, fallas, diaclasas, rift, entre otras estructuras. Estas estructuras geológicas son esenciales para comprender la evolución de la Tierra y son clave en la identificación y explotación de recursos naturales, como yacimientos minerales y fuentes de energía (Tarbuck, E. J., & Lutgens, F. K. 2005).

2.4. Formulación de hipótesis

2.4.1. Hipótesis general

La aplicación de algoritmos de Machine Learning permitirá predecir modelos de mineralización en la concesión minera Gus del Aire, integrando datos geológicos, geoquímicos, estructurales y espectrales, lo que facilitará la identificación de zonas potenciales para la exploración minera.

2.4.2. Hipótesis específicas

- Relación entre la Matriz de Confusión y la Eficiencia Predictiva, Si se construyen matrices de confusión para evaluar los algoritmos de Machine Learning, entonces se podrá determinar la precisión y capacidad predictiva de los modelos, especialmente en relación con datos geológicos como concentraciones geoquímicas, proximidad a fallas y características litológicas.
- Impacto de la Curva ROC en la Validación de Algoritmos Predictivos, Si se generan y comparan curvas ROC, entonces será posible identificar el modelo con mayor sensibilidad y especificidad, optimizando su capacidad para pronosticar zonas de mineralización favorable basándose en datos geológicos y espectrales recolectados.
- Efecto del Machine Learning en la Eficiencia de la Exploración Minera, Si se integran datos geológicos y estructurales en algoritmos de Machine Learning, entonces se reducirá significativamente el tiempo de análisis, mejorando la toma de decisiones y optimizando la eficiencia en la exploración minera.

2.5. Identificación de variables

2.5.1. Variables independientes

Machine learning (Aprendizaje automático).

2.5.2. Variables dependientes

Pronóstico de mineralización.

2.5.3. Variables intervinientes

Geoquímica, geología, tectónica, imágenes satelitales de alta resolución.

2.6. Definición operacional de variables e indicadores

La definición operativa de estas variables define las dimensiones necesarias en el área de investigación, y los indicadores de proceso son:

Tabla 3 Operacionalización de variables

Variable	Tipo de variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores
Machine learning (Aprendizaje automático).	Independiente	Es una rama de la inteligencia artificial (IA) que implica el desarrollo de algoritmos y modelos que permiten a las computadoras aprender patrones y realizar tareas sin estar programadas explícitamente (Goodfellow et al., 2016).	El Machine Learning se basa en aplicar diferentes tipos de modelos y/o algoritmos, desde la definición de problemas, recolección, preparación y división de datos, selección de algoritmo, entrenamiento y la evaluación del modelo asegurando que el modelo sea preciso y robusto.	Eficiencia de algoritmos y modelos	Rendimiento de la matriz de confusión Área de la curva ROC
				Resolución espacial	Tamaño de pixeles
Pronóstico de mineralización.	Dependiente	El pronóstico de mineralización es una evaluación que se realiza para determinar la probabilidad de que un área específica contenga depósitos minerales económicamente explotables. Este proceso involucra una combinación de estudios geológicos, geoquímicos, geofísicos y, a menudo, de técnicas de modelado y simulación. Melgarejo, J. C., Proenza, J. A., Galí, S., & Llovet, Xavier. (2010). Técnicas de caracterización mineral y su aplicación en exploración y explotación minera.	El pronóstico de mineralización se proyecta en encontrar depósitos minerales en áreas determinadas a través de pasos y procedimientos a través de la recolección de datos, análisis de datos, modelado y simulación, y evaluación de potencial.	Ambiente geológico	Tipo de roca
				Anomalías geoquímicas	Background y Threshold
				Controles estructurales	Complejidad estructural
				Resolución espacial	Tamaño de pixeles

Fuente: Elaboración propia.

CAPITULO III

METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación

De acuerdo con la aplicabilidad del estudio que se llevará a cabo la investigación será de tipo de investigación aplicada, no experimental y tecnológica puesto que se enfoca en resolver un problema práctico, considerando que no se manipulan las variables y que se aplica herramienta de ML para un caso geológico como tecnología. Campos, G. y Sosa, V. (2011). Ya que recogerá información primaria y secundaria para poder generar modelos predictivos utilizando el diagrama desarrollado.

3.2. Nivel de investigación

La investigación corresponde a nivel descriptivo-correlacional y predictivo, pues describe las características geológicas y establece la relación entre variables de ML y predictivo ya que busca predecir zonas favorables de mineralización usando modelos ya establecidos. Carrasco, D. (2006). Metodología de la investigación científica (pp. 41 – 42).

3.2.1. Descriptivo

Este nivel se enfoca en identificar y caracterizar las variables relacionadas con la mineralización en la zona de estudio. Una descripción clara de las variables y su distribución espacial, apoyada con mapas geológicos, tablas de datos entre otros mapas.

3.2.2. Correlacional

Este nivel busca establecer relaciones entre las variables geológicas, geoquímicas y estructurales para determinar cómo interactúan y contribuyen a la búsqueda mineralización. Establecimiento de relaciones directas o indirectas entre variables clave y la identificación de variables predictivas relevantes para los modelos.

3.2.3. Predictivo

En este nivel, se desarrollan modelos capaces de predecir la ubicación de depósitos mineralizados utilizando las relaciones identificadas previamente. A través de modelos robustos con alta precisión para predecir mineralización en la zona de estudio mediante mapas predictivos que guíen futuras investigaciones y exploraciones en el área.

3.3. Métodos de investigación

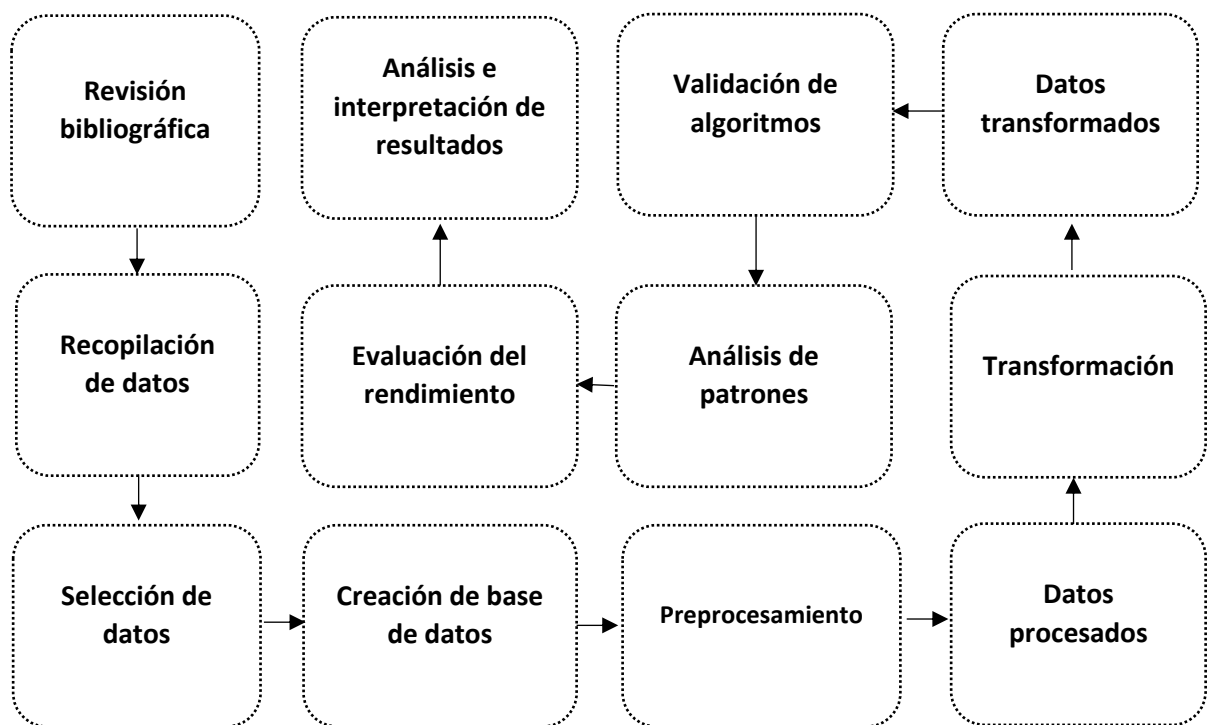
Se tendrá una metodología del tipo deductivo ya que se recopilará información y posteriormente buscará probar la hipótesis mediante la recolección y análisis de datos llevando esto a la validación y verificación de cada modelo.

De igual forma se aplicará el método cuantitativo se adapta perfectamente al problema, ya que se basa en datos medibles y herramientas analíticas que puedan permitir predicciones confiables, como: datos numéricos, análisis de datos, modelo predictivo basado en algoritmos y los resultados medibles.

3.4. Diseño de la investigación

Para cumplir los objetivos planteados es primordial definir el esquema que se debe de seguir para obtener resultados deseados. Se presenta de manera general un esquema de la metodología usada para el desarrollo de esta investigación.

Figura 9 Esquema de metodología de trabajo



Fuente: Elaboración propia.

La descripción de este esquema es lo siguiente:

- Revisión bibliográfica (Descrita en el cap. III).
- Recopilación de datos, ya sea información primaria y secundaria.
- Selección de datos, el cual se filtrará información verificando la importancia para el futuro procesamiento.
- Creación de base de datos, se agrupará la información seleccionada para correlación uno con otro.

- Preprocesamiento, se realizará un primer procesamiento de la data cruda como ciertas correcciones de información.
- Datos procesados, esta información muchas veces se llaman minería de datos el cual se podrá identificar los patrones entre sí.
- Transformación, es donde ya validada los patrones de la información estos se llevan a un mismo factor en este caso será en imágenes ráster y shape a través de una relación espacial.
- Datos transformados, estos se mantendrán dentro de una extensión el cual se podría integrar a un tipo de algoritmo del Machine Learning.
- Validación de algoritmos, se plantearán los diferentes algoritmos del ML, ya sea árbol de decisiones, bosques aleatorios o redes neuronales artificiales.
- Análisis de patrones, se corrobora el análisis a través de la matriz de confusión o el ROC.
- Evaluación del rendimiento, Dentro de la evaluación del rendimiento veremos cuál de los algoritmos tiene una mejor eficacia al predecir la mineralización en la zona de estudio.
- Análisis e interpretación de resultados, teniendo el mejor resultado de los algoritmos aplicados, estos se correlacionarán con la información de campo validando así la predicción de mineralización.

3.5. Población y muestra

3.5.1. Población

La población está representada por la información geológica regional como geología, geoquímica, metalogenia, dominio estructural, tectónica, geomorfología, ocurrencia mineral y proyectos mineros dentro y cercanos a la concesión minera Gus del Aire representada por 1.24 millones de píxeles.

3.5.2. Muestra

La muestra estará representada por 12416 pixeles donde los datos están entrenados y evaluados conformadas por la proximidad de cuerpos intrusivos, concentraciones geoquímicas, proximidad de fallas y las anomalías espectrales dentro y cerca a la concesión.

3.6. Técnicas e instrumentos de recolección de datos

3.6.1. Recolección de datos

Técnicas de recolección de datos.

- **La observación.** La técnica de la observación se utilizará para recopilar características litológicas, estratigráficas y estructurales, entre otras. Como inspección de afloramientos, identificación de tipos de roca y estructuras tectónicas y características importantes de las notas de campo.
- **Análisis documental.** Revisión de fuentes secundarias, mapas geológicos, registros tectónicos, entre otros de instituciones como INGEMMET y GEOCATMIN.
- **Muestreo Geológico.** Obtención de muestras geoquímicas en la zona de estudio.
- **Procesamiento de datos satelitales.** Uso de imágenes satelitales de alta resolución para identificar anomalías espectrales y áreas potenciales de interés.

Instrumentos de recolección de datos.

- **Instrumentos de campo.**

- Planos topográficos. Para ubicar coordenadas precisas en el área de trabajo. Además, se proporcionarán descripciones breves utilizando símbolos y abreviaturas específicos de la geología.
- GPS. Para georreferenciar puntos específicos.
- Brújula geológica. Para medir las orientaciones de los afloramientos y/o estructuras geológicas.

- **Software y herramientas digitales.**

- Microsoft Excel y Minitab: Procesamiento estadístico de data geoquímica. Resultados la correlación de Spearman y Pearson.
- Anaconda Navigator y Jupyter Notebook. Uso para programación y modelo predictivo mediante modelos Machine Learning.
- Arcgis: Para creación de mapas y análisis espacial.
- Python. Para el diseño y aprendizaje de modelos predictivos, análisis de matriz de confusión y generación de curvas ROC.

- **Procesamiento General.**

- Recolección de información primaria y secundaria. Recopilación de datos del área de estudio mediante observación directa y revisión documental.
- Creación de base de datos. Clasificación y organización de la información para análisis digital.
- Procesamiento digital. Limpieza, transformación y validación de datos para garantizar la calidad.

- Generación de mapas y modelos predictivos. Uso de herramientas GIS y algoritmos de Machine Learning para identificar el área de interés.
- Validación de datos. Uso de matrices de confusión y curvas ROC para verificar el desempeño de algoritmos predictivos.

3.7. Técnicas de procesamiento y análisis de datos

3.7.1. Técnicas de procesamiento

Procesamiento manual

- Consiste en el tratamiento directo de los datos recolectados en campo para su preparación y análisis.
- Elaboración de planos geológicos locales a través de mapeos de campo.
- Organización de muestreos.

Procesamiento electrónico y/o digital

- Uso de herramientas y softwares especializados para manejar grandes volúmenes de datos geológicos y geoquímicos.
- Uso de Microsoft Excel y Minitab para el tratamiento inicial de datos geoquímicos. Ver resultados en el anexo I.
- Modelado con Machine Learning a través de los diferentes algoritmos utilizando Python a través del Anaconda Navigator y el Jupyter Notebook. Donde se implementó algoritmos como Bosques Aleatorios (RF) y Redes Neuronales Artificiales (ANN).
- Generación de mapas temáticos y predicción mediante Arcgis.
- Creación de matrices de confusión y de la curva ROC para evaluar el desempeño de los modelos.

Preprocesamiento de datos

- Preparación de los datos crudos para su uso en modelos predictivos y análisis estadísticos. A través de limpieza de datos, eliminando valores atípicos y corrección de errores en la base de datos, estandarización y normalización, transformación de datos a un formato homogéneo para garantizar la compatibilidad entre capas de datos y la conversión a formatos digitales. Digitalización de datos de campo a formatos ráster y vectoriales.

3.7.2. Análisis de datos

Análisis estadístico

- Identificar patrones y correlaciones entre variables. Mediante el cálculo de estadística descriptiva (Media, mediana y desviación estándar).
- Análisis de correlación utilizando la matriz de Spearman y/o Pearson.

Análisis de patrones geoespaciales

- Identificar la distribución espacial en la mineralización, a través de la superposición de capas ráster y vectoriales en Arcgis y su identificación de anomalías geológicas y geoquímicas.

Modelado predictivo

- Predecir áreas de mineralización favorable mediante Machine Learning, con la implementación de algoritmos como:
- Bosques aleatorios (RF) sirve para manejar datos no lineales y evaluar la importancia de las variables predictivas de entrada.

- Redes neuronales artificiales (ANN) para establecer relaciones complejas entre variables e identificar patrones ocultos.
- La validación de modelos se realiza con la matriz de confusión que sirve para evaluar el rendimiento y precisión, también la curva ROC que ayuda analizar el equilibrio entre sensibilidad y especificidad.
- Evaluación del rendimiento de los modelos
- Tiene como objetivo seleccionar el modelo con mejor desempeño para la predicción de mineralización, mediante la comparación métrica como: Precisión, sensibilidad, especificidad y área bajo la curva (AUC) ROC.

3.8. Tratamiento estadístico

Para la aplicación de la estadística descriptiva o tratamiento estadístico de los datos y/o píxeles de los mapas a predecir se tuvo las variables como: Au, Ag, Zn, Pb, Cu, Mo, Geología local, Fallas, anomalías espectrales y el área de interés, donde a través del Anaconda Navigator a través del Notebook Jupyter obtener su tratamiento de datos. Ver figura 10.

Figura 10 Tratamiento estadístico del estudio.

In [19]: df_filter = df print(df_filter.describe())						
	Au	Ag	Pb	Zn	Cu	\
count	12416.000000	12416.000000	12416.000000	12416.000000	12416.000000	
mean	-0.538376	-1.502537	0.525743	0.049148	0.380924	
std	0.532087	1.756727	0.654163	0.092616	0.478901	
min	-1.576033	-4.493131	-0.877148	-0.185670	-0.303617	
25%	-0.960875	-2.952560	0.000000	-0.005102	0.000000	
50%	-0.493862	-1.086818	0.503370	0.044981	0.350923	
75%	-0.015235	0.000000	0.970814	0.109064	0.723858	
max	0.471179	1.332183	2.077592	0.315077	1.427978	
	Mo	Geologia	Fallas	Anomalias espectrales	\	
count	12416.000000	12416.000000	12416.000000	12416.000000		
mean	0.588225	2.569353	67.030631	42.452667		
std	0.738604	16.673995	131.316497	122.725045		
min	-0.369070	0.000000	0.000000	0.000000		
25%	0.000000	0.000000	0.000000	0.000000		
50%	0.426311	0.000000	0.000000	0.000000		
75%	1.054217	0.000000	72.691292	0.000000		
max	2.332602	110.767677	487.627991	397.205963		
	AOI					
count	1.241600e+04					
mean	-inf					
std	inf					
min	-1.797693e+308					
25%	-1.797693e+308					
50%	-1.797693e+308					
75%	-1.797693e+308					
max	1.000000e+00					

Fuente: Elaboración propia.

3.9. Orientación ética, filosófica y epistémica

Para desarrollar el trabajo de investigación actual, se seguirán y respetarán todas las normas éticas del campo de la investigación científica. Además, todos los elementos que se presentarán, como cuadros, gráficos, planos geológicos, resultados y otros datos, serán datos reales obtenidos de la investigación. Además, las definiciones y conceptos teóricos se obtendrán de su fuente principal.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del área de trabajo

4.1.1. Ubicación

La concesión minera “Gus del Aire” se ubica al este del Perú al Sur Este de la ciudad de Huánuco, entre el límite de los departamentos de Huánuco y Pasco, políticamente pertenece al distrito de Panao, provincia de Pachitea del departamento de Huánuco. Ver plano de ubicación en el anexo I. Se encuentra en un rango de altitud entre los 3000 a 4000 m.s.n.m. Se ubica aproximadamente en las coordenadas siguientes:

Tabla 4 *Coordenadas WGS84 de la concesión minera Gus del Aire*

VERTICE	NORTE	ESTE
1	8,882,000	416,000
2	8,879,000	416,000
3	8,879,000	413,000
4	8,880,000	413,000
5	8,880,000	415,000
6	8,882,000	415,000

Fuente: Elaboración propia.

4.1.2. Accesibilidad

La vía de acceso principal es por la carretera central tramo Lima-Huánuco de 358 km (asfaltado), Huánuco-Churubamba de 21.6 km (asfaltada), Churubamba-Panao de 42.3 km (Trocha carrozable), y Panao-área de estudio de 175 km (Trocha carrozable) haciendo un total de 596.9 km.

Tabla 5 *Vías de acceso al proyecto de investigación “Gus del Aire”*

Acceso a la zona de estudio			
RUTA	km.	Vía	Horas (vehículo)
Lima-Huánuco	358	Asfaltada	7 h 41 min
Huánuco-Churabamba	21.6	Asfaltada	34 min
Churubamba-Panao	42.3	Trocha carrozable	1 h 19 min
Panao-área de estudio	175	Trocha carrozable	4 h 00 min
Total	596.9	-	13 h 34 min

Nota. Única vía de acceso al proyecto de investigación.

4.1.3. Clima y vegetación

El clima en esta zona se caracteriza por dos estaciones distintas: noviembre-marzo con fuertes lluvias y abril-octubre sin lluvia. La precipitación media es de 650 mm. año y la temperatura promedio es de 17°C.

Ahora de acuerdo a la clasificación climática del Perú elaborado por SENAMHI, el clima en el área del proyecto es frío y templado, con una humedad relativa anual de 84.0% y una temperatura mínima de 4.1°C a una máxima de 19°C, con mayor variación en la estación seca y más estable en la estación lluviosa. La temperatura está influenciada por la proximidad de la selva.

La flora está limitada por plantas como aliso lambran o rambrash, el maíz, hichu, paja, ayrampo, acelga, verbena, taya entre otras.

La fauna está conformada por pocas especies como zorzal gris llamado también chihuaco o yukish.

4.1.4. Marco geológico

Geología Regional

La geología regional está representada por rocas volcánicas y metamórficas, las rocas más antiguas están representadas por rocas metamórficas compuesta por esquistos neoproterozoicos del complejo marañón. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996).

El paleozoico contiene las rocas volcánicas que se evidencia en la parte superior con evidencias de cizallamiento, con pocas deformaciones donde se encuentra el granito de Paucartambo. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996).

El Neógeno producto de secuencias piroclásticas se da la formación Larontache que se encuentra consolidado por brechas angulosas de origen andesítico. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996). Ver plano en el anexo I.

Geología local

Las rocas predominantes en la zona de estudio corresponden a metamórficas del Neoproterozoico, suprayaciendo se tiene a los volcánicos del Pérmico-Paleozoico, suprayaciendo a estas se encuentran sedimentarios con piroclásticos del Neogeno-Cenozoico. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996). Geología de los cuadrángulos de Aguaytía, Panao y Pozuzo. Coronando la secuencia depósitos glaciares. Ver plano geológico local en el anexo I.

Estratigrafía

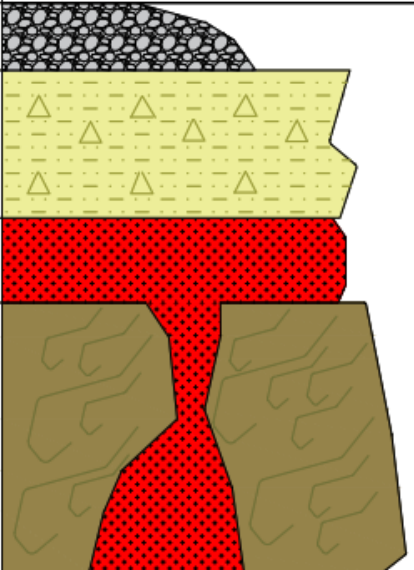
Se describe a continuación las siguientes formaciones rocosas desde la más antigua hasta la más reciente: Ver figura 11.

- **Complejo Basalto Marañón (NP).** Se presenta rocas metamórficas en el sur de Huánuco de igual manera entre el límite entre la región de Huánuco y Pasco. Esta secuencia se debe al basamento más antiguo de la zona que pertenece al Neoproterozoico.. Compuesta por esquistos y gneis, donde los esquistos se encuentran más al NW de la zona de estudio, presentan características de color verdosos a color gris oscuro, evidencia vetillas de cuarzo. Ahora los Gneis se presentan en franjas largas quien se caracteriza por presentar micas-cuarzo-plagioclasas y tiene formas lenticulares. El cual tiene una edad entre 600 m.a. y 50 m.a. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996). Geología de los cuadrángulos de Aguaytía, Pano y Pozuzo.
- **Granito Paucartambo (PET-gr).** Representada por rocas plutónicas de composición graníticas ácidas teniendo una granulometría de medio a fino con un diaclasamiento moderado, que contiene feldespático potásico, cuarzo, plagioclasas y poca presencia de biotita y hornblenda. En algunas zonas del estudio se presenta poca presencia de granodioritas. El cual se estima entre 259 m.a. y 200 m.a. De la Cruz, J.; Valencia, M. & Boulanger, E. (1996). Geología de los cuadrángulos de Aguaytía, Pano y Pozuzo.
- **Formación Lantorache (N-I).** Se presenta rocas piroclásticas y sedimentos tobáceos que se extiende dentro de la zona de estudio. Compuesta por una secuencia de conglomerados de matriz arenosa semiconsolidada con intercalaciones de sedimentos tobáceos de composición riolítica poca presencia de brechas y aglomerados con fragmentos angulosos y redondeados con composición andesítica. Esta formación descansa sobre rocas intrusivas del Paleozoico con un grosor de 200 m. El cual tiene entre

5.33 m.a. y 2.58 m.n., que pertenece Neogeno-Cenozoico. De la Cruz, J.; Valencia, M. & Boulangger, E. (1996). Geología de los cuadrángulos de Aguaytía, Panao y Pozuzo.

- **Deposito glaciario (Q-gl).** Está constituida por brechas subangulosas producto de la deglaciación que fueron erosionadas y trasladadas que se encuentra en la parte central de la zona de estudio. De la Cruz, J.; Valencia, M. & Boulangger, E. (1996). Geología de los cuadrángulos de Aguaytía, Panao y Pozuzo.

Figura 11 Columna estratigráfica de la geología local

ERATEMA	SISTEMA	SERIE	UNIDADES ESTRATIGRAFICAS	COLUMNA ESTRATIGRAFICA	ESPESOR (mts)	DESCRIPCIÓN
CENOZOICO	CUATERNARIO	Pleistoceno	Depositos glaciares			Representada por grava arenosa y clastos subredondeados.
	NEOGENO	Mioceno	Formación Lantorache		200	Representada piroclastos intercalados con conglomerados arenosos de igual manera con areniscas limosas y limoarcillitas.
PALEOZOICO	PERMIICO	Superior	Granito de Paucartambo		300	Representada rocas graníticas y granodioritas de composición acida, con presencia de un diaclasamiento moderado a intenso.
NEOPROTEROZOICO			Complejo basal del Marañón			Representada por esquistos y gneis de color verdoso a color gris oscuro con evidencias de vetillas de cuarzo.

Fuente: Elaboración propia.

Geología estructural

Se presenta una asociación de fracturas donde se tiene ocurrencias de Au, Pb y Zn en una serie de vetas y vetillas posiblemente de una intrusión calcoalcalino del granito de Paucartambo con una desmovilización de las rocas encajonantes. Ver plano estructural en el anexo I.

- **Fracturamiento.** La zona se haya fuertemente fracturada debido al tectonismo, estas fracturas siguen un patrón estructural derivado de la dirección de los esfuerzos el cual se presentan en sistemas de fracturamiento con dirección de NE a SW y Este a Oeste.
- **Fallas.** En la zona de estudio predomina dos sistemas de fallas uno con sentido NE - SW y de E - W:

Sistema de fallamiento NW-SE (longitudinal); son fallas premineralización de desplazamiento tipo normal originando ensanchamiento a la veta y otras desplazando las rocas encajonantes.

Sistema de fallamiento E – W (diagonal) de rumbo E a W y buzamiento alto de 80° el cual se presenta en la parte norte del estudio.

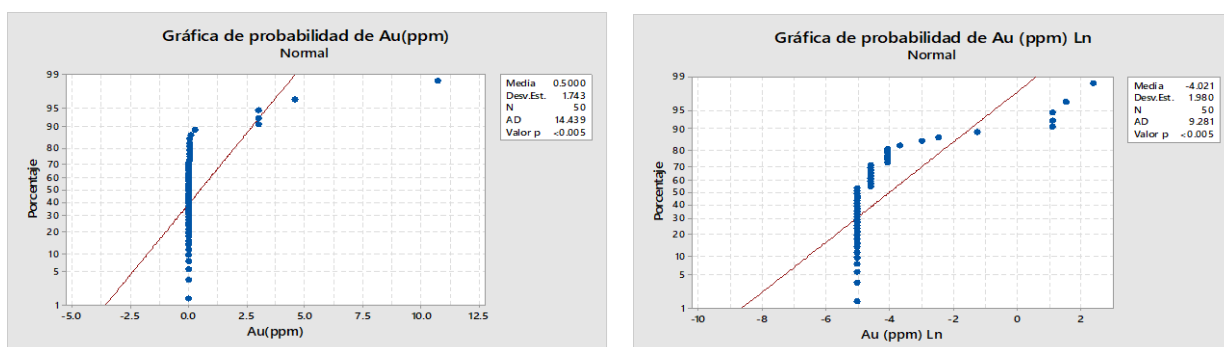
Análisis geoquímico

En la zona de estudio se realizó una campaña de exploración donde se obtuvo 50 muestras, donde a través de un análisis estadístico de datos geoquímicos se realizó una estadística multivariable con el objetivo de determinar zonas anómalas de algún posible depósito.

- a. **Distribución de normalidad.** Los datos obtenidos en el muestreo al ser tratado a través del análisis estadístico multivariado se tienen que tiene altos valores atípicos y de igual manera al hacerle una prueba de normalidad estas no tienen una distribución normal por el cual se deberá obtener una distribución normal el cual aplicaremos logaritmo natural (Ln) para obtener una distribución normal y poder realizar el análisis multivariado, de igual forma calcular el Background y el Treshhold. El cual fue desarrollado con el software Mintab 18.

- **Au:** en la figura N°12 se observa que no se tiene una buena distribución normal, aplicando el logaritmo natural se puede observar una mejor distribución donde el índice de Anderson es 9.281.

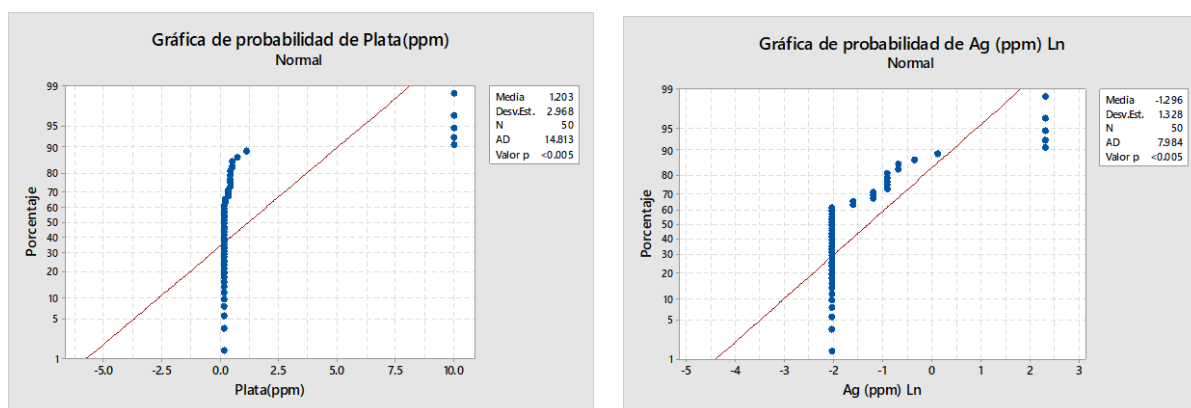
Figura 12 Relación entre la distribución no normalizada y normalizada de Au



Fuente: Elaboración propia.

- **Ag:** en la figura N°13 se observa que no se tiene una buena distribución normal, aplicando el logaritmo natural se puede observar una mejor distribución donde el índice de Anderson es 7.984.

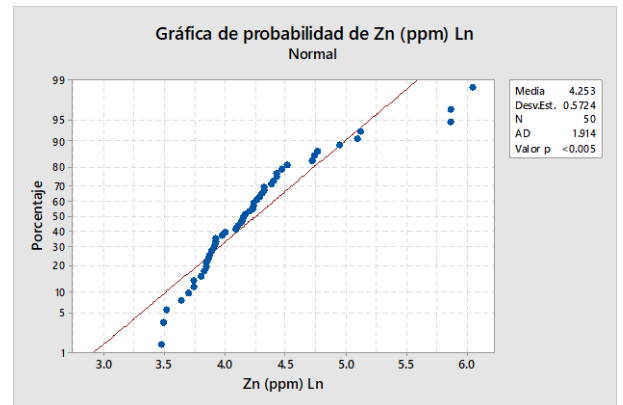
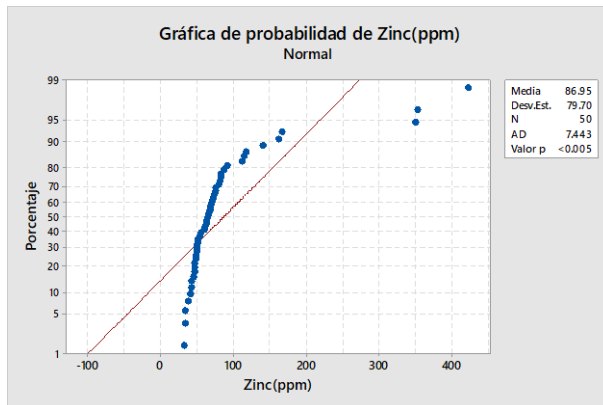
Figura 13 Relación entre la distribución no normalizada y normalizada de Ag



Fuente: Elaboración propia.

- **Zn:** en la figura N°14 se observa que no se tiene una buena distribución normal, aplicando el logaritmo natural se puede observar una mejor distribución donde el índice de Anderson es 1.914.

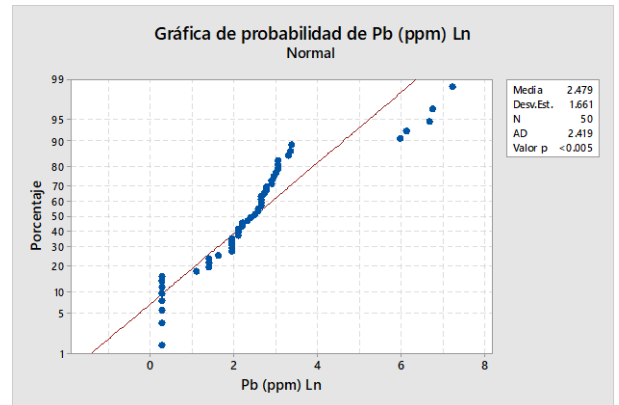
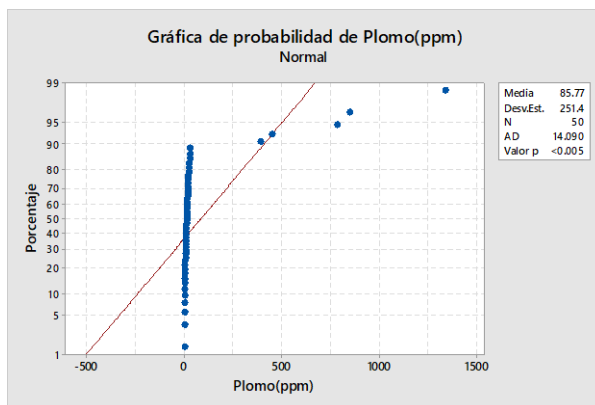
Figura 14 Relación entre la distribución no normalizada y normalizada de Zn



Fuente: Elaboración propia.

- **Pb:** en la figura N°15 se observa que no se tiene una buena distribución normal, aplicando el logaritmo natural se puede observar una mejor distribución donde el índice de Anderson es 2.419.

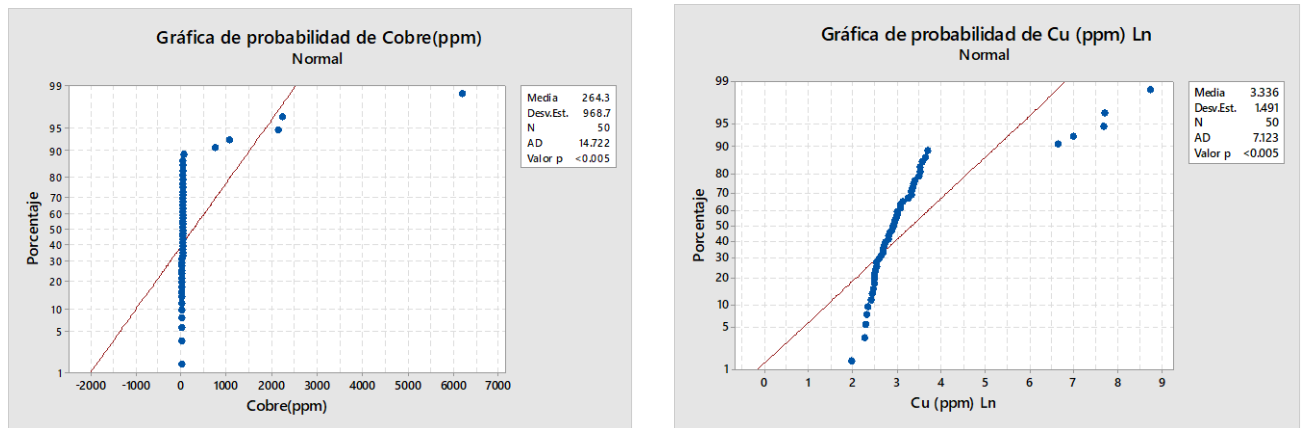
Figura 15 Relación entre la distribución no normalizada y normalizada de Pb



Fuente: Elaboración propia.

- **Cu:** en la figura N°16 se observa que no se tiene una buena distribución normal, aplicando el logaritmo natural se puede observar una mejor distribución donde el índice de Anderson es 7.123.

Figura 16 Relación entre la distribución no normalizada y normalizada Cu



Fuente: Elaboración propia.

- b. Distribución de elementos geoquímicos.** Se obtendrán los resultados de los parámetros estadísticos de los elementos.

Tabla 6 Estadística descriptiva

Variable	Media	Desv.Est.	Varianza	Mínimo	Q1	Mediana	Q3	Máximo
Au(ppm)	0.5	1.743	3.039	0.006	0.006	0.006	0.017	10.66
Aluminio(ppm)	19222	8874	78750322	7700	12300	16450	24825	52000
Calcio(ppm)	1693	6755	45627345	0	0	1	1	36100
Cromo(ppm)	74.7	33.48	1120.83	28	47.5	71.5	92.25	163
Cobre(ppm)	264	969	938385	7	12	18	29	6192
Hierro(ppm)	47586	33112	1096383269	16100	27775	36900	51425	150000
Potasio(ppm)	3320	4020	16162041	800	1400	2150	3325	24500
Magnesio(ppm)	8028	5700	32493486	700	4900	6800	9800	32700
Manganeso(ppm)	853	1353	1829885	215	516	618	740	9936
Sodio(ppm)	830	1271	1616020	100	300	500	625	7100
Fosforo(ppm)	29.7	93.6	8767.4	0	0	0.1	0.1	394
Estroncio(ppm)	41.63	31	961.22	8.6	20.03	33.3	56.33	142.3
Titanio(ppm)	1794	837	700167	300	1200	1800	2325	4300
Vanadio(ppm)	81.52	48.06	2309.64	12	43	64.5	111	225
Zinc(ppm)	87	79.7	6351.5	32.2	47.9	63.9	83	422
Circonio(ppm)	7.08	13.22	174.66	0.33	1.37	2.9	4.58	63.7
Plata(ppm)	1.203	2.968	8.807	0.13	0.13	0.13	0.4	10
Arsénico(ppm)	218	780	608717	2	6	13	22	4432
Berilio(ppm)	0.609	0.2787	0.0776	0.3	0.325	0.6	0.8	1.3
Cobalto(ppm)	11.57	7.31	53.4	3.4	7	10	14	45
Lantano(ppm)	24.12	18.91	357.41	3	11.1	18.1	31.05	76.7
Molibdeno(ppm)	4.7	4.868	23.7	0.65	2	3	5	24.19
Níquel(ppm)	22.54	12.95	167.72	6.6	13	18.5	30.25	69
Plomo(ppm)	85.8	251.4	63207.3	1.3	4.8	11.5	19.3	1335
Antimonio(ppm)	33.1	145.7	21226.4	3.3	3.3	3.3	9	1026
Sc(ppm)	6.018	2.874	8.261	1.2	3.775	5.5	7.825	12.8
Itrio(ppm)	13.5	8.25	68.04	4.2	7.77	9.85	17.15	37.2

Fuente: Elaboración propia. La data cruda se encuentra en el anexo I. Ver base de datos de la geoquímica cruda.

- c. **Cálculo del background y del Threshold.** Calcular el background y el threshold de los elementos geoquímicos a partir de la estadística descriptiva, donde calcularemos el umbral, y anomalías bajas y altas. Donde se usará la información de la media y la desviación estándar. El cual se aplicará la siguiente formula:

$$Media = Background$$

$$Media + Desv.est. = Threshold$$

$$Media + 2(Desv.est.) = Threshold$$

$$Media + 3(Desv.est.) = Threshold$$

Tabla 7 Cálculo de background y Threshold

Variable	Media	Desv.Est.	BG_ppm	T1_ppm	T2_ppm	T3_ppm
Au(ppm)	0.5	1.743	0.5	2.243	3.986	5.729
Aluminio(ppm)	19222	8874	19222	28096	36970	45844
Calcio(ppm)	1693	6755	1693	8448	15203	21958
Cromo(ppm)	74.7	33.48	74.7	108.18	141.66	175.14
Cobre(ppm)	264	969	264	1233	2202	3171
Hierro(ppm)	47586	33112	47586	80698	113810	146922
Potasio(ppm)	3320	4020	3320	7340	11360	15380
Magnesio(ppm)	8028	5700	8028	13728	19428	25128
Manganeso(ppm)	853	1353	853	2206	3559	4912
Sodio(ppm)	830	1271	830	2101	3372	4643
Fosforo(ppm)	29.7	93.6	29.7	123.3	216.9	310.5
Estroncio(ppm)	41.63	31	41.63	72.63	103.63	134.63
Titanio(ppm)	1794	837	1794	2631	3468	4305
Vanadio(ppm)	81.52	48.06	81.52	129.58	177.64	225.7
Zinc(ppm)	87	79.7	87	166.7	246.4	326.1
Circonio(ppm)	7.08	13.22	7.08	20.3	33.52	46.74
Plata(ppm)	1.203	2.968	1.203	4.171	7.139	10.107
Arsénico(ppm)	218	780	218	998	1778	2558
Berilio(ppm)	0.609	0.2787	0.609	0.8877	1.1664	1.4451
Cobalto(ppm)	11.57	7.31	11.57	18.88	26.19	33.5
Lantano(ppm)	24.12	18.91	24.12	43.03	61.94	80.85
Molibdeno(ppm)	4.7	4.868	4.7	9.568	14.436	19.304
Níquel(ppm)	22.54	12.95	22.54	35.49	48.44	61.39
Plomo(ppm)	85.8	251.4	85.8	337.2	588.6	840
Antimonio(ppm)	33.1	145.7	33.1	178.8	324.5	470.2
Sc(ppm)	6.018	2.874	6.018	8.892	11.766	14.64
Itrio(ppm)	13.5	8.25	13.5	21.75	30	38.25

Fuente: Elaboración propia.

d. Correlación multivariable

- **Correlación de Sperman.** Se observa el análisis del comportamiento y correlación geoquímica en PPM, cual se empleará con el software Minitab 18. Ver la tabla correlación de Sperman en el anexo II.

Donde se observa una buena correlación baja entre el Au y la Ag con un valor de 0.424 de igual forma el Au con el Pb con un valor de 0.411

También se observa una buena correlación entre el Cu y el Zn con un valor de 0.877 y ya con una moderada correlación tendremos entre el Cu y Ag con un valor de 0.649.

- **Correlación de Pearson.** Se observa el análisis del comportamiento y correlación geoquímica en PPM normalizado, cual se empleará el software Minitab 18. Ver la tabla de correlación de Pearson en el anexo II.

Donde se observa una muy buena correlación entre el Au y Ca con un valor de 0.905, de igual forma el Au – P con un valor de 0.91.

También se observa una correlación buena de Au – Cu con un valor de 0.895. de igual forma Au – Ag con un valor de 0.857 y Au – Pb con un valor de 0.802. Con una moderada correlación tenemos al Au – Zn con un valor 0.722.

Ahora se tiene una buena muy buena correlación entre el Cu – Ag con un valor de 0.925 y una buena correlación del Cu – Pb con un valor de 0.802 y el Cu – Zn con un valor de 0.873.

- e. Análisis de anomalías mediante el índice de enriquecimiento local (LREI).** De acuerdo el análisis aplicado aquí se usará la ventana deslizante

el cual se tuvo un criterio simétrico con las muestras el cual se clasifico en 4.
Ver la figura 17.

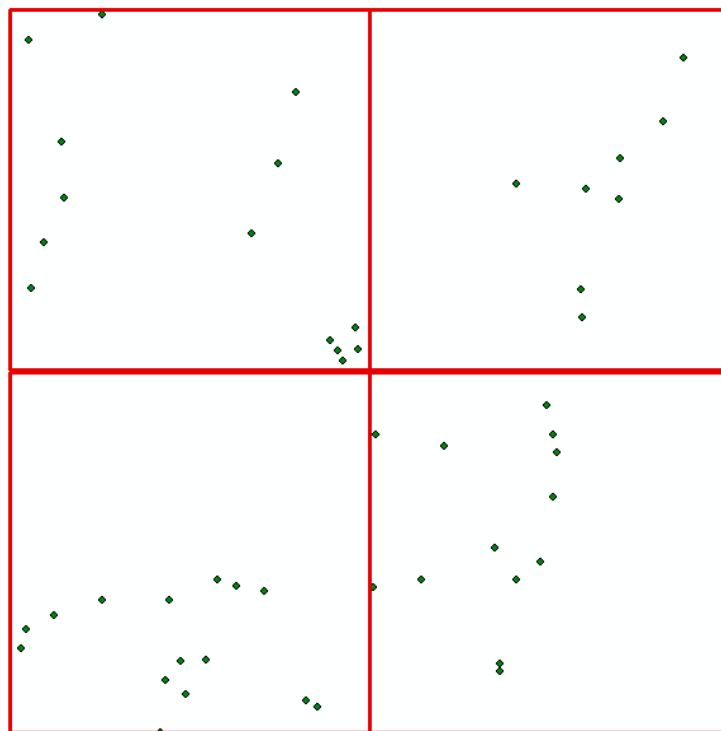
Posteriormente a ello el índice de enriquecimiento local (LREI) en el cual implica el uso de los datos normalizados, donde se obtendrá la mediana de cada ventana deslizante y posteriormente se aplicará lo siguiente:

$$LREI = \left(\frac{Ln[Metal]}{Mediana[Metal]} \right) - 1 \dots (1)$$

$Ln[Metal]$ = Valor del elemento normalizado.

$Mediana[Metal]$ = Mediana de una ventana deslizante del elemento normalizado.

Figura 17 Aplicación de la ventana deslizante



Fuente: Elaboración propia.

A continuación, obtenido los datos LREI para cada elemento aplicando la formula (1), se clasificará entre valores negativos y positivos siendo el cero

un valor medio. Donde el cual los valores mayores a cero nos indicaran anomalías sobre cada elemento planteado. Ver plano geoquímico del Au, Ag, Zn, Pb, Cu en el anexo I.

Análisis espectral

Las anomalías espectrales identificadas corresponden a óxidos de hierro encontradas cerca y dentro del área de estudio. El reconocimiento de campo nos dio a conocer de las anomalías espectrales identificadas, algunas de ellas corresponden correctamente a alteraciones en rocas volcánicas. Las zonas de oxidación están en las zonas de cizalla donde se presenta rocas metamórficas y sedimentarias con intercalaciones volcánicas los cuales pasaron por un intenso cizallamiento y/o fracturamiento. Ver plano de anomalías espectrales en el anexo I.

Ocurrencias minerales

En la zona de estudio se observan vetas y vetillas de sulfuros, donde el oro se encuentra erosionado producto del intemperismo en la zona. La pirita es el mineral que abunda, presentándose en vetas fracturadas producto del cizallamiento en la zona. La galena y la esfalerita se asocian en las zonas de oxidación de los filones. Las vetas tienen diferente potencia de centímetros, donde también podemos encontrar cajas falsas con presencia de espejos de fallas. Las vetas están asociadas con cuarzo grisáceo a lechoso con patina oxidada el cual la ser muestreado presentan altas cantidades de ppm de Au, pero mucho macroscópicamente.

Su emplazamiento principal se debe a las rocas graníticas y granodioríticas cual es el responsable de la mineralización el granito y

granodiorita de Paucartambo, cuyas soluciones mineralizantes circularon a través de las fracturas preexistentes producto del cizallamiento ocurrido en la zona.

Por el Norte al NW de la zona de estudio se encontró otros yacimientos informales de oro el cual nos indica una relación entre sí para correlacionar la predicción mineralógica en el área.

El cual se puede observar en la fotografía 1 y 2 donde se presenta dos mineras informales el cual no sé muestra en GEOCATMIN.

Fotografía 1 *Campamentos mineros por la ruta hacia la zona de estudio.*



Fotografía 2 *Garita de acceso hacia una bocamina cerca de la zona de estudio.*



4.2. Presentación, análisis e interpretación de resultados

4.2.1. Presentación y análisis de la información geológica, geoquímica, estructural y anomalías espectrales con la ocurrencia de depósitos minerales.

En este capítulo, se llevó a cabo la revisión exhaustiva de la información integrada en la base de datos diseñada específicamente para la predicción y/o pronóstico de ocurrencias mineralizantes en la concesión minera. Esta base de datos fue creada y gestionada utilizando Jupyter Notebook, una plataforma interactiva que permite combinar código, texto y resultados visuales de manera eficiente. Estructurando los datos recolectados (geológicos, geoquímicos, estructurales y espectrales) para su procesamiento mediante Python,

aprovechando bibliotecas avanzadas como Scikit-learn (Sklearn). Estas herramientas permitieron el desarrollo de modelos predictivos basados en algoritmos de aprendizaje automático, como Bosques Aleatorios y Redes Neuronales Artificiales, con el propósito de generar mapas de prospectividad mineral.

Mapas de predicción mineral con Python y Sklearn

Importemos las librerías principales

Dentro de las librerías que se van a utilizar está Rasterio que permite la lectura de los metadatos de la información de entrada a fin de obtener la georeferencia de las mismas.

```
In [1]: #Importing the initial few python packages as their usual aliases
import pandas as pd
import numpy as np
from skimage import io
import matplotlib.pyplot as plt
from IPython.display import Markdown as md
```

```
In [2]: import rasterio as rio
```

Habiendo realizado las importaciones de las librerías a usar y los ráster como geoquímica, geología, fallas geológicas y anomalías espectrales se pasará a ver la resolución espacial.

4.2.2. Análisis y selección adecuada de un tamaño de celda unitaria (pixel).

Ya que al predecir mapas en un formato ráster es importante ver la resolución tanto de la información geológica y la escala de trabajo el cual es una metodología dada por Carranza (2009) el cual implica en seleccionar un tamaño de pixel para los algoritmos a usar, puesto que cada pixel dará un valor que se diferenciará de otros, cual se puede ver en los planos del Anexo III. Donde los

Importemos los rasters de entrada

Para este proyecto se seleccionaron la geoquímica, Geología, Fallas, Anomalías espectrales Comencemos con la geoquímica:

NOTA: dentro del paréntesis escribir los elementos a utilizar, siempre y cuando hayamos nombrado nuestros archivos solo con el nombre del elemento.

```
In [3]: Au = io.imread('./Python/Au_f.tif')
Ag = io.imread('./Python/Ag_f.tif')
Pb = io.imread('./Python/Pb_f.tif')
Zn = io.imread('./Python/Zn_f.tif')
Cu = io.imread('./Python/Cu_f.tif')
Mo = io.imread('./Python/Mo_f.tif')
```

```
In [4]: #Revisemos qué se importó para saber bien con qué estamos trabajando
plt.figure(figsize=(12, 20))
plt.imshow(Cu)
```

planos a través del Arcgis, pueden ser exportados a diferente tipo de resolución espacial de pixeles al criterio de trabajo del usuario, puede aumentar o disminuir el tamaño del pixel.

```
In [10]: #Analysing the shape of each numpy array,
print('Au shape is',Au.shape)
print('Ag shape is',Ag.shape)
print('Pb shape is',Pb.shape)
print('Zn shape is',Zn.shape)
print('Cu shape is',Cu.shape)
print('Mo shape is',Mo.shape)
print('Geo is',geo.shape)
print('Falla shape is',Falla.shape)
print('Oxi shape is',Oxi.shape)
print('Area of Interest shape is',AOI.shape)]

Au shape is (97, 128)
Ag shape is (97, 128)
Pb shape is (97, 128)
Zn shape is (97, 128)
Cu shape is (97, 128)
Mo shape is (97, 128)
Geo is (97, 128)
Falla shape is (97, 128)
Oxi shape is (97, 128)
Area of Interest shape is (97, 128)
```

4.2.3. Análisis de depósitos de interés.

Tiene una gran importancia ya que la selección y su análisis al plantear el número de ocurrencias de depósitos minerales conocidas es mucho menor que las ubicaciones no mineralizadas. Aunque el algoritmo implica obtener buenos resultados.

Tabla 8 *Capas de entrada usadas para generar mapas de prospectividad mineral (MPM) de depósitos tipo vetiforme usando Machine Learning.*

Algoritmos	Franja Metalogenetica	Deposito	Capas de entrada
RF y ANN	Franja de depósitos de Li-U-W-Sn-Mo,	Vetiformes	- Geología local.
	Au-Cu-Pb-Zn relacionados con intrusivos		- Proximidad de fallas geológicas.
	y pórfidos-skarn de Cu-Ag del Pérmico		- Planos geoquímicos (Au, Ag, Pb, Zn, Cu y Mo).
	Triásico.		- Plano de anomalías espectrales (óxidos de Fe).

Fuente: Elaboración propia. *RF. Bosques aleatorios, ANN. Red neuronal artificial.*

Tabla 9 Capas de entrada para los algoritmos RF y ANN

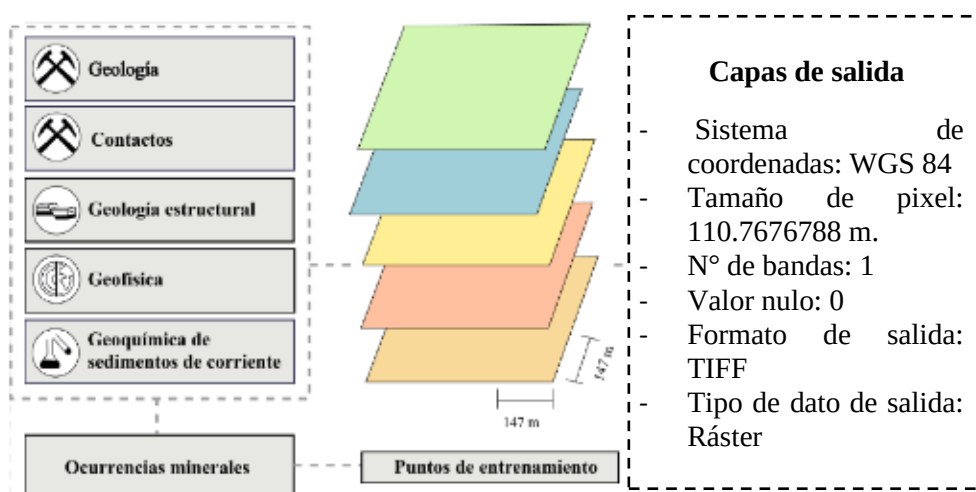
Proxi	Factor condicionante
Geología local	- Proximidad a cuerpos intrusivos. - Proximidad a rocas volcánicas.
Geología estructural	- Proximidad de fallas geológicas. - Proximidad de contactos litológicos y tectónicos.
Geoquímica	- Planos de concentración geoquímica de Au, Ag, Pb, Zn, Cu y Mo.
Anomalías espectrales	- Planos de anomalías espectrales (Óxidos de Fe).

Fuente: Elaboración propia.

4.2.4. Creación de base de datos

Teniendo la información requerida como la geología, geoquímica, geología estructural y anomalías espectrales. Se procesará las capas de entrada incluyendo la geología económica como ocurrencia para generar los puntos de entrenamiento. El procesamiento de las capas se realizará ajustando los ráster con un mismo sistema de coordenadas, tamaño de celdas, numero de filas y columnas, extensión, numero de bandas, valores nulos y formatos de salida. (Figura 18).

Figura 18 Características de las capas de salida.



Fuente: Elaboración propia.

Geología

Se tuvieron en cuenta las rocas volcánicas y plutónicas el cual está a escala 1: 50.000 de composición félsica e intermedia. Donde se calculó su distancia euclidiana dentro de un margen. Ver plano en el anexo III.

Geoquímica

Los valores geoquímicos de Au, Cu, Pb, Zn, Cu y Mo fueron tratados para el procesamiento el cual se inició con la eliminación de datos debajo del límite de detección, posteriormente se continuo con un análisis estadístico a través de la revisión de histogramas de cada elemento, que llevo a una normalización a través de la transformación logarítmica y terminando con la interpolación de Kriging para los planos geoquímicos. Ver plano en el anexo III.

Fallas

Se tuvieron en cuenta las fallas para ver la proximidad que se tiene dentro del ambiente del depósito el cual fueron definidas a 1: 50.000 a través de un buffer de 500 m. Ver plano en el anexo III.

Anomalías espectrales

Habiendo extraído anomalías espectrales de imágenes de alta resolución y habiendo validado cada una de ellas se aplicó una distancia euclidiana de 500 m para verificar su proximidad en relación con los depósitos. Ver plano en el anexo III.

Ocurrencias minerales

Dentro y cerca de la zona de estudio se encontró 2 ocurrencias minerales que pertenecen a depósitos vetiformes de oro y polimetálicos, el cual se usaron como apoyo para la prospectividad de mineralización. Ver plano en el anexo III.

Se seleccionaron una cantidad de depósitos y no depósitos, el cual incluye a 12416 datos de entrenamiento donde se clasificó como 1 depósitos o zonas favorable y 0 no depósitos zonas no favorables.

Los planos de entrada que representan la información geológica para los algoritmos de RF y ANN. Donde las variables fueron clasificadas como se muestra en la Tabla para entender los modelos de aprendizaje. Ver planos en el anexo V.

Tabla 10 *Clasificación de las variables de entrada para los algoritmos de Machine Learning*

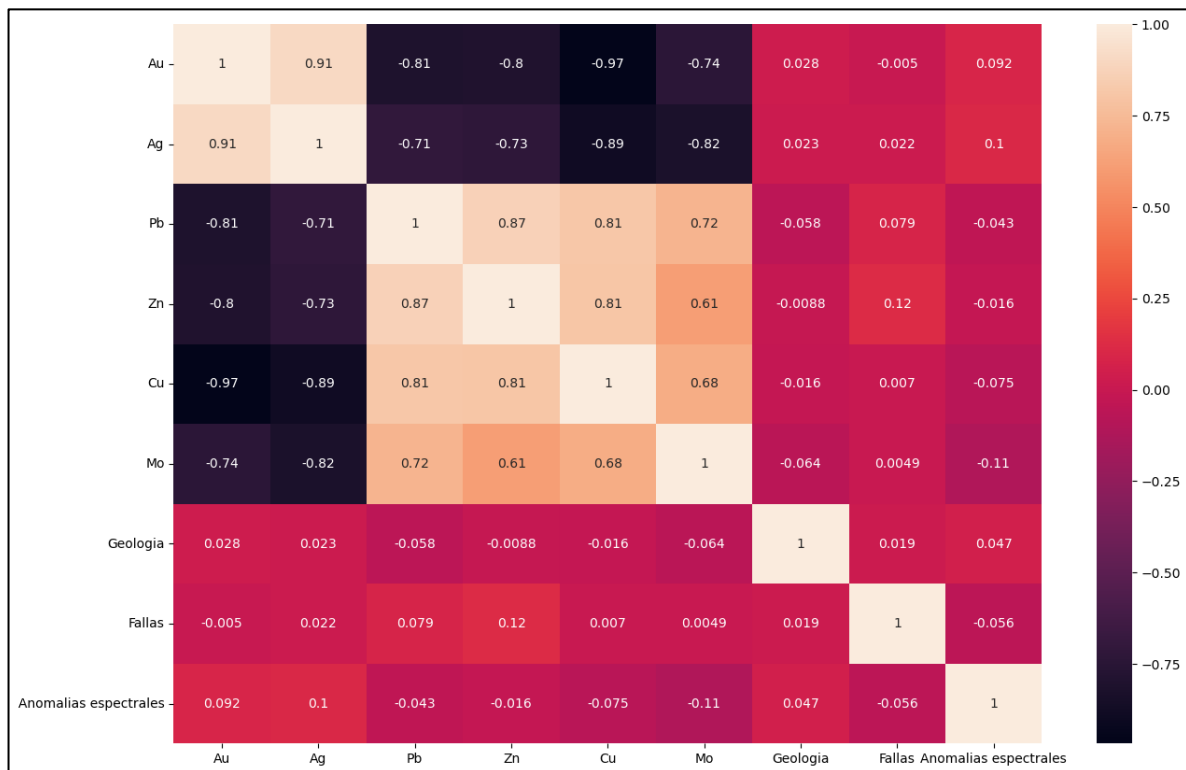
Variable	Tipo de Variable	Descripción
X1 (Geología)	Continua	Proximidad a las rocas plutónicas y volcánicas
X2 (Fallas)	Continua	Proximidad a las fallas geológicas
X3 (Anomalías espectrales)	Continua	Proximidad a las anomalías espectrales
X4 (Au)	Continua	Concentración geoquímica de Au
X5 (Ag)	Continua	Concentración geoquímica de Ag
X6 (Pb)	Continua	Concentración geoquímica de Pb
X7 (Zn)	Continua	Concentración geoquímica de Zn
X8 (Cu)	Continua	Concentración geoquímica de Cu
X9 (Mo)	Continua	Concentración geoquímica de Mo
Ocurrencias minerales	Nominal/ Binaria	Ocurrencias minerales de depósitos y no depósitos

Fuente: Elaboración propia.

4.2.5. Análisis entre las variables de entrada

Para ver la relación de una variable con otra tendremos que elaborar una Matriz de correlación para las variables predictoras el cual se muestra en la figura 19.

Figura 19 Matriz de correlación para las variables predictorias



Fuente: Elaboración propia.

Para explicar la figura 18 tendremos en cuenta que se realiza del tratamiento de 1.2 millones de píxeles que teníamos inicialmente, solo se tiene 12416 píxeles de información el cual al crear la matriz de correlación de los píxeles entrenados el cual detalla que la variable “y” será el objetivo de predecir depósitos, mientras que los valores “x” son los atributos de entrada, generando una matriz definida para el modelo. Además, se representa como un mapa de color llamado heatmap. El cual se elabora correlacionando las diferentes variables ya que cada celda tiene un coeficiente de correlación de la siguiente manera.

- 1: Correlación positiva perfecta.
- 0: Sin correlación.
- -1: Correlación negativa perfecta.

Tabla 11 Interpretación de los valores de coeficientes de la matriz de correlación

Valor de r	Fuerza de correlación
$ r < 0.3$	Ninguna o débil
0.3 – 0.5	Débil
0.5 – 0.7	Moderada
$ r > 0.7$	Fuerte

Fuente. Tomado de Mindrila & Balentyne (2017).

4.2.6. Diseño de los modelos de Machine Learning usando bosques aleatorios y redes neuronales artificiales en el lenguaje de programación Python.

El testeo y entrenamiento del Machine Learning conduce a la clasificación de depósitos minerales tipos vetiformes y los no depósitos, por tanto, nos llevará a la creación del mapa de prospectividad mineral dentro y cerca de la zona de estudio. Se aplicará dos algoritmos los cuales son los bosques aleatorios (RF) y las redes neuronales artificiales (ANN), el testeo y el entrenamiento de los algoritmos y llevo a cabo usando el lenguaje de programación Python a través de Jupyter Notebook de Anaconda Navigator con la Librería Pandas (<https://pandas.pydata.org/>). Los datos entrenados se deberán medir métricas como la matriz de confusión y las curvas ROC.

Bosques aleatorios (RF)

Para el modelo de bosques aleatorios se empleó la librería Scikit-learn. Donde se definió los hiperparámetros de la Tabla 12. El algoritmo es de una categoría BOOTSTRAP que usa el índice de Accuracy para analizar y calcular los datos.

Tabla 12 Hiperparámetros definidos del modelo RF

Hiperparámetro	Descripción
Numero de arboles	500
Algoritmo categórico	Bootstrap: True
Función de perdida	Entropía
Métricas	Matriz de confusión, ROC
Valor	0.99914 (99.91%)

Fuente: Elaboración propia.

```
In [120]: #Importing the grid search module
from sklearn.model_selection import GridSearchCV

#Creating our initial test classifier
RF = RandomForestClassifier()

#Selecting our algorithm variables that will be tested on our dataset and running them through a grid search
grid_param_RF = { 'n_estimators': [100,200,300,500,800,1000], 'criterion': ['gini', 'entropy'], 'bootstrap': [True, False]}
gd_sr_RF = GridSearchCV(estimator=RF, param_grid=grid_param_RF, scoring='accuracy',cv=10, n_jobs=-1)
gd_sr_RF.fit(X_train, y_train)

#Assigning a variable to the best parameters and the classification accuracy
best_parameters_RF = gd_sr_RF.best_params_
best_result_RF = gd_sr_RF.best_score_

print(best_parameters_RF)
print(best_result_RF)

{'bootstrap': False, 'criterion': 'entropy', 'n_estimators': 500}
0.9991409394116808
```

Otra forma de definir la precisión de nuestro modelo es utilizando la matriz de confusión. Como su nombre lo indica (y no me refiero a que la idea es confundirnos) es indicar que tan "confundido" está nuestro modelo con nuestra información de entrada para predecir la información de salida.

Donde observamos que la precisión alcanzada con la matriz de confusión del algoritmo usado para los bosques aleatorios RF es de 99.96% donde la figura

```
In [126]: from sklearn.metrics import confusion_matrix
from sklearn import metrics

disp = metrics.plot_confusion_matrix(model, X_test, y_test)
disp.figure_.suptitle("Confusion Matrix RFC")
print(accuracy_score(y_test, predictions))
print(confusion_matrix(y_test, predictions))
print(classification_report(y_test, predictions))

C:\Users\ASUS\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function plot_confusion_matrix is deprecated; Function 'plot_confusion_matrix' is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: ConfusionMatrixDisplay.from_predictions or ConfusionMatrixDisplay.from_estimator.
  warnings.warn(msg, category=FutureWarning)

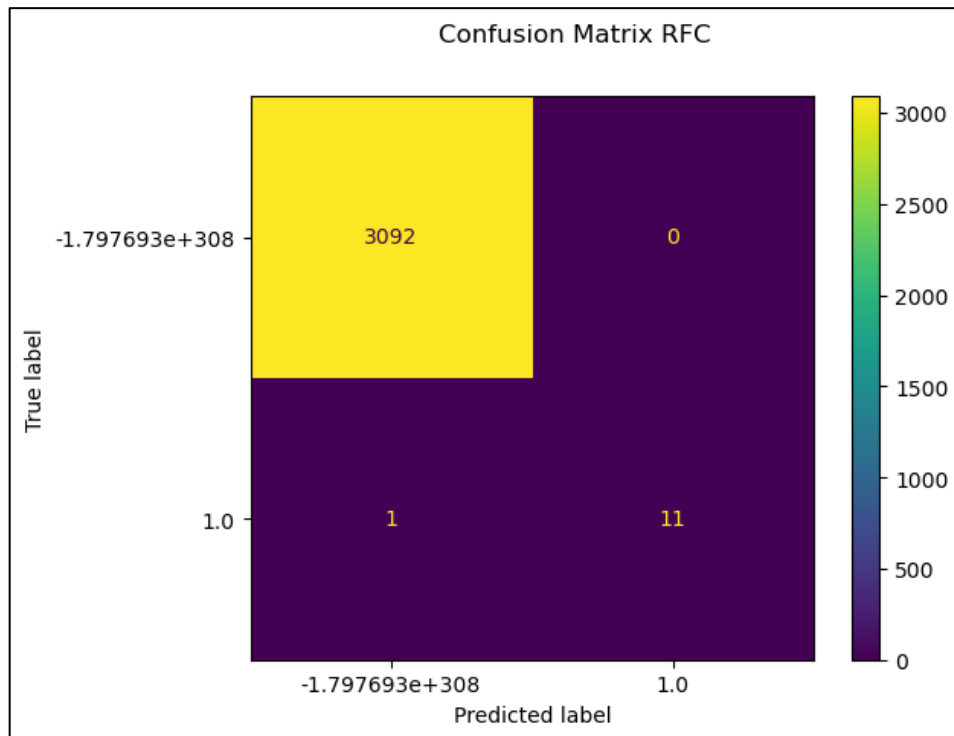
0.9996778350515464
[[3092  0]
 [ 1 11]]
      precision    recall  f1-score   support

-1.797693e+308      1.00      1.00      1.00       3092
         1.0      0.92      0.96         12

   accuracy      1.00      0.96      0.98       3104
  macro avg      1.00      0.96      0.98       3104
 weighted avg      1.00      1.00      1.00       3104
```

20 muestra que clasifico 3092 como verdaderos positivos (VP), 0 como falsos positivos (FP), 1 falso negativo (FN) y 11 verdaderos negativos (VN). Por el cual muestra esa precisión.

Figura 20 Matriz de confusión para le modelo RF



Fuente: Elaboración propia.

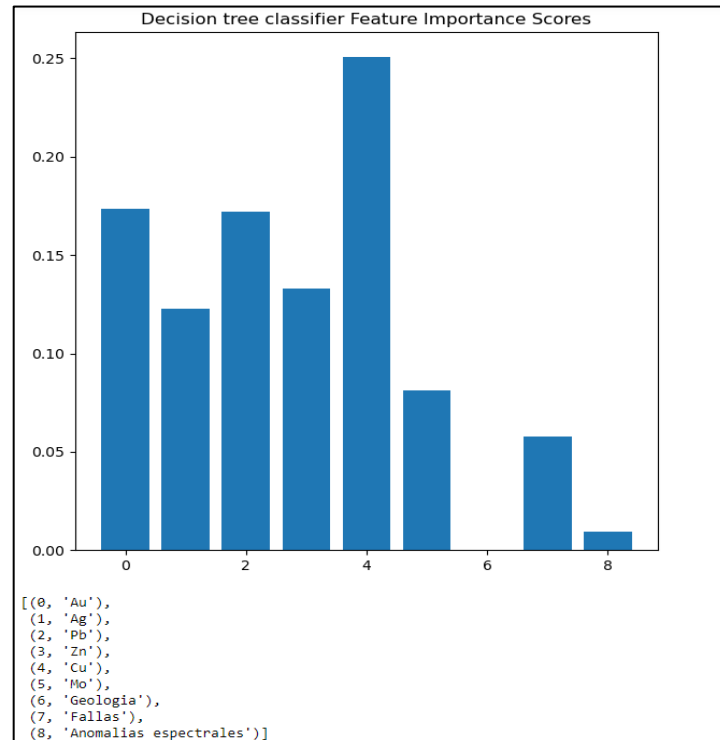
Ahora, suponiendo que tenemos un buen modelo y estamos contentos con los resultados, podemos extraer los predictores más importantes de nuestro modelo.

```
In [129]: importance = model.feature_importances_
k = plt.figure(figsize=(7,7))
plt.bar([x for x in range(len(importance))], importance)
plt.title('Decision tree classifier Feature Importance Scores')
plt.show()
list(enumerate(df_model.columns))
```

En este caso vemos que nuestro modelo catalogó la información detallada de los intrusivos con poca relevancia, ¿podemos retirar esta información de nuestro modelo? Por otro lado, no es sorpresa que el cobre sea importante, al igual que el oro, fallas y anomalías espectrales. Pero, ¿la densidad de las fallas debería

ser importante? Este tipo de puntos es donde debemos ser críticos con nuestros parámetros de entrada y hacer cambios sí o no (ver figura 21).

Figura 21 Características importantes de las variables para el modelo RF



Fuente: Elaboración propia.

Para corroborar con la precisión que se tiene se aplicara la curva ROC el cual es el siguiente, donde también se nos muestra un valor de 99.90%. Donde el

```

In [130]: from collections import Counter
from imblearn.under_sampling import RandomUnderSampler
from imblearn.over_sampling import SMOTE
from sklearn.metrics import roc_auc_score

# summarize class distribution
clf_before = model.fit(X_train, y_train)
print("Before undersampling: ", Counter(y_train))
print("ROC AUC score for untampered data: ", roc_auc_score(y_test, predictions))

# define undersampling strategy
undersample = RandomUnderSampler(sampling_strategy='majority')
X_train_under, y_train_under = undersample.fit_resample(X_train, y_train)
print("After undersampling: ", Counter(y_train_under))
clf_under = model.fit(X_train_under, y_train_under)
pred_under = clf_under.predict_proba(X_test)
y_prob_split1 = np.split(pred_under, 2, axis=1)
print("ROC AUC score for undersampled data Undersampling: ", roc_auc_score(y_test, y_prob_split1[1]))

# define oversampling strategy
SMOTE = SMOTE()
X_train_SMOTE, y_train_SMOTE = SMOTE.fit_resample(X_train, y_train)
print("After oversampling: ", Counter(y_train_SMOTE))
clf_SMOTE = model.fit(X_train_SMOTE, y_train_SMOTE)
pred_SMOTE = clf_SMOTE.predict_proba(X_test)
y_prob_split2 = np.split(pred_SMOTE, 2, axis=1)
print("ROC AUC score for oversampled SMOTE data: ", roc_auc_score(y_test, y_prob_split2[1]))

Before undersampling: Counter({'-1.797693e+308': 9276, '1.0': 36})
ROC AUC score for untampered data: 0.9583333333333333
After undersampling: Counter({'-1.797693e+308': 36, '1.0': 36})

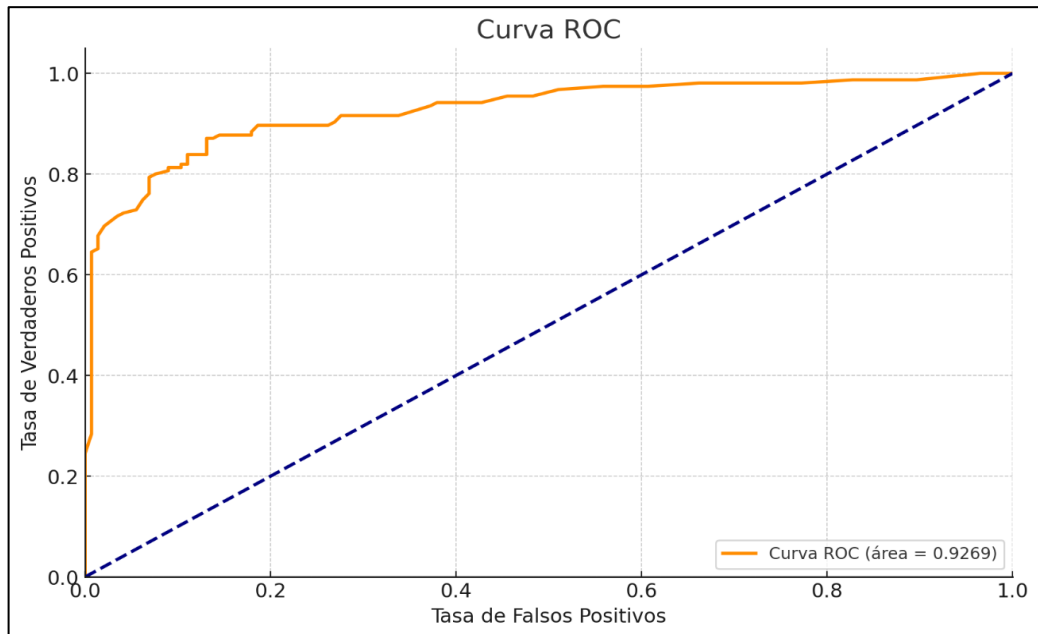
C:\Users\ASUS\anaconda3\lib\site-packages\numpy\core\fromnumeric.py:86: RuntimeWarning: overflow encountered in reduce
return ufunc.reduce(obj, axis, dtype, out, **passkwargs)
C:\Users\ASUS\anaconda3\lib\site-packages\numpy\core\fromnumeric.py:86: RuntimeWarning: overflow encountered in reduce
return ufunc.reduce(obj, axis, dtype, out, **passkwargs)

ROC AUC score for undersampled data Undersampling: 0.9990297542043984
After oversampling: Counter({'-1.797693e+308': 9276, '1.0': 9276})
ROC AUC score for oversampled SMOTE data: 1.0

```

cual se puede ver en la figura 22. Aquí es claro que nuestro modelo es muy bueno ya que está cerca de la esquina superior izquierda. Sin embargo, podemos ver que, al hacer un balance en nuestros datos, si hay diferencia.

Figura 22 Curva ROC de RF



Fuente: Elaboración propia.

Redes Neuronales Artificiales (ANN)

El algoritmo usado para las redes neuronales artificiales fue realizado a través de la librería de Scikit-learn, el cual tiene los siguientes hiperparámetros que se mostrara en la Tabla 13. Donde se usaron 10 variables de entrada, teniendo una precisión de 99.81%.

Tabla 13 *Hiperparámetros definidos del modelo ANN*

Hiperparámetro	Descripción
Variable de entrada	10
Capas ocultas	Dos capas con 100 y 50
Dropout	No aplica
Neuronas requeridas en cada capa oculta	10
Función de activación capas ocultas	ReLu
Función de activación capa de salida	Sigmoide
Optimizador	Adam
Tasa de aprendizaje	0.001
Función de perdida	Entropía cruzada
Métricas	Matriz de confusión, ROC
Épocas	200

Fuente: Elaboración propia.

```
In [132]: from sklearn.neural_network import MLPClassifier

#Creating our initial test classifier
ANN = MLPClassifier()

#Selecting our algorithm variables that will be tested on our dataset and running them through a grid search
grid_param_ANN = {'solver': ['lbfgs', 'sgd', 'adam'], 'max_iter': [50,100,125,150,175,200,250,500,1000], 'activation': ['identity', 'tanh', 'relu']}
gd_sr_ANN = GridSearchCV(estimator=ANN, param_grid=grid_param_ANN, scoring='accuracy', cv=10, n_jobs=-1)
gd_sr_ANN.fit(X_train, y_train)

#Assigning a variable to the best parameters and the classification accuracy
best_parameters_ANN = gd_sr_ANN.best_params_
best_result_ANN = gd_sr_ANN.best_score_

print(best_parameters_ANN)
print(best_result_ANN)
```

```
{'activation': 'relu', 'max_iter': 250, 'solver': 'lbfgs'}
0.9981743521894867
```

```
In [135]: disp = metrics.plot_confusion_matrix(clfann, X_test, y_test)
disp.figure_.suptitle("Confusion Matrix ANN")
print(accuracy_score(y_test, predictionsann))
print(confusion_matrix(y_test, predictionsann))
print(classification_report(y_test, predictionsann))
```

C:\Users\ASUS\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function plot_confusion_matrix is deprecated; Function 'plot_confusion_matrix' is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: ConfusionMatrixDisplay.from_predictions or ConfusionMatrixDisplay.from_estimator.
warnings.warn(msg, category=FutureWarning)

```
0.9980670103092784
[[3088  4]
 [ 2 10]]
      precision    recall  f1-score   support

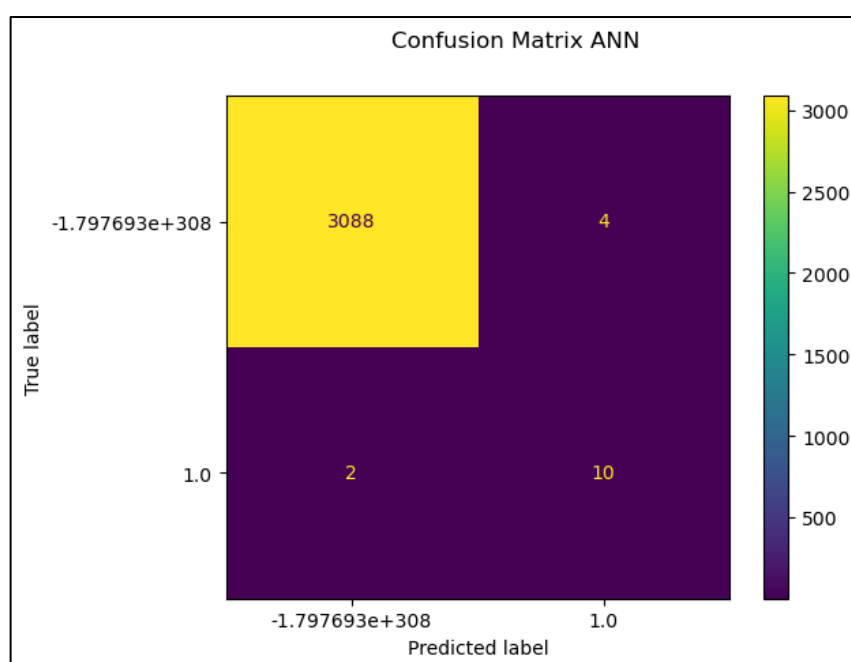
-1.797693e+308      1.00      1.00      1.00      3092
         1.0      0.71      0.83      0.77         12

   accuracy      0.86
  macro avg      0.86
 weighted avg      0.86
```

Para validar el modelo se realizó 10 veces el entrenamiento de la red neuronal, donde las métricas evaluadas muestran un buen desempeño del modelo el cual a través de la matriz de confusión se tiene lo siguiente:

Donde observamos que la precisión alcanzada con la matriz de confusión del algoritmo usado para los bosques aleatorios RF es de 99.80% donde la figura 23 muestra que clasifico algoritmo clasifico en 3088 puntos como verdaderos positivos (VP), 4 falsos positivos (FP), 2 falsos negativos (FN) y finalmente 10 verdaderos negativos (VN). Por el cual muestra esa precisión.

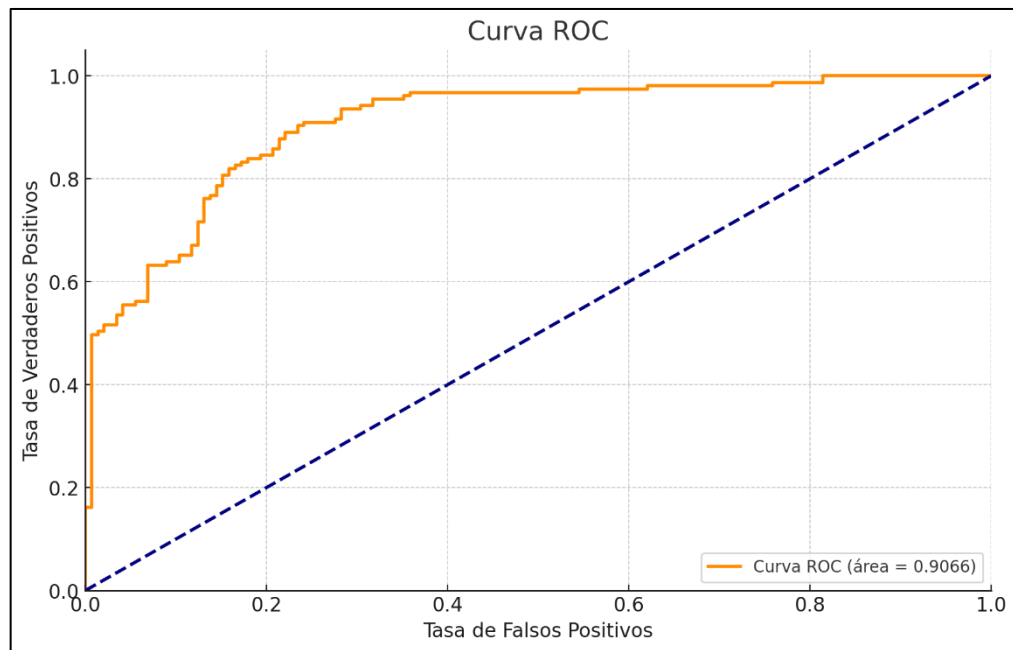
Figura 23 Matriz de confusión para le modelo ANN



Fuente: Elaboración propia.

Para corroborar con la precisión que se tiene se aplicara la curva ROC el cual es el siguiente, donde también se nos muestra un valor de 99.65%. Donde el cual se puede ver en la figura 24. Aquí es claro que nuestro modelo es muy bueno ya que está cerca de la esquina superior izquierda. Sin embargo, podemos ver que, al hacer un balance en nuestros datos, si hay diferencia.

Figura 24 *Curva ROC de ANN*



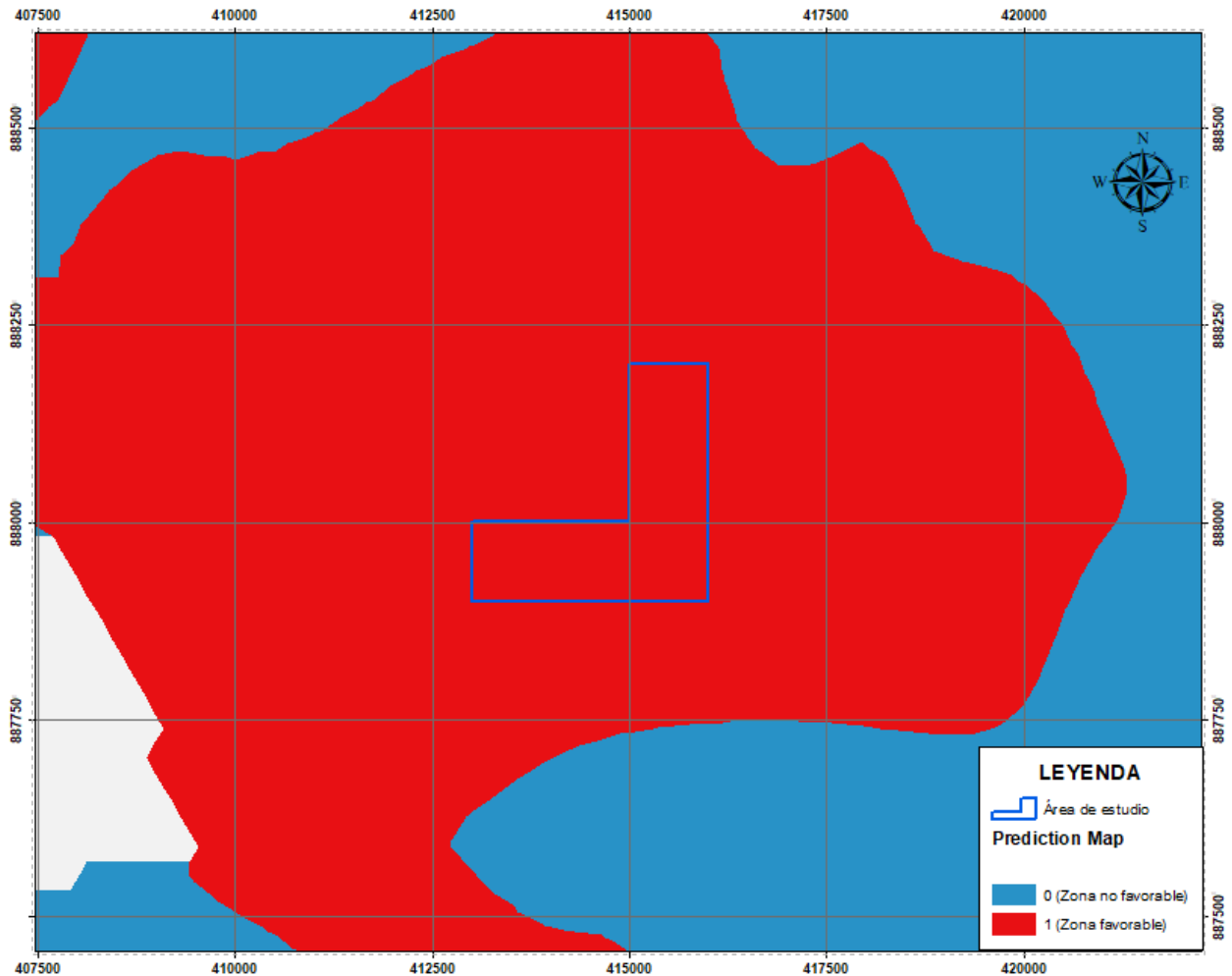
Fuente: Elaboración propia.

4.2.7. Selección de un modelo utilizando los algoritmos de Machine Learning y la información geológica para predecir la probabilidad de ocurrencia de depósitos.

Los mapas de predicción a través del uso de las variables geológicas fueron elaborados de forma correcta y efectiva usando dos modelos que presentan diferentes criterios de decisión. El cual se presenta de la siguiente manera:

Mapa de predicción con Bosques aleatorios (RF)

Figura 25 Mapa de predicción con RF



Fuente: Elaboración propia.

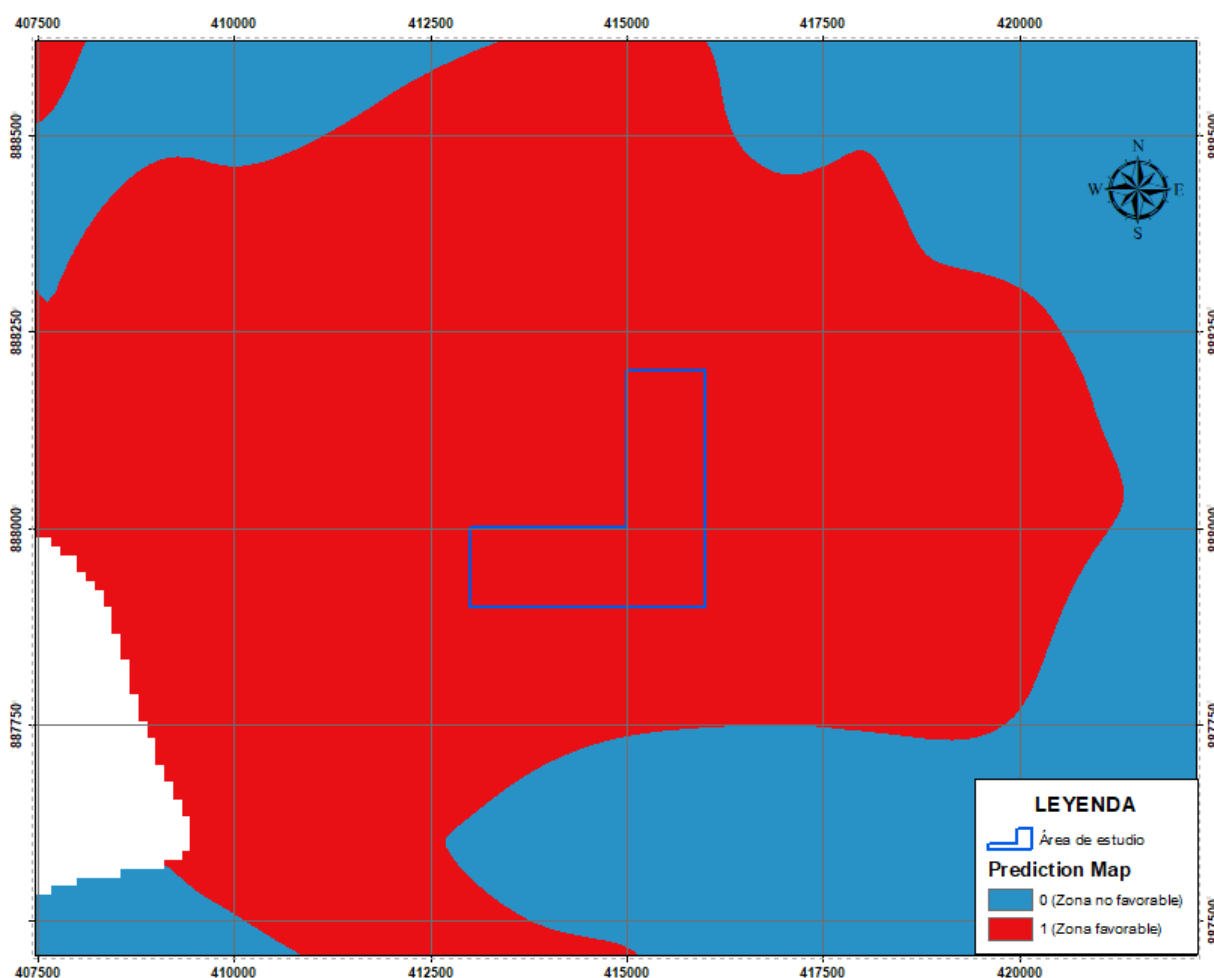
En la figura 25 se puede observar la zona delimitada teniendo zonas favorables en un 80% al área de estudio o la concesión minera “Gus del Aire” pues se muestra una mineralización favorable por el N, NW, NE y SW cerca la zona de estudio donde el cual esta codificada por el color rojo. Donde se presenta un sistema de fallas con dirección NE a SW. Conformada por el intrusivo de Paucartambo. Ver mapa de predicción con los bosques aleatorios RF en el anexo V.

Los polígonos de color rojo son zonas favorables donde mayormente se encuentras rocas plutónicas y volcánicas cómo son la diorita y la granodiorita, lo

cual es de suma importancia ya que son intrusivos de Paucartambo que tienen una gran importancia ya que es donde se hospeda la mayoría de los depósitos tipo vetiforme, es decir, el algoritmo nos da una alta predicción en especial en unidades ígneas y metamórficas cerca y en el área de estudio.

Mapa de predicción con Redes neuronales artificiales (ANN)

Figura 26 Mapa de predicción con ANN



Fuente: Elaboración propia.

En la figura 26 se puede observar la zona delimitada teniendo zonas favorables en un 80% al área de estudio o la concesión minera “Gus del Aire” pues se muestra una mineralización favorable por el N, NW, NE y SW cerca la zona de estudio donde el cual esta codificada por el color rojo. Donde se tiene un sistema de fallas con dirección NE a SW. Ver plano del anexo I. Conformada por

el intrusivo de Paucartambo. Ver mapa de predicción con los Redes neuronales artificiales ANN en el anexo V.

Los polígonos de color rojo son zonas favorables donde mayormente se encuentran rocas plutónicas y volcánicas cómo son la diorita y la granodiorita, lo cual es de suma importancia ya que son intrusivos de Paucartambo que tienen una gran importancia ya que es donde se hospeda la mayoría de los depósitos tipo vetiforme, es decir, el algoritmo nos da una alta predicción en especial en unidades ígneas y metamórficas cerca y en el área de estudio.

Ahora si nos damos cuenta la primera con la segunda predicción tienen mucha similitud el cual se debe a que la precisión de pronóstico no se aleja mucho el cual al usar las mismas variables nos da algo similar.

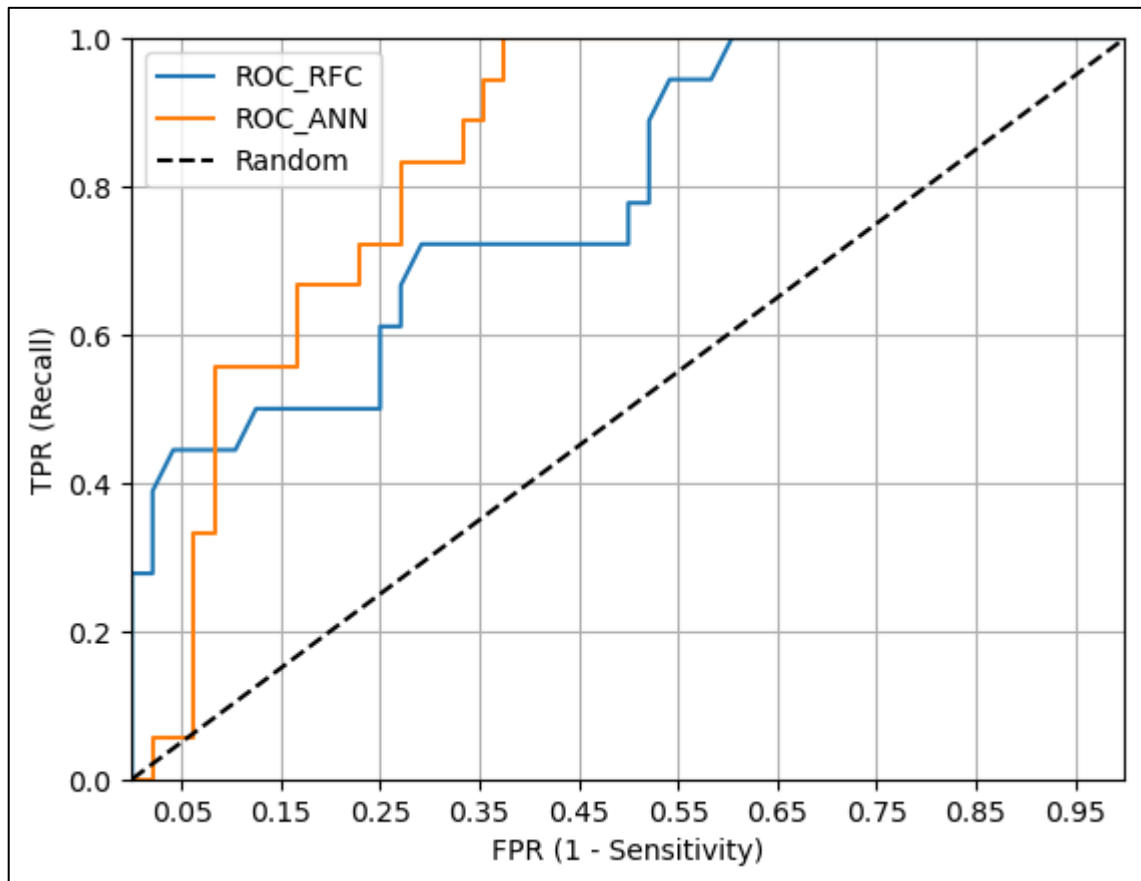
4.2.8. Análisis y comparación de los mapas de predicción

Para comparar las zonas pronosticadas de una metodología con la otra de los algoritmos de Machine Learning usados. Los resultados obtenidos relacionados con la información geológico y puntos conocidos tiene el mismo comportamiento como variables.

```
In [138]: def roc_curve_plot1(y_test, pred_proba_c1, pred_under):  
fprs, tprs, thresholds = roc_curve(y_test, pred_proba_c1, pos_label='2.0')  
fprs1, tprs1, thresholds1 = roc_curve(y_test, pred_under, pos_label='2.0')  
  
plt.plot(fprs, tprs, label='ROC_RFC')  
plt.plot(fprs1, tprs1, label='ROC_ANN')  
plt.plot([0,1], [0,1], 'k--', label='Random')  
  
start, end = plt.xlim()  
plt.xticks(np.round(np.arange(start, end, 0.1),2))  
plt.xlim(0,1); plt.ylim(0,1)  
plt.xlabel('FPR (1 - Sensitivity)'); plt.ylabel('TPR (Recall)')  
plt.legend(); plt.grid()  
  
roc_curve_plot1(y_test, predictionsRF[:,1], predictionsANN[:,1])
```

En este caso para ver cuál de los dos modelos predice mejor aplicaremos una comparación de áreas de sus curvas ROC el cual se muestra a continuación: Ver figura 27.

Figura 27 Curva ROC de comparación entre RF y ANN



Fuente: Elaboración propia.

En resumen, en los dos algoritmos de Machine Learning usados nos resulta una relación entre sí que indica un pronóstico muy bueno ya que las áreas para el RF y ANN son los siguientes 95.83% y 91.60% respectivamente. Por tanto, se tendrá más criterio al modelo RF puesto que la precisión se acerca más al 100%.

```
In [139]: print("Área bajo la curva para el bosque aleatorio: ", roc_auc_score(y_test, predictions))
          print("Área bajo la curva para la red neuronal: ", roc_auc_score(y_test, predictionsann))

Área bajo la curva para el bosque aleatorio: 0.9583333333333333
Área bajo la curva para la red neuronal: 0.9160198361362657
```

4.3. Prueba hipótesis.

4.3.1. Hipótesis general

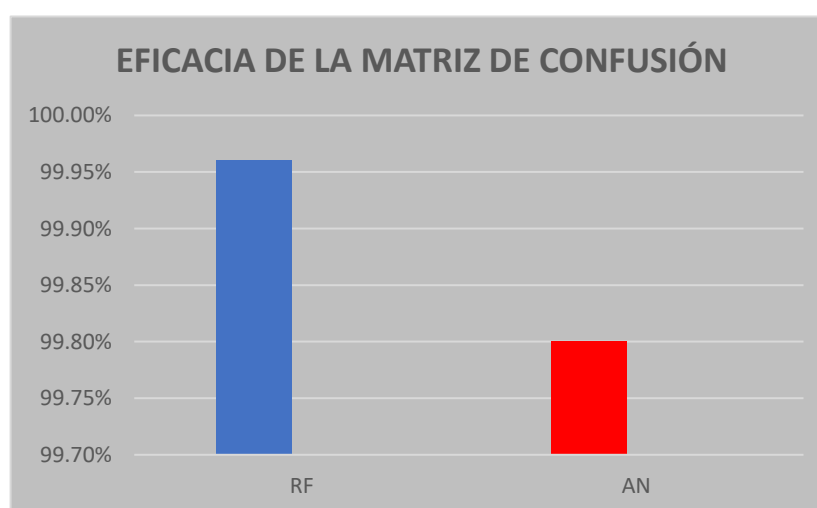
La hipótesis general del trabajo de investigación nos indica que la aplicación de algoritmos de Machine Learning proporciona una predicción favorable de mineralización de la concesión minera Gus del Aire habiendo usado los algoritmos RF y ANN el cual nos proporcionó mapas de predicción precisos

bajado en variables geológicas. Nos menciona que las predicciones muestran estructuras mineralizadas en las zonas de cizallamiento, confirmando la hipótesis general de que los algoritmos de Machine Learning son efectivos para el pronóstico de mineralización el cual se efectúa de acuerdo a las variables dependientes e independientes.

4.3.2. Hipótesis específicas

La primera hipótesis nos indica que la matriz de confusión nos proporcionará una evaluación detallada de cada modelo predictivo en la determinación de la presencia o ausencia de mineralización. Los cuales estarán evaluadas a través de herramientas métricas como la precisión, recall, especificidad y F1-scores. Se puede observar en la Figura 18 y la Figura 21 se tiene lo siguiente:

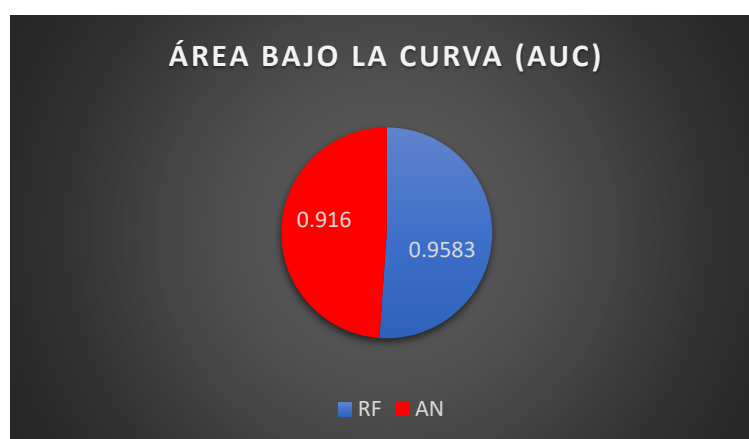
Modelo	Predicción	Valores	Precisión	Recall	F1- Scores	Especificidad
RF	99.96%	-1.797693e+308	1	1	1	1
		1	1	0.92	0.96	
ANN	99.80%	-1.797693e+308	1	1	1	0.9987
		1	0.71	0.83	0.77	



Ambos modelos muestran una buena predicción, confirmando que la matriz de confusión es una herramienta muy útil para evaluar la eficacia de los modelos predictivos.

La segunda hipótesis nos indica que la curva ROC proporcionará una evaluación integral del rendimiento de cada algoritmo y facilitará la selección del algoritmo más eficaz, puesto que al comparar las curvas ROC permitirá identificar un mejor algoritmo como se muestra en la Figura 25. Será evaluada a través de las tasas de verdaderos positivos y la de falsos positivos, llamado también TPR y FPR el cual a través de esta herramienta métrica cuantitativa nos da lo siguiente:

	RF	AN
AUC	0.9583	0.916



Por lo tanto, la curva ROC del modelo RF tiene un AUC superior, indicando un mejor rendimiento general, lo cual confirma la hipótesis de que la curva ROC es útil para seleccionar el algoritmo más eficaz.

La tercera hipótesis nos menciona que el efecto del Machine Learning ayudará a descubrir patrones en datos geológicos, contribuyendo a una toma de decisiones más precisa y rápida. A través de los diferentes algoritmos que puedan describir patrones ocultos en grandes conjuntos de datos, en este caso se aplicó RF y ANN a datos geológicos que ayudara a identificar patrones relevantes en la exploración minera mediante la recolección de datos geológicos como: Anomalías de Au, Ag, Pb, Zn, Cu, Mo, Geología, Fallas y anomalías espectrales los cuales son variables que corresponden para la predicción de mapas y lo cual

se muestra en la Figura 20. Donde se ve la gran importancia de las concentraciones de Au y Pb en relación con las fallas. Por lo tanto, las variables nos muestran que los algoritmos RF y ANN reducen errores humanos y mejoran la visualización de las variables favorables para la exploración minera a través del Machine Learning.

4.4. Discusión de resultados

La implementación de algoritmos de Machine Learning ha mostrado mejoras significativas en los resultados de la exploración geológica, tanto en precisión como en eficiencia. (Rodríguez, R.; Coba, L. & Mamani, M. 2021).

Pero los resultados obtenidos en este estudio destacan la efectividad de los algoritmos de Machine Learning aplicados, específicamente Bosques Aleatorios (RF) y Redes Neuronales Artificiales (ANN), para el pronóstico de mineralización en la concesión minera Gus del Aire. Sin embargo, existen factores que podrían estar influyendo en los resultados y que deben ser considerados:

- 1. Calidad y disponibilidad de datos geológicos:** Una de las limitaciones identificadas fue la escasez de datos geológicos locales detallados, lo que obligó a complementar la información con datos secundarios obtenidos de fuentes como GEOCATMIN e INGEMMET. Esto podría haber introducido ciertas inconsistencias o imprecisiones en la modelización debido a diferencias en las escalas y niveles de detalle de los datos.
- 2. Sensibilidad de los algoritmos a las características geológicas:** Los modelos predictivos dependieron en gran medida de la calidad y relevancia de las variables geológicas seleccionadas, como concentraciones geoquímicas, proximidad a fallas y características litológicas. Aunque se

lograron precisiones altas (99.91% para RF y 99.81% para ANN), es necesario reconocer que estas cifras pueden estar sesgadas si alguna variable clave no fue incluida o si ciertas capas de datos presentaron valores atípicos.

3. **Comparación de algoritmos y desempeño relativo:** La comparación entre las curvas ROC de RF y ANN demostró que ambos algoritmos poseen un rendimiento sobresaliente, pero RF presentó un ligero mejor equilibrio entre sensibilidad y especificidad. Este resultado sugiere que RF es más robusto en la clasificación de zonas mineralizadas bajo las condiciones específicas de este estudio. Sin embargo, el comportamiento de ANN podría mejorar con un ajuste más refinado de los hiperparámetros.
4. **Efecto de los tamaños de celdas y resolución espacial:** El análisis del tamaño de celda unitaria (pixel) utilizado en los modelos mostró que resoluciones más finas permitieron una mayor precisión en la predicción, pero a costa de mayores tiempos de procesamiento. Este aspecto plantea un desafío en términos de balance entre precisión y eficiencia computacional, especialmente al trabajar con grandes volúmenes de datos.
5. **Limitaciones en la validación de los modelos:** Aunque se usaron matrices de confusión y curvas ROC para evaluar el rendimiento de los algoritmos, la validación basada exclusivamente en datos históricos podría no capturar completamente las variaciones naturales de la mineralización en el área de estudio. Esto sugiere la necesidad de realizar validaciones adicionales en campo para corroborar las predicciones generadas.
6. **Impacto del Machine Learning en la exploración minera:** Los resultados resaltan que el uso de Machine Learning no solo redujo significativamente

los tiempos de análisis, sino que también permitió identificar patrones complejos que habrían sido difíciles de discernir mediante métodos convencionales. Sin embargo, la dependencia de recursos computacionales avanzados y personal capacitado podría ser un desafío para la implementación a gran escala en otros proyectos.

CONCLUSIÓN

El presente trabajo de investigación permitió predecir con alta precisión el modelo de mineralización en la concesión minera Gus del Aire, Huánuco 2023, mediante la aplicación de algoritmos de Machine Learning como son el de bosques aleatorios y las redes neuronales artificiales, cumpliendo el objetivo general planteado. La metodología empleada integró información geológica, geoquímica, estructural y espectral, lo que permitió generar un modelo predictivo confiable y reproducible para la exploración de zonas con potencial mineralógico.

En cumplimiento del primer objetivo específico, el análisis del desempeño de los algoritmos de Bosques Aleatorios (Random Forest) y Redes Neuronales Artificiales (Artificial Neural Networks) evidenció una alta precisión y capacidad de predicción, alcanzando valores de 99.91 % para el modelo Random Forest y 99.81 % para el modelo de Redes Neuronales Artificiales, según las métricas obtenidas mediante la matriz de confusión. Estos resultados demuestran que ambos modelos son eficientes en la clasificación de áreas mineralizadas, destacando el algoritmo Random Forest por su mejor desempeño y estabilidad.

En relación con el segundo objetivo específico, la comparación de las curvas ROC (Receiver Operating Characteristic) mostró áreas bajo la curva (AUC) de 0.9583 para el modelo Random Forest y 0.916 para el modelo de Redes Neuronales Artificiales, confirmando una excelente sensibilidad y especificidad en la predicción de zonas de interés mineral. Estos valores validan la eficacia de los algoritmos utilizados y su potencial para ser aplicados en futuros estudios de exploración geológica en contextos similares.

Respecto al tercer objetivo específico, la aplicación del Machine Learning demostró una notable reducción en los tiempos de análisis geológico y una mayor eficiencia en la integración de datos multivariados, optimizando la identificación de áreas prioritarias para la exploración. Asimismo, la automatización del procesamiento de información geocientífica contribuyó a disminuir los costos operativos y los riesgos económicos asociados a la etapa inicial de prospección minera.

En síntesis, se concluye que la aplicación de algoritmos de Machine Learning representa una herramienta innovadora, precisa y replicable en la exploración geológica, permitiendo pronosticar de manera confiable las zonas de mineralización y fortalecer la toma de decisiones estratégicas en proyectos mineros. La metodología desarrollada constituye un aporte significativo para la minería moderna en el Perú, promoviendo la digitalización, automatización y eficiencia en los procesos de exploración de recursos minerales.

RECOMENDACIONES

Se recomienda aplicar otros modelos y/o algoritmos de Machine Learning por ejemplo algoritmos de regresión, bayesianos, Componentes principales, máquinas de soporte vectorial, entre otros. Con el fin de comparar y evaluar la calidad de predicción de un modelo con otro, lo cual contribuirá en el flujo de trabajo propuesto en este trabajo de investigación.

En el futuro es más probable que el Machine Learning sea más priorizado en las diferentes ramas y áreas de la geología, el cual se recomienda aprender sobre ello teniendo en cuenta la aplicación y el conocimiento de los lenguajes de programación como el Python.

Realizar un estudio geológico más profundo a mayor escala para la verificación total de los mapas predictivos cerca de la zona de estudio para así poder llegar a una inversión en ese sector.

Se propone actualizarse a los cambios tecnológicos en la geología puesto que el Machine Learning no será capaz de reemplazar a los geólogos profesionales sino a optimizar los futuros estudios conllevando a mejores profesionales al mercado, ya que el criterio humano estará más arriba de cualquier modelo de Machine Learning.

REFERENCIA BIBLIOGRAFICA

- Agencia Nacional de Minería. (s.f.). Glosario Minero. Recuperado de <https://www.anm.gov.co/sites/default/files/DocumentosAnm/glosariominero.pdf>
- Aldás Núñez, R. J., Tuz Chamorro, K. V., Vega Ocaña, J. A., Velasco Haro, M. S., & Mejía Escobar, C. I. (2022). Delimitación automática de ceniza volcánica en imágenes satelitales mediante Deep Learning. FIGEMPA: Investigación y Desarrollo, 13(1), 48-58. <https://doi.org/10.29166/revfig.v13i1.3121>.
- Andrade, J. A. C., & Celis, J. M. C. (2021). Modelo geometalúrgico de la caracterización de arcillas para la eficiencia de la flotación usando la metodología de machine learning. Revista del Instituto de investigación de la Facultad de minas, metalurgia y ciencias geográficas, 24(48), 55-68.
- Barlow, H. B. (1989). Unsupervised learning. Neural computation, 1(3), 295-311.
- Bazan Santa Cruz, O. (2021). Innovaciones estadísticas y redes neuronales artificiales aplicadas a la prospección geoquímica en los cuadrángulos de Chaparra y Atico, Caravelí – Arequipa. Revista Del Instituto De Investigación De La Facultad De Minas, Metalurgia Y Ciencias Geográficas, 24(48), 35-44. <https://doi.org/10.15381/iigeo.v24i48.19212>.
- Brandmeier, M., Cabrera Zamora, IG, Nykänen, V. y Middleton, M. (2020). Impulso al modelado de prospectividad mineral: una nueva caja de herramientas SIG. Investigación de recursos naturales, 29 (1), 71-88. <https://doi.org/10.1007/s11053-019-09483-8>.
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Campos, G. y Sosa, V. (2011). Estrategias metodológicas para la elaboración de tesis de posgrado. México, Porrúa.

Carrasco Pardo, D. (2021). La inteligencia artificial en la investigación minera. ResearchGate.
https://www.researchgate.net/publication/353998343_LA_INTELIGENCIA_ARTIFICIAL_EN_LA_INVESTIGACION_MINERA.

Castillo Liberato, O. A. (2022). Desarrollo de modelos predictivos de regresión en la industria minera mediante el uso de algoritmo de machine learning. Cybertesis repositorio de tesis digitales.
<https://cybertesis.unmsm.edu.pe/handle/20.500.12672/18458>.

Cerda, R., & Cifuentes, L. (2012). Uso de Curvas ROC en Investigación Clínica: Aspectos Teórico-Prácticos. Revista Chilena de Infectología, 29(2), 138-141.

De la Cruz, J.; Valencia, M. & Boulanger, E. (1996) - Geología de los cuadrángulos de Aguaytía, Panao y Pozuzo. Hoja: 19-I, 20-I, 21-I. INGEMMET, Boletín, Serie A: Carta Geológica Nacional, 80, 134 p. <https://hdl.handle.net/20.500.12544/36>.

Geología Viva. (2021). Muestreo de Minerales. Recuperado de <https://www.geologiaviva.info/wp-content/uploads/2021/07/Muestreo-de-Minerales.pdf>.

Goldschmidt, V. M. (1954). Chemistry. Oxford, England: Oxford University Press.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. <https://www.deeplearningbook.org/>.

- Hastie, T., Tibshirani, R. & Friedman, J.H. (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd edn, Springer Series in Statistics, Springer, New York, p. 533.
- Hidalgo Gajardo, A. E. (2023). Métodos de machine learning aplicados a la química mineral de cloritas para su uso en exploración de pórfidos cupríferos. Repositorio bibliotecas UdeC. <http://repositorio.udec.cl/jspui/handle/11594/11463>.
- Kohavi, R., & Provost, F. (1998). Glossary of Terms. Machine Learning Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, 30, 271-274.
- Kuncheva, L. (2004). Combining Pattern Classifiers: Methods and Algorithms, Wiley, Hoboken, New Jersey, p. 376. 'Learning to recognize volcanoes on Venus', Machine Learning, vol. 30, no. 2, pp. 165-194.
- Lasaga, A.C., & Ohmoto, H. (2002). The geochemical oxygen cycle: dynamics and stability. *Geochimica et Cosmochimica Acta*, 66(3), 361-381. [https://doi.org/10.1016/s0016-7037\(01\)00685-8](https://doi.org/10.1016/s0016-7037(01)00685-8).
- Macheyeki, A. S., Li, X., Kafumu, D. P. & Yuan, F. (2020). Applied geochemistry: advances in mineral exploration techniques. Cambridge: Elsevier.
- Mamani Sacsi, J. (2022). Aplicación de redes neuronales como algoritmos de machine learning para la predicción de recuperaciones auríferas mediante cianuración por agitación en Yauca – Caravelí. Repositorio Institucional – Universidad Nacional de San Agustín de Arequipa. <http://hdl.handle.net/20.500.12773/15482>.
- Mantilla Dulcey, A., (2023). Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático. ResearchGate.

https://www.researchgate.net/publication/370126251_Prediccion_de_la_ocurrencia_de_depositos_minerales_tipo_porfido_usando_tecnicas_de_aprendizaje_automatico.

Martín, E. (28 de noviembre de 2017). Por qué "machine learning" será la tecnología más importante en 2018. El País.

Melgarejo, Joan Carles, Proenza, Joaquín A., Galí, Salvador, & Llovet, Xavier. (2010). Técnicas de caracterización mineral y su aplicación en exploración y explotación minera. Boletín de la Sociedad Geológica Mexicana, 62(1), 1-23. Recuperado en 25 de julio de 2024, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-33222010000100002&lng=es&tlng=es.

Morales Madariaga, O. A. (2023). Aplicación de métodos de Machine Learning para la exploración geoquímica de tierras raras en yacimientos tipo IOGG, distritos Candelaria -- Punta del Cobre y Mantoverde, región de Atacama, Chile. Repositorio Académico UDA. <https://repositorioacademico.uda.cl/handle/20.500.12740/16763>.

Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. International Journal of Morphology, 35(1), 227-232. <https://dx.doi.org/10.4067/S0717-95022017000100037>.

Rodríguez, R., Coba, L., & Mamani, M. (2021). El machine learning de la Carta Geológica Nacional 50K y la exploración minera. In Congreso Internacional de Prospectores y Exploradores, 12, proExplo 2021: Recursos minerales para un futuro sostenible. Compendio (pp. 123-127). Lima: Instituto de Ingenieros de Minas del Perú.

- Rodríguez-Galiano, VF, Chica-Olmo, M. y Chica-Rivas, M. (2014). Modelado predictivo del potencial aurífero con la integración de información multifuente basada en bosque aleatorio: un estudio de caso en la zona de Rodalquilar, sur de España. *Revista internacional de ciencia de la información geográfica*, 28 (7), 1336–1354. <https://doi.org/10.1080/13658816.2014.885527>.
- Rudnick, R.L., & Gao, S. (2003). The Composition of the Continental Crust. In H.D. Holland & K.K. Turekian (Eds.), *Treatise on Geochemistry*, Vol. 3, The Crust (pp. 1-64). Elsevier-Pergamon. <https://doi.org/10.1016/b0-08-043751-6/03016-4>.
- Sánchez Muñoz, José. (2016). Análisis de calidad cartográfica mediante el estudio de la matriz de confusión. *Revista: Pensamiento Matemático*. <https://dialnet.unirioja.es/servlet/articulo?codigo=5998855>.
- Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15.
- Tar buck, E. J., & Lutgens, F. K. (2016). *Ciencias de la Tierra* (8va ed.). Recuperado de <https://www.xeologosdelmundu.org/wp-content/uploads/2016/03/TARBUCK-y-LUTGENS-Ciencias-de-la-Tierra-8va-ed.-1.pdf>.
- Tumialán De la Cruz, P.H. (2003). *Compendio de yacimientos minerales del Perú*. Sistema de Bibliotecas. Recuperado de <https://sisbib.unmsm.edu.pe/Bibvirtual/Libros/2007/cyacimientos/ficha.htm>.
- Turekian, K. K., & Wedepohl, K. H. (1961). Distribution of the Elements in Some Major Units. *Geological Society of America Bulletin*, 72(2), 175-192. Recuperado de

<https://pubs.geoscienceworld.org/gsa/gsabulletin/article/72/2/175/5262/Distribution-of-the-Elements-in-Some-Major-Units>.

VALLS, R. (2019). Nueva Clasificación Geoquímica de los Elementos. ResearchGate.
https://www.researchgate.net/publication/336240885_Nueva_Clasificacion_Geoquimica_de_los_Elementos.

ANEXOS

ANEXO I. INSTRUMENTOS DE CAMPO

BASE DE DATOS DE LA GEOQUIMICA CRUDA

ID	Este (X)	Norte (Y)	Au(ppb)	Au(ppm)	Al(%)	Br(ppm)	Ba(ppm)	Ca(ppm)	Cr(ppm)	Cu(ppm)	Fe(%)	K(%)	Li(ppm)	Mg(%)	Mn(ppm)	Na(%)	P(ppm)	Sr(ppm)	Ti(%)	V(ppm)	Zn(ppm)	Zr(ppm)	Ag(ppm)	As(ppm)
M1	402253.350	8884024.645	<10	N.R.	1.65	N.R.	N.R.	0.39	93	16.4	2.78	0.33	N.R.	0.59	637	0.06	0.04	22.9	0.11	50	48	0.9	<0.2	23
M2	407515.581	8869132.883	<10	N.R.	2.24	N.R.	N.R.	1.44	108	14.6	5.18	0.25	N.R.	1.05	770	0.13	0.09	82.4	0.29	145	63.3	5.3	<0.2	12
M3	413696.169	8864713.328	<10	N.R.	3.78	N.R.	N.R.	0.83	84	33.2	4.29	0.7	N.R.	1.24	738	0.04	0.06	39.1	0.18	110	83	4.4	0.2	29
M4	412784.847	8890269.454	<10	N.R.	1.24	N.R.	N.R.	0.45	92	15.4	2.55	0.19	N.R.	0.63	549	0.05	0.06	16.3	0.18	39	48.8	1.1	<0.2	20
M5	421102.033	8871325.339	10	N.R.	1.46	N.R.	N.R.	0.81	48	12.4	2.77	0.09	N.R.	0.68	486	0.03	0.07	32.5	0.21	45	52.9	3.4	<0.2	<3
M6	401625.499	8892425.714	<10	N.R.	2.96	N.R.	N.R.	0.12	132	29.5	4.28	0.79	N.R.	0.74	847	0.02	0.05	12.1	0.14	70	81	1.6	<0.2	5
M7	404711.607	8869133.173	17	N.R.	2.57	N.R.	N.R.	1.24	88	13.1	3.51	0.28	N.R.	1.13	621	0.11	0.09	77.6	0.24	90	60.5	3.8	<0.2	7
M8	408203.774	8865212.509	<10	N.R.	1.2	N.R.	N.R.	0.19	56	12.6	2.1	0.34	N.R.	0.27	294	0.02	0.03	12.1	0.08	54	33.5	1.7	<0.2	17
M9	409053.170	8866663.271	10	N.R.	1.2	N.R.	N.R.	0.1	67	11.7	1.61	0.35	N.R.	0.21	255	0.01	0.02	8.6	0.05	27	32.2	1	<0.2	28
M10	412066.665	8887272.237	<10	N.R.	1.02	N.R.	N.R.	0.42	60	11	2.02	0.13	N.R.	0.53	396	0.04	0.06	15.2	0.15	32	42	0.8	<0.2	24
M11	428998.280	8891697.149	<10	N.R.	1.42	N.R.	N.R.	1.14	38	28.2	3.39	0.23	N.R.	0.93	518	0.05	0.05	52.5	0.13	61	69	4.9	<0.2	6
M12	403106.310	8858884.662	283	N.R.	1.6	N.R.	N.R.	0.38	89	16.4	3.17	0.29	N.R.	0.54	652	0.04	0.04	21.5	0.14	64	45.7	0.9	<0.2	13
M13	407155.452	8863642.876	10	N.R.	2.46	N.R.	N.R.	0.78	79	20.2	3.89	0.33	N.R.	0.86	554	0.05	0.08	37.8	0.14	114	62.7	1.6	<0.2	31
M14	404716.733	8893503.281	10	N.R.	1.98	N.R.	N.R.	0.18	72	19.4	2.88	0.53	N.R.	0.52	589	0.02	0.04	14.7	0.1	44	68.7	1.1	<0.2	12
M15	422014.083	8869998.266	<10	N.R.	2.1	N.R.	N.R.	0.8	57	21.6	3.62	0.23	N.R.	0.89	496	0.07	0.06	31.7	0.23	99	59.5	4	0.5	12
M16	423276.688	8877210.394	83	N.R.	1.64	N.R.	N.R.	0.4	38	17.8	2.66	0.2	N.R.	1.21	518	0.06	0.03	21.1	0.13	40	75.1	3.2	<0.2	6
M17	423537.948	8873401.719	<10	N.R.	1.9	N.R.	N.R.	0.36	43	38.1	3.81	0.13	N.R.	1.51	608	0.03	0.04	20.1	0.11	63	86.9	2.5	0.4	16
M18	422018.662	8886423.909	<10	N.R.	1.19	N.R.	N.R.	0.62	43	12	2.29	0.16	N.R.	0.68	441	0.06	0.04	35.8	0.11	30	46.5	2	0.2	3
M19	424723.524	8882038.864	10	N.R.	2.04	N.R.	N.R.	0.53	29	32.6	5.89	0.29	N.R.	1.78	954	0.05	0.05	52.7	0.24	114	140	6.9	0.4	20
M20	402715.445	8868522.869	<10	N.R.	1.25	N.R.	N.R.	0.89	71	12.3	3.37	0.14	N.R.	0.43	667	0.07	0.05	66.9	0.22	81	44.6	3.2	<0.2	12
M21	408004.081	8866613.001	<10	N.R.	1.78	N.R.	N.R.	0.4	100	18.2	2.46	0.18	N.R.	0.41	415	0.04	0.04	34.1	0.11	55	166	0.6	<0.2	17
M22	411453.330	8869502.343	<10	N.R.	1.37	N.R.	N.R.	0.9	79	14.5	4.6	0.16	N.R.	0.64	608	0.08	0.07	55.2	0.19	118	49.9	3.3	<0.2	14
M23	413218.069	8864963.160	<10	N.R.	3.02	N.R.	N.R.	0.46	65	28.6	3.76	0.62	N.R.	1.11	617	0.03	0.05	23.1	0.13	90	75.1	2.7	1.1	18
M24	412293.155	8866170.812	<10	N.R.	2.98	N.R.	N.R.	0.5	54	20.2	3.61	0.47	N.R.	0.72	657	0.03	0.04	25	0.25	97	72.4	1.4	<0.2	<3
M25	422992.811	8870712.503	17	N.R.	1.8	N.R.	N.R.	0.7	67	18.6	5.13	0.19	N.R.	0.72	520	0.06	0.06	26.3	0.28	176	62.3	4.5	<0.2	4
M26	401306.361	8867147.781	17	N.R.	2.55	N.R.	N.R.	0.6	139	21.6	6.44	0.43	N.R.	0.96	838	0.05	0.07	36.3	0.15	171	83.1	1.8	<0.2	21
M27	403014.679	8888164.517	<10	N.R.	1.54	N.R.	N.R.	0.32	79	11.9	11.24	0.29	N.R.	0.39	1257	0.02	0.06	31.2	0.29	225	70.4	10	<0.2	9
M28	409493.649	8870002.714	10	N.R.	1.05	N.R.	N.R.	0.82	55	10.1	3.62	0.11	N.R.	0.36	651	0.06	0.05	59.7	0.22	92	40.1	3.6	<0.2	14
M29	410922.127	8884374.558	<10	N.R.	1.24	N.R.	N.R.	0.47	80	16.7	2.38	0.13	N.R.	0.56	510	0.05	0.07	15.7	0.18	38	49.7	1	<0.2	42
M30	424754.435	8880887.871	10	N.R.	1.72	N.R.	N.R.	0.62	51	22.8	5.2	0.14	N.R.	1.04	566	0.04	0.05	41.6	0.25	103	64.5	4.1	0.3	7
M31	426297.944	8885781.684	50	N.R.	2.45	N.R.	N.R.	1.26	37	27.6	5.99	0.14	N.R.	2.3	980	0.04	0.05	135.5	0.43	184	114	11.6	0.7	18
M32	428144.415	8889014.110	<10	N.R.	1.86	N.R.	N.R.	0.83	45	35.3	3.81	0.17	N.R.	1.36	745	0.03	0.07	39.9	0.23	63	90.9	4.8	0.3	<3
M33	401551.671	8867942.423	<10	N.R.	1.19	N.R.	N.R.	0.83	46	10.4	3.17	0.12	N.R.	0.41	619	0.06	0.05	62.3	0.21	77	42.2	2.5	<0.2	22
M34	416106.037	8876024.336	10	N.R.	1.46	N.R.	N.R.	0.51	75	9.7	2.95	0.09	N.R.	0.62	613	0.03	0.06	23	0.17	48	46.6	1	0.3	27
M35	419012.402	8875525.752	<10	N.R.	1.03	N.R.	N.R.	0.39	73	7.1	2.28	0.08	N.R.	0.39	526	0.03	0.05	19.6	0.12	31	37.9	0.6	<0.2	<3
M36	423680.667	8875291.718	17	N.R.	2.28	N.R.	N.R.	0.39	43	33.6	4.74	0.13	N.R.	1.83	725	0.02	0.08	17.5	0.13	58	79.7	3.3	0.4	10
M37	424891.171	8886213.314	25	N.R.	1.18	N.R.	N.R.	0.69	39	11.4	2.57	0.15	N.R.	0.75	455	0.05	0.04	36.4	0.12	37	50.4	2.3	0.4	<3
M38	426325.045	8887526.560	<10	N.R.	2.87	N.R.	N.R.	1	48	40.3	5.08	0.15	N.R.	3.27	965	0.02	0.07	59.7	0.42	140	112	5.4	0.5	<3
M39	407335.881	8865823.940	<10	N.R.	1.57	N.R.	N.R.	1	72	11.9	4.12	0.18	N.R.	0.73	641	0.08	0.07	63.7	0.2	109	50.3	2.8	<0.2	9
M40	410315.825	8869731.686	<10	N.R.	1.57	N.R.	N.R.	1.08	95	13.8	4.82	0.18	N.R.	0.68	708	0.09	0.08	71.6	0.22	128	54.7	3.8	<0.2	20
M41	418034.865	8870004.746	10	N.R.	2.79	N.R.	N.R.	0.75	28	25.6	4.85	0.29	N.R.	0.79	751	0.05	0.06	34.5	0.22	139	68.3	2.4	0.4	<3
M42	423513.696	8876015.612	<10	N.R.	1.06	N.R.	N.R.	0.33	92	9.6	1.8	0.1	N.R.	0.26	385	0.03	0.04	19.8	0.09	26	32.7	<0.5	<0.2	<3
M43	401756.815	8882083.751	<10	N.R.	1.49	N.R.	N.R.	0.44	100	14.9	3.14	0.33	N.R.	0.57	669	0.06	0.04	24.5	0.14	65	47.6	1.5	<0.2	13
M44	416012.583	8869693.708	<10	N.R.	2.63	N.R.	N.R.	0.83	32	27.6	5.27	0.28	N.R.	0.76	679	0.05	0.06	36.5	0.24	151	73.7	3	<0.2	10
M45	421305.600	8866482.659	17	N.R.	2.71	N.R.	N.R.	0.45	52	19.2	3.37	0.43	N.R.	0.66	608	0.03	0.04	23.4	0.24	91	66.3	1.3	<0.2	<3
M46	415265.171	8880477.445	4550	4.55	0.84	N.R.	13	3400	120	2211	15	0.2	24	0.07	215	0.33	236	17.4	0.03	26	422	63.7	10	4432
M47	415371.933	8879539.938	3000	3	4.03	N.R.	316	30500	140	1061	11.91	1.74	29	0.51	1865	0.55	394	113.6	0.19	59	162	47.2	10	2094
M48	414234.247	8879953.642	10660	10.66	0.77	N.R.	92	11000	134	6192	15	0.29	33	0.16	9936	0.26	165	104.9	0.03	12	352	26.9	10	N.R.
M49	414521.171	8879503.238	3000	3	1.18	N.R.	19	3600	163	2129	15	0.4	29	0.11	215	0.16	352	13.8	0.06	32	350	37.8	10	2638
M50	414755.864	8879066.474	3000	3	5.2	N.R.	407	36100	145	752.1	8.56	2.45	33	0.58	2101	0.71	334	142.3	0.25	63	117	44.7	10	1099

Be(ppm)	Bi(ppm)	Cd(ppm)	Cs(ppm)	Co(ppm)	Ce(ppm)	Ga(ppm)	Ge(ppm)	Hf(ppm)	Hg(ppm)	In(ppm)	La(ppm)	Lu(ppm)	Mo(ppm)	Nb(ppm)	Ni(ppm)	Pb(ppm)	Rd(ppm)	Sb(ppm)	Sc(ppm)	Se(ppm)	Sn(ppm)	Ta(ppm)
0.6	<5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	30.4	N.R	6	N.R	46	9	N.R	<5	5	N.R	<10	N.R
0.5	<5	<1	N.R	9	N.R	N.R	N.R	N.R	<0.05	N.R	33	N.R	4	N.R	37	<2	N.R	<5	10.6	N.R	<10	N.R
1.1	<5	<1	N.R	15	N.R	N.R	N.R	N.R	<0.05	N.R	35.6	N.R	4	N.R	40	10	N.R	<5	12.8	N.R	<10	N.R
<0.5	<5	<1	N.R	11	N.R	N.R	N.R	N.R	<0.05	N.R	12.9	N.R	5	N.R	23	5	N.R	<5	3.6	N.R	<10	N.R
0.5	8	<1	N.R	10	N.R	N.R	N.R	N.R	<0.05	N.R	7.6	N.R	3	N.R	8	<2	N.R	<5	6.6	N.R	<10	N.R
1.2	<5	<1	N.R	14	N.R	N.R	N.R	N.R	<0.05	N.R	19.2	N.R	5	N.R	69	21	N.R	<5	8.5	N.R	<10	N.R
0.5	<5	<1	N.R	9	N.R	N.R	N.R	N.R	<0.05	N.R	22.4	N.R	3	N.R	22	8	N.R	<5	10.1	N.R	<10	N.R
0.9	<5	<1	N.R	5	N.R	N.R	N.R	N.R	<0.05	N.R	18.7	N.R	3	N.R	16	14	N.R	<5	2.4	N.R	<10	N.R
1.1	<5	<1	N.R	4	N.R	N.R	N.R	N.R	<0.05	N.R	16	N.R	3	N.R	34	14	N.R	<5	1.9	N.R	<10	N.R
<0.5	7	<1	N.R	10	N.R	N.R	N.R	N.R	<0.05	N.R	12	N.R	3	N.R	17	7	N.R	<5	2.9	N.R	<10	N.R
0.6	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	19.7	N.R	3	N.R	12	11	N.R	13	4.7	N.R	<10	N.R
0.6	<5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	49.7	N.R	5	N.R	25	16	N.R	<5	5.2	N.R	<10	N.R
0.9	<5	<1	N.R	13	N.R	N.R	N.R	N.R	<0.05	N.R	18.2	N.R	3	N.R	30	19	N.R	<5	6.4	N.R	<10	N.R
0.8	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	26.7	N.R	5	N.R	26	15	N.R	<5	4.4	N.R	<10	N.R
0.6	<5	<1	N.R	13	N.R	N.R	N.R	N.R	<0.05	N.R	13.4	N.R	1	N.R	17	<2	N.R	14	6.8	N.R	<10	N.R
0.6	<5	<1	N.R	9	N.R	N.R	N.R	N.R	<0.05	N.R	10.8	N.R	4	N.R	20	21	N.R	<5	4.7	N.R	<10	N.R
<0.5	<5	<1	N.R	13	N.R	N.R	N.R	N.R	<0.05	N.R	7.9	N.R	3	N.R	12	7	N.R	16	5.4	N.R	12	N.R
0.5	<5	<1	N.R	6	N.R	N.R	N.R	N.R	<0.05	N.R	13.7	N.R	3	N.R	21	13	N.R	8	3.8	N.R	<10	N.R
<0.5	<5	<1	N.R	14	N.R	N.R	N.R	N.R	<0.05	N.R	10.7	N.R	2	N.R	14	8	N.R	17	7.6	N.R	11	N.R
<0.5	<5	<1	N.R	5	N.R	N.R	N.R	N.R	<0.05	N.R	70.9	N.R	6	N.R	33	14	N.R	<5	4.8	N.R	<10	N.R
0.6	<5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	28.1	N.R	5	N.R	42	29	N.R	<5	3.5	N.R	<10	N.R
<0.5	<5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	45.6	N.R	5	N.R	20	7	N.R	<5	5.3	N.R	<10	N.R
1	<5	<1	N.R	14	N.R	N.R	N.R	N.R	<0.05	N.R	18	N.R	2	N.R	23	16	N.R	<5	9.2	N.R	<10	N.R
0.8	<5	<1	N.R	13	N.R	N.R	N.R	N.R	<0.05	N.R	11.2	N.R	3	N.R	13	18	N.R	<5	9.4	N.R	<10	N.R
0.5	<5	<1	N.R	12	N.R	N.R	N.R	N.R	<0.05	N.R	24.9	N.R	2	N.R	14	14	N.R	9	5.9	N.R	<10	N.R
0.9	<5	<1	N.R	13	N.R	N.R	N.R	N.R	<0.05	N.R	45.2	N.R	3	N.R	30	21	N.R	<5	7.8	N.R	<10	N.R
0.7	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	76.7	N.R	2	N.R	31	18	N.R	<5	6.5	N.R	<10	N.R
<0.5	<5	<1	N.R	4	N.R	N.R	N.R	N.R	<0.05	N.R	75.3	N.R	4	N.R	13	<2	N.R	<5	4.5	N.R	<10	N.R
0.5	<5	<1	N.R	11	N.R	N.R	N.R	N.R	0.05	N.R	14.7	N.R	3	N.R	21	12	N.R	<5	3.5	N.R	<10	N.R
0.6	<5	<1	N.R	17	N.R	N.R	N.R	N.R	<0.05	N.R	15	N.R	2	N.R	13	4	N.R	8	6	N.R	22	N.R
<0.5	<5	<1	N.R	20	N.R	N.R	N.R	N.R	<0.05	N.R	3	N.R	<1	N.R	18	8	N.R	<5	11.5	N.R	<10	N.R
0.6	<5	<1	N.R	12	N.R	N.R	N.R	N.R	<0.05	N.R	14.2	N.R	4	N.R	19	13	N.R	12	5.6	N.R	<10	N.R
<0.5	<5	<1	N.R	5	N.R	N.R	N.R	N.R	<0.05	N.R	67.5	N.R	5	N.R	12	9	N.R	<5	4.5	N.R	<10	N.R
<0.5	<5	<1	N.R	10	N.R	N.R	N.R	N.R	<0.05	N.R	13.4	N.R	2	N.R	17	<2	N.R	<5	3.7	N.R	14	N.R
<0.5	5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	12.6	N.R	3	N.R	12	<2	N.R	7	3.2	N.R	<10	N.R
0.7	<5	<1	N.R	23	N.R	N.R	N.R	N.R	<0.05	N.R	7.7	N.R	3	N.R	15	28	N.R	10	5.8	N.R	<10	N.R
0.5	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	17.3	N.R	2	N.R	7	27	N.R	8	3.9	N.R	<10	N.R
0.6	<5	<1	N.R	23	N.R	N.R	N.R	N.R	<0.05	N.R	4.1	N.R	3	N.R	27	<2	N.R	9	10.1	N.R	<10	N.R
<0.5	<5	<1	N.R	7	N.R	N.R	N.R	N.R	<0.05	N.R	41.4	N.R	5	N.R	18	7	N.R	<5	6.4	N.R	<10	N.R
<0.5	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	52.7	N.R	6	N.R	38	20	N.R	<5	6.5	N.R	<10	N.R
0.9	<5	<1	N.R	17	N.R	N.R	N.R	N.R	<0.05	N.R	9.1	N.R	1	N.R	8	<2	N.R	6	11.6	N.R	<10	N.R
<0.5	12	<1	N.R	6	N.R	N.R	N.R	N.R	<0.05	N.R	19.3	N.R	2	N.R	16	4	N.R	<5	3	N.R	<10	N.R
0.5	<5	<1	N.R	8	N.R	N.R	N.R	N.R	<0.05	N.R	40.2	N.R	3	N.R	44	7	N.R	<5	5.4	N.R	<10	N.R
0.8	<5	<1	N.R	17	N.R	N.R	N.R	N.R	<0.05	N.R	9.4	N.R	<1	N.R	9	4	N.R	<5	11.9	N.R	<10	N.R
0.8	<5	<1	N.R	11	N.R	N.R	N.R	N.R	<0.05	N.R	10.4	N.R	1	N.R	12	3	N.R	<5	8.8	N.R	<10	N.R
0.3	14	1.19	10.8	5.5	5.89	7.2	2.5	1.7	N.R	2.08	5.7	0.09	23.08	1.4	7.8	1335	15.8	110.4	1.2	2	8.5	0.63
1.3	53.9	0.41	41.63	45	14.8	11.7	1.2	1.43	N.R	0.94	23.7	0.25	24.19	7.1	44.9	781.4	75.1	133.4	6.2	11	5.9	1.23
0.4	1015	3.7	40.63	15.7	3.83	4	2.9	0.81	N.R	2.63	18.3	0.33	12.32	1.9	14.3	387.6	13.9	1026	2.4	3	2.8	0.38
0.6	16	0.72	9.74	3.4	9.37	8.7	2.1	1.2	N.R	1.69	5	0.07	16.05	2.3	6.6	843.7	32.4	72.4	1.5	2	6.4	0.46
1.3	35.5	0.4	46.26	31.1	15.9	13.1	0.6	1.56	N.R	0.66	26.7	0.26	13.07	8.9	39.5	449.6	98	70.3	7.9	5	5.9	0.79

Te(ppm)	Th(ppm)	Tl(ppm)	U(ppm)	W(ppm)	Y(ppm)
N.R	N.R	N.R	N.R	<10	13.5
N.R	N.R	N.R	N.R	<10	27.3
N.R	N.R	N.R	N.R	<10	13.6
N.R	N.R	N.R	N.R	<10	8.7
N.R	N.R	N.R	N.R	<10	7.1
N.R	N.R	N.R	N.R	<10	8.4
N.R	N.R	N.R	N.R	<10	21.4
N.R	N.R	N.R	N.R	14	5.1
N.R	N.R	N.R	N.R	13	4.2
N.R	N.R	N.R	N.R	<10	7.7
N.R	N.R	N.R	N.R	<10	8.3
N.R	N.R	N.R	N.R	<10	17.3
N.R	N.R	N.R	N.R	<10	9.5
N.R	N.R	N.R	N.R	14	11.4
N.R	N.R	N.R	N.R	<10	7.8
N.R	N.R	N.R	N.R	<10	9.8
N.R	N.R	N.R	N.R	<10	7.3
N.R	N.R	N.R	N.R	<10	9.9
N.R	N.R	N.R	N.R	<10	9.1
N.R	N.R	N.R	N.R	<10	37.2
N.R	N.R	N.R	N.R	11	10.9
N.R	N.R	N.R	N.R	<10	24.6
N.R	N.R	N.R	N.R	<10	9.1
N.R	N.R	N.R	N.R	<10	6.7
N.R	N.R	N.R	N.R	<10	9
N.R	N.R	N.R	N.R	<10	17.1
N.R	N.R	N.R	N.R	<10	20.5
N.R	N.R	N.R	N.R	<10	36
N.R	N.R	N.R	N.R	<10	9.5
N.R	N.R	N.R	N.R	<10	15.7
N.R	N.R	N.R	N.R	<10	10.6
N.R	N.R	N.R	N.R	<10	11
N.R	N.R	N.R	N.R	13	33.9
N.R	N.R	N.R	N.R	<10	7.7
N.R	N.R	N.R	N.R	<10	6.1
N.R	N.R	N.R	N.R	<10	8.8
N.R	N.R	N.R	N.R	<10	11
N.R	N.R	N.R	N.R	<10	12.8
N.R	N.R	N.R	N.R	<10	23.3
N.R	N.R	N.R	N.R	<10	28.1
N.R	N.R	N.R	N.R	<10	9.6
N.R	N.R	N.R	N.R	<10	6.9
N.R	N.R	N.R	N.R	<10	16.9
N.R	N.R	N.R	N.R	<10	9.2
N.R	N.R	N.R	N.R	<10	6.3
3.15	0.6	0.17	1.3	240.4	6.1
7.02	5.9	0.37	2.3	88.1	20.1
2.17	1	0.14	5.8	143.7	16.5
1.19	0.7	0.29	1.6	166.8	4.5
3.69	6.6	0.45	2.1	20.8	21.8

LEYENDA

DEPARTAMENTO Area de estudio

PROVINCIA

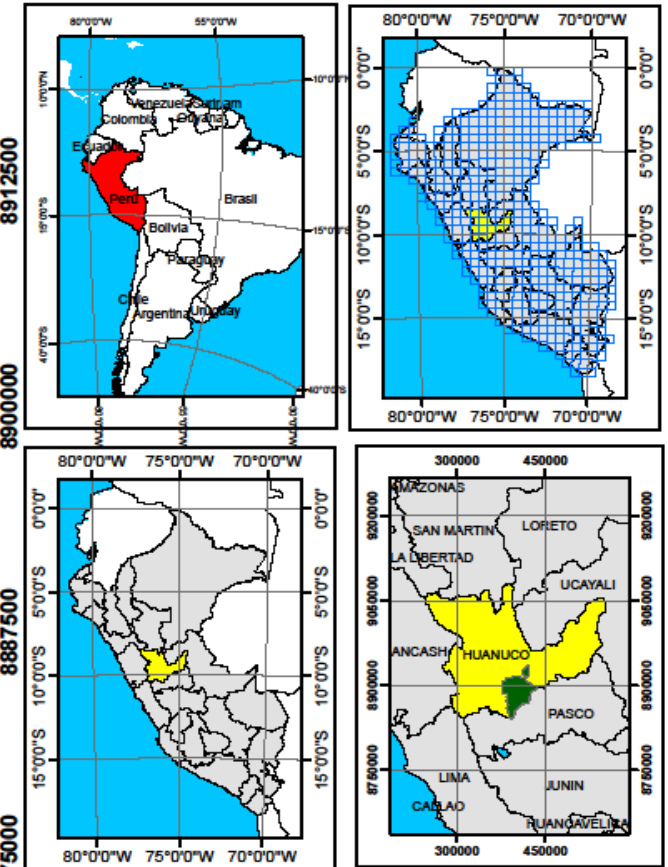
HUÁNUCO Vías de Acceso

PACHITEA

DISTRITO

PUNO

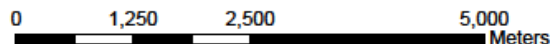
VERTICE	NORTE	ESTE
1	8,882,000	416,000
2	8,879,000	416,000
3	8,879,000	413,000
4	8,880,000	413,000
5	8,880,000	415,000
6	8,882,000	415,000



Acceso a la zona de estudio			
RUTA	VIA DE ACCESO	Via (Km)	Horas (Vehículo)
Lima-Huanuco	Asfaltada	358	7 h 41 min
Huánuco-Churabamba	Asfaltada	21.6	34 min
Churubamba-Panao	Trocha carrozable	42.30	1 h 19 min
Panao-área de estudio	Trocha carrozable	175.00	4 h 00 min
TOTAL	-	596.9	13 h 34 min



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN						
FACULTAD DE INGENIERÍA						
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA						
NOMBRE DEL PROYECTO:	UBICACIÓN:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huancayo 2023	DISTRITO: PANAO PROVINCIA: PACHITEA REGIÓN: HUANUCO	Bach. Gianmarco Castro Melendez REVISADO POR: Mg. Eder Guido ROBLES MORALES	PLANO DE UBICACIÓN	PROYECCIÓN: UTM ZONA: 18 SUR COORDENADAS: WGS - 84	1/250000 FECHA: 1/07/2024	01



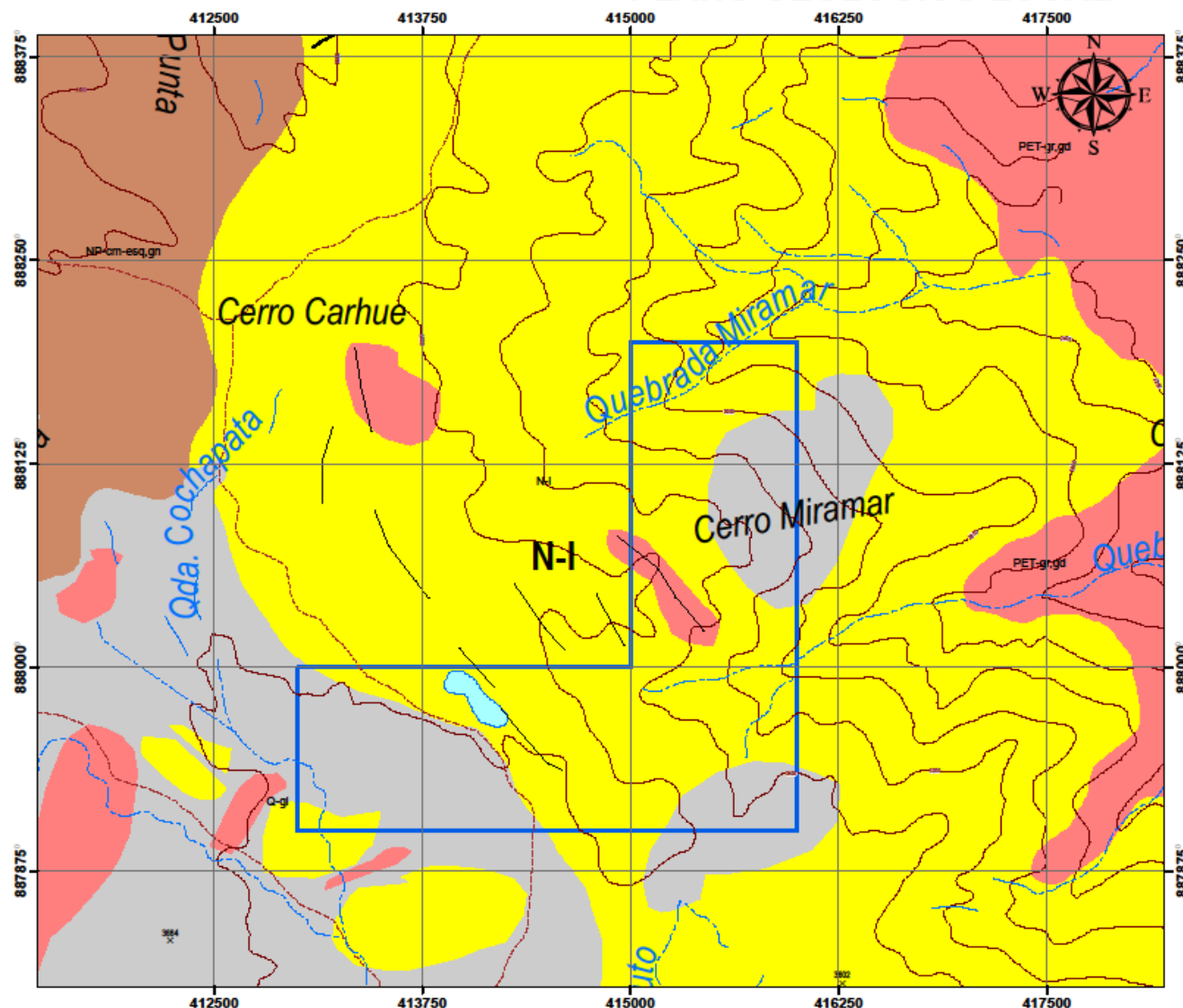
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERÍA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 3/223	Bach. Gianmarco Castro Melendez	PLANO GEOLOGICO REGIONAL	PROYECCIÓN: UTM	1/50000	02
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

LEYENDA		ROCAS ÍNTUSAS Y SUBVOLCÁNICAS	
UNIDADES LITOTRÁFMICAS			
Deposito aluvial	Qh-z	Depositos pedregosos con matriz arcillosa	
Deposito glacial, fangal	Q-gf	Depositos pedregosos, heterométricos subredondeados con matriz arcillosa a arenosa	
Deposito glacial	Q-g	Depositos pedregosos, clastos subangulares con matriz arenosa	
Panoraso Llanero	N-l	Conglomerado de matriz arenosa arenomolada formada por sedimentos llaneros que transcurrieron de conglomerado reciente	
Panoraso Ventanoso	P-y	Arenas y/o grutas, arcillas y/o grutas, arenas arcillosas, vagués y conglomerado de grutas	
Panoraso Vilán	Ks-y	Arenas con vagués llaneros a medianos, grutas y/o arenos, conglomerados de matriz arenosa con vagués arenos	P-p Pelto
Panoraso Chirico	Kis-y	Lutitas y/o arcillas, lutillas con lutitas de arcillas arenosas, conglomerados con lutillas y/o grutas a arenas arcillosas recientes	KD-qf Grutas
Dringo Chirico	Ks-c	Arenas arcillosas limpias laminadas con limonitas grises	
Panoraso Nueveojos	Js-c	Arenas heterométricas grises de grano medio a grueso, pedregos arcillosos de conglomerado y limonías grises	
Dringo Vilán	Js-c	Cajas que chispa y vacuolas en lutillas grises y arcillosas arenolitas	
Dringo Vilán	Js-c	Grutas arcillosas, arcillosas, lutinosas y moladas, intercaladas con lutillas grises y arenas y vagués	MP-gf Montañas, domos
Dringo Vilán	Js-c	Arenas heterométricas con limonías y conglomerado arcillosos, lutinos y arenas arcillosas y moladas	PE-qg Grutas, granitoides
Grupos Mestizo	N-cm-esslo	Grutas, grutas	PE-m-ql Grutas, granitoides
Grupos Mestizo	NP-gf-qg	Grutas	
Grupos Mestizo	NP-tu-qf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf	Grutas	
Grupos Mestizo	NP-gf		

SIMBOLOGIA

	Area de estudio		Rumbo y buzamiento de estratos fotointerpretado < 30°
×	Cota comprobada		Rumbo y buzamiento de foliación
⊙	Capital de departamento		Rumbo y buzamiento de estratos fotointerpretado 30° - 60°
⦿	Localidad común		Líneas de sección
	Rio principal		Líneas de sección completa
	Rio secundario		Flèches
	Quebrade		Falla normal
	Quebrade seca una parte del afo		Falla inferia
	Curva de Nivel Primaria		Falla inversa
	Carratera afirmada, transitable		Falla de rumbo dextral
	Camino de herradura		Eje de anticlinal
	Rumbo y buzamiento de estratos		Eje de sinclinal
			Contacto geológico
			Indicadores
			Laguna

PLANO GEOLOGICO LOCAL



LEYENDA

Geologia Local

UNIDADES LITOESTRATIGRAFICAS

- Deposito glaciar (Q-gl)
- Formación Lantorache (N-i)
- Granito de Paucartambo (PET-gr, gd)
- Complejo Maraion (NP-esq,gn)

SIMBOLOGIA

- Area de estudio
- Cota comprobada
- Falla normal
- Capital de departamento
- Localidad común
- Rio principal
- Rio secundario
- Quebrada
- Quebrada seca una parte del año
- Curva de Nivel Primaria
- Camino afirmado, transitable
- Centro de hemadura
- Rumbo y buzamiento de estratos
- Rumbo y buzamiento de estratos fotointerpretado < 30°
- Rumbo y buzamiento de foliación
- Rumbo y buzamiento de estratos fotointerpretado 30° - 60°
- Linea de sección
- Linea de sección completa
- Flachas
- Falla normal
- Falla inferida
- Falla inversa
- Falla de rumbo dextral
- Eje de anticlinal
- Eje de sincinal
- Laguna

0 625 1,250 2,500 Meters

1:25,000



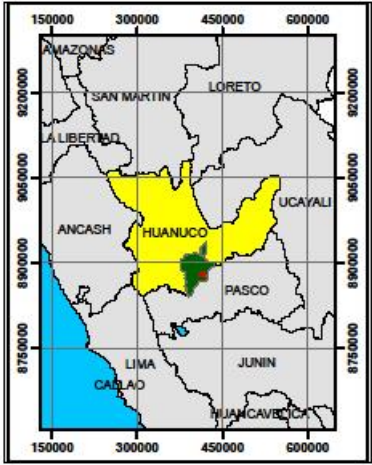
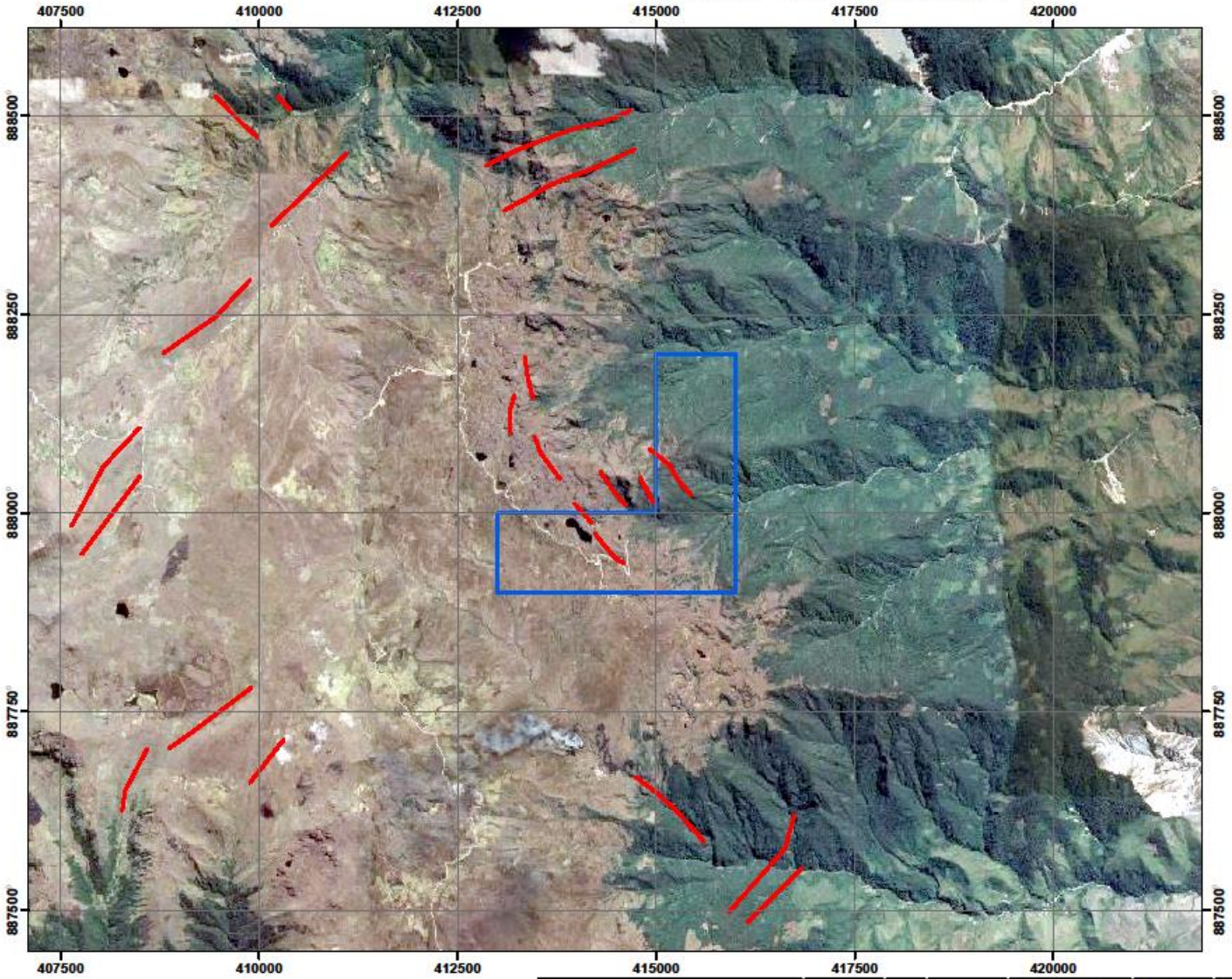
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA

NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOLOGICO LOCAL	PROYECCIÓN: UTM	1/25000	03
	Mg. Eder Guido ROBLES MORALES		ZONA: 18 SUR	FECHA: 2/08/2024	
			COORDENADAS: WGS - 84		

PLANO ESTRUCTURAL



LEYENDA

— Falla Normal

▭ Area_de_estudio

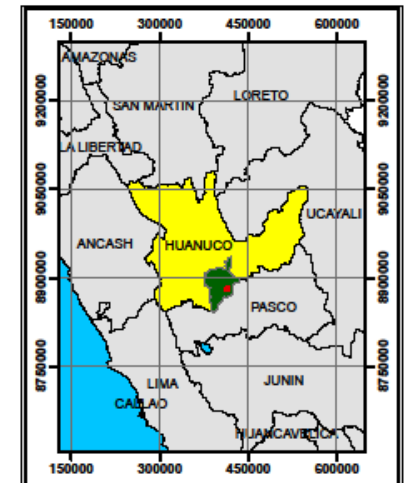
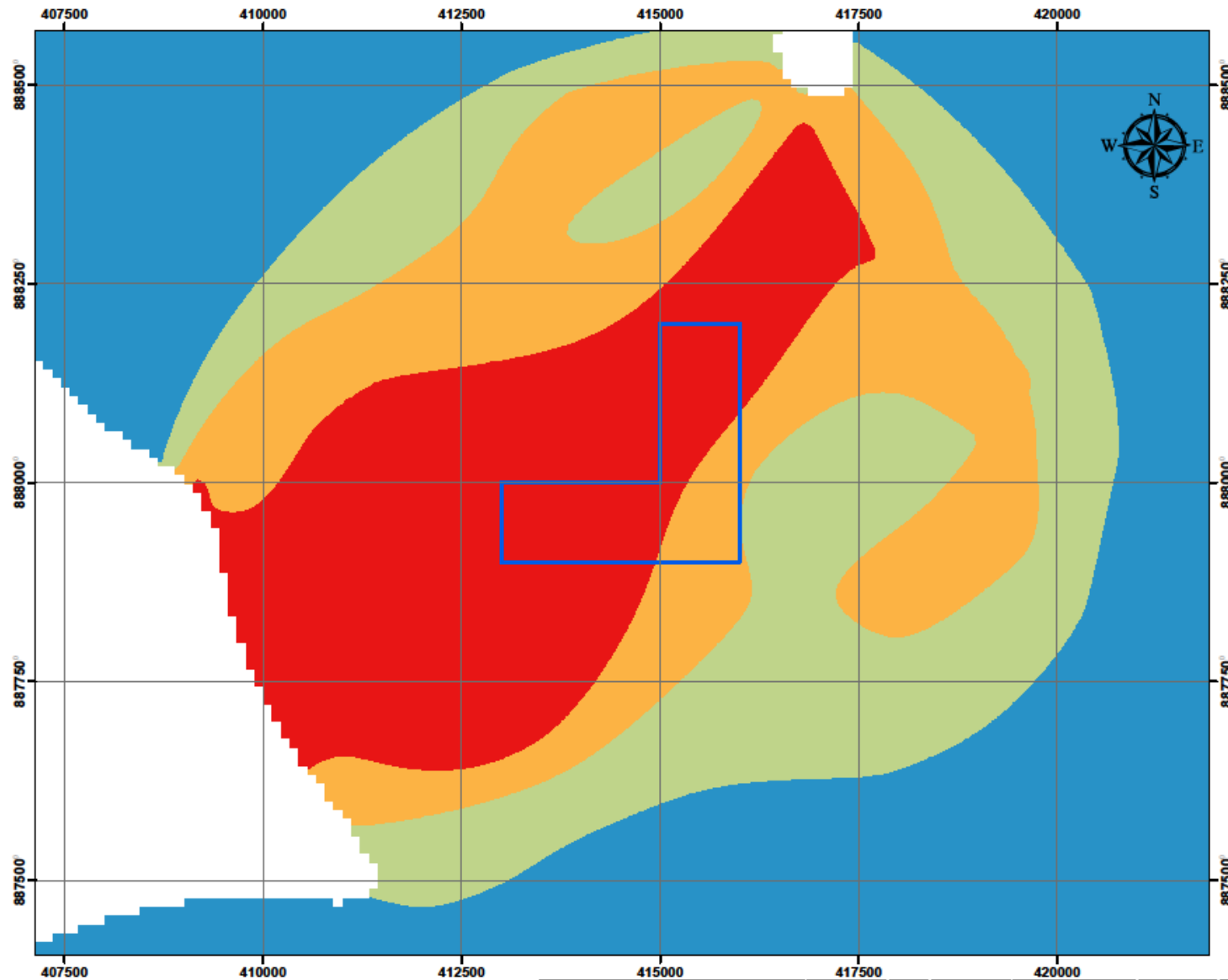


1:50,000



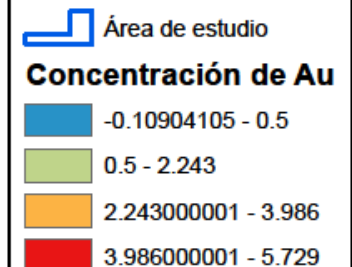
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huancayo 2023	Bach. Gianmarco Castro Melendez	PLANO ESTRUCTURAL	PROYECCIÓN: UTM	1/50000	04
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO GEOQUIMICO DE AU



Background	0.5
Treshold 1	2.243
Treshold 2	3.986
Treshold 3	5.729

LEYENDA



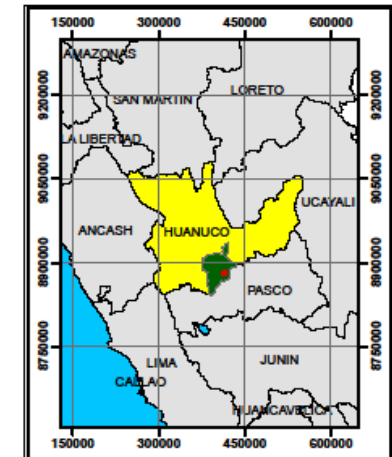
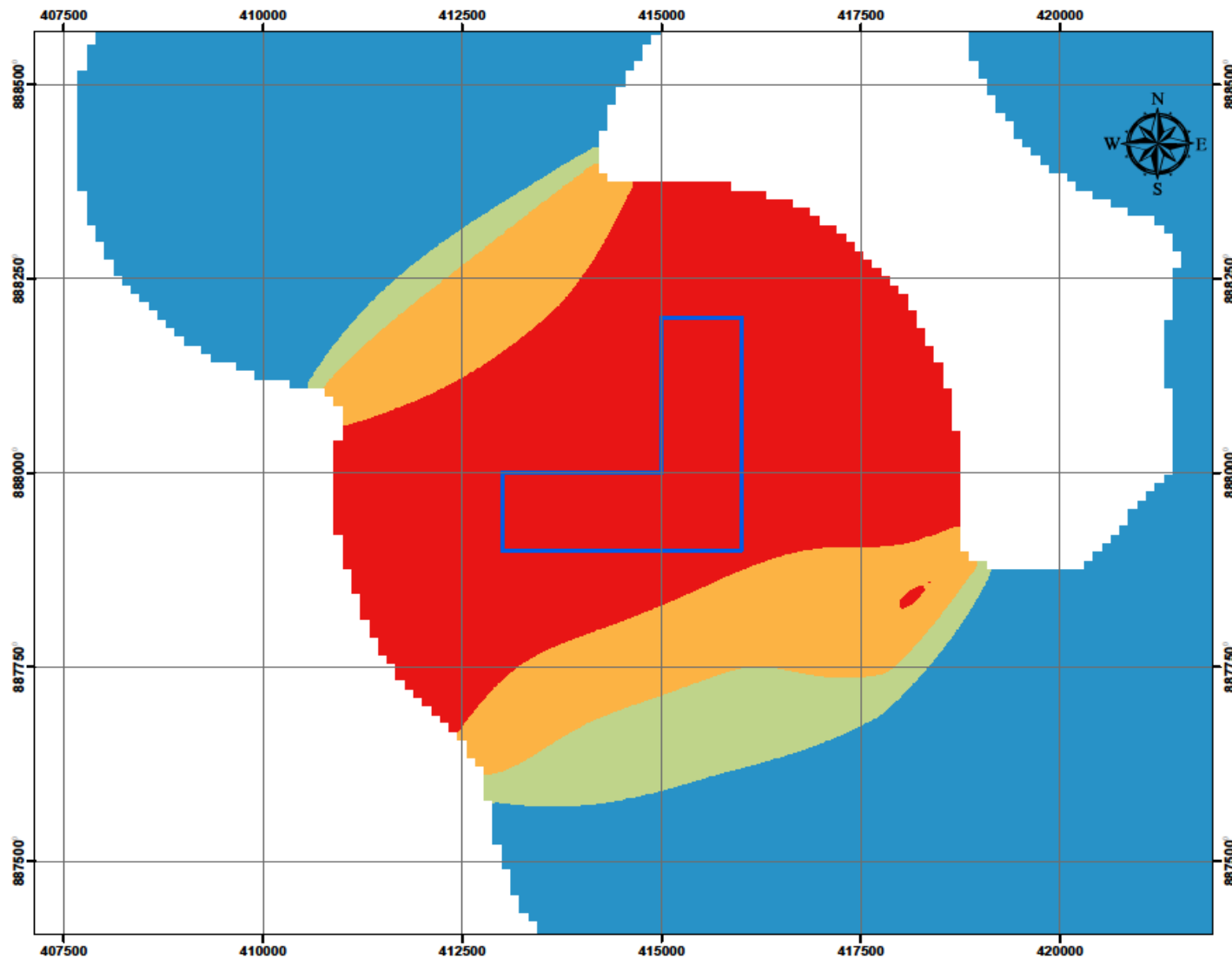
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DEL AU	PROYECCIÓN: UTM	1/50000	05
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO GEOQUIMICO DE AG



Background	1.203
Treshold 1	4.171
Treshold 2	7.139
Treshold 3	10.107

LEYENDA	
	Área de estudio
Concentración de Ag	
	-0.285414398 - 1.203
	1.203000001 - 4.171
	4.171000001 - 7.139
	7.139000001 - 10.107

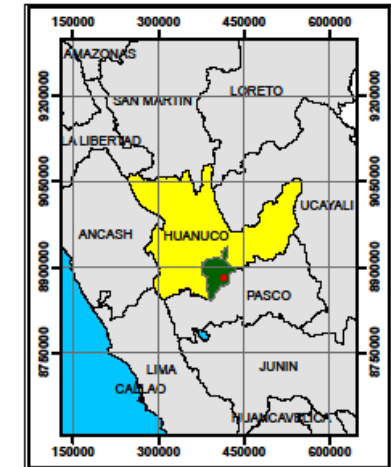
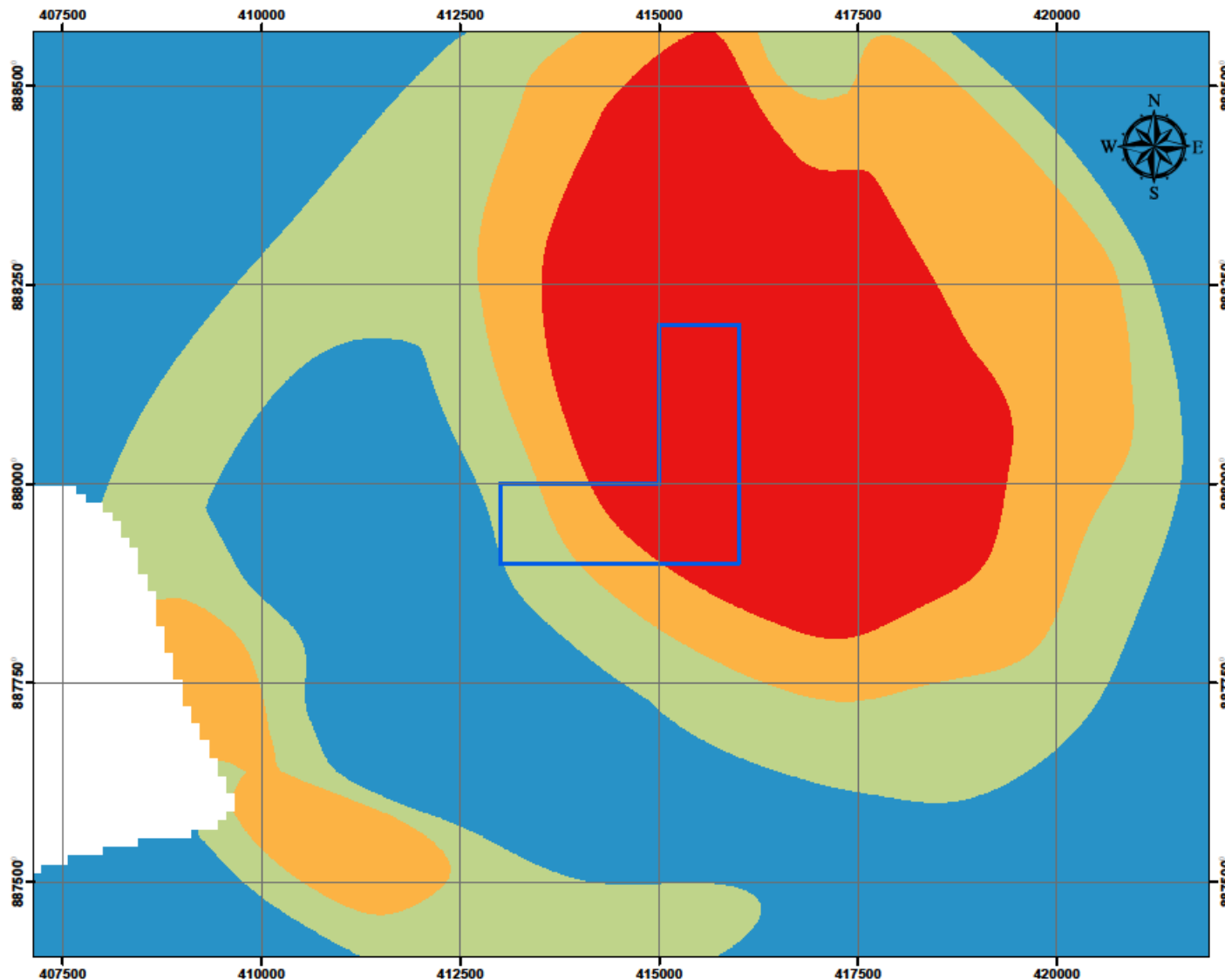
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DEL AG	PROYECCIÓN: UTM	1/50000	06
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO GEOQUIMICO DE PB



Background	85.800
Treshold 1	337.200
Treshold 2	588.600
Treshold 3	840.000

LEYENDA



Área de estudio

Concentración de Pb

Blue	-236.2518311 - 85.8
Green	85.80000001 - 337.2
Orange	337.20000001 - 588.6
Red	588.60000001 - 840

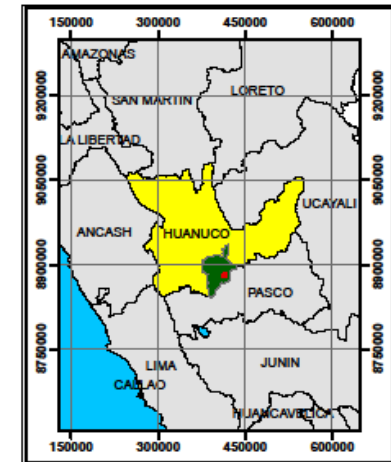
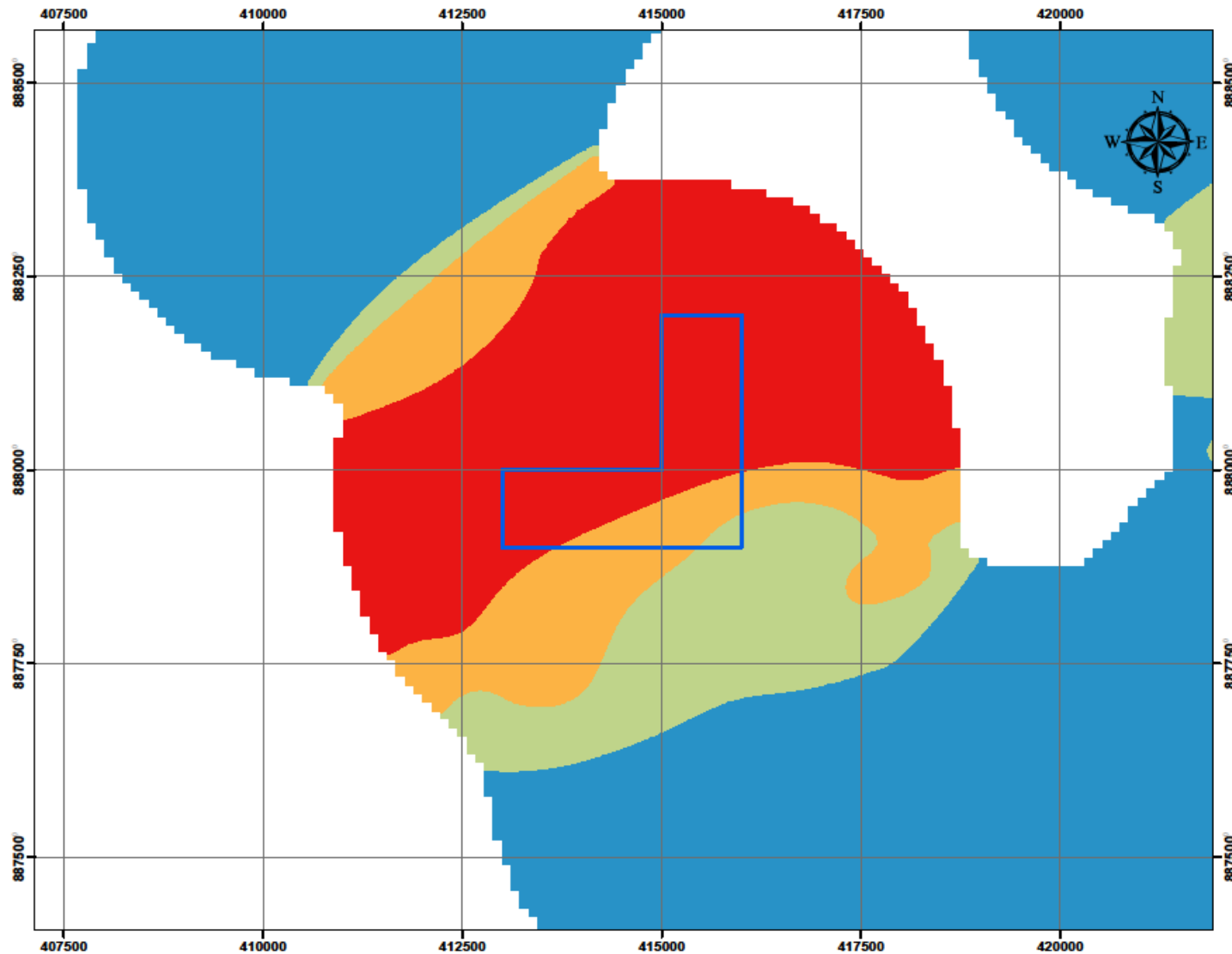
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN				
FACULTAD DE INGENIERÍA				
ESCUELA PROFESIONAL DE INGENIERÍA GEOLOGICA				
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DEL PB	PROYECCIÓN: UTM	07
	REVISADO POR:		ZONA: 18 SUR	
	Mg. Eder Guido ROBLES MORALES		FECHA: 1/07/2024	
			COORDENADAS: WGS - 84	

PLANO GEOQUIMICO DE ZN



Background	87.000
Treshold 1	166.700
Treshold 2	246.400
Treshold 3	326.100

LEYENDA



Área de estudio

Concentración de Zn

Blue	-2.304814577 - 87
Green	87.00000001 - 166.7
Orange	166.70000001 - 246.4
Red	246.40000001 - 326.1

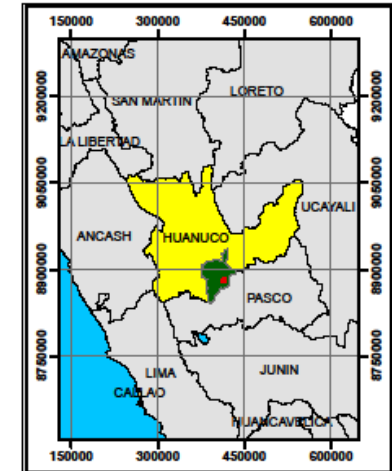
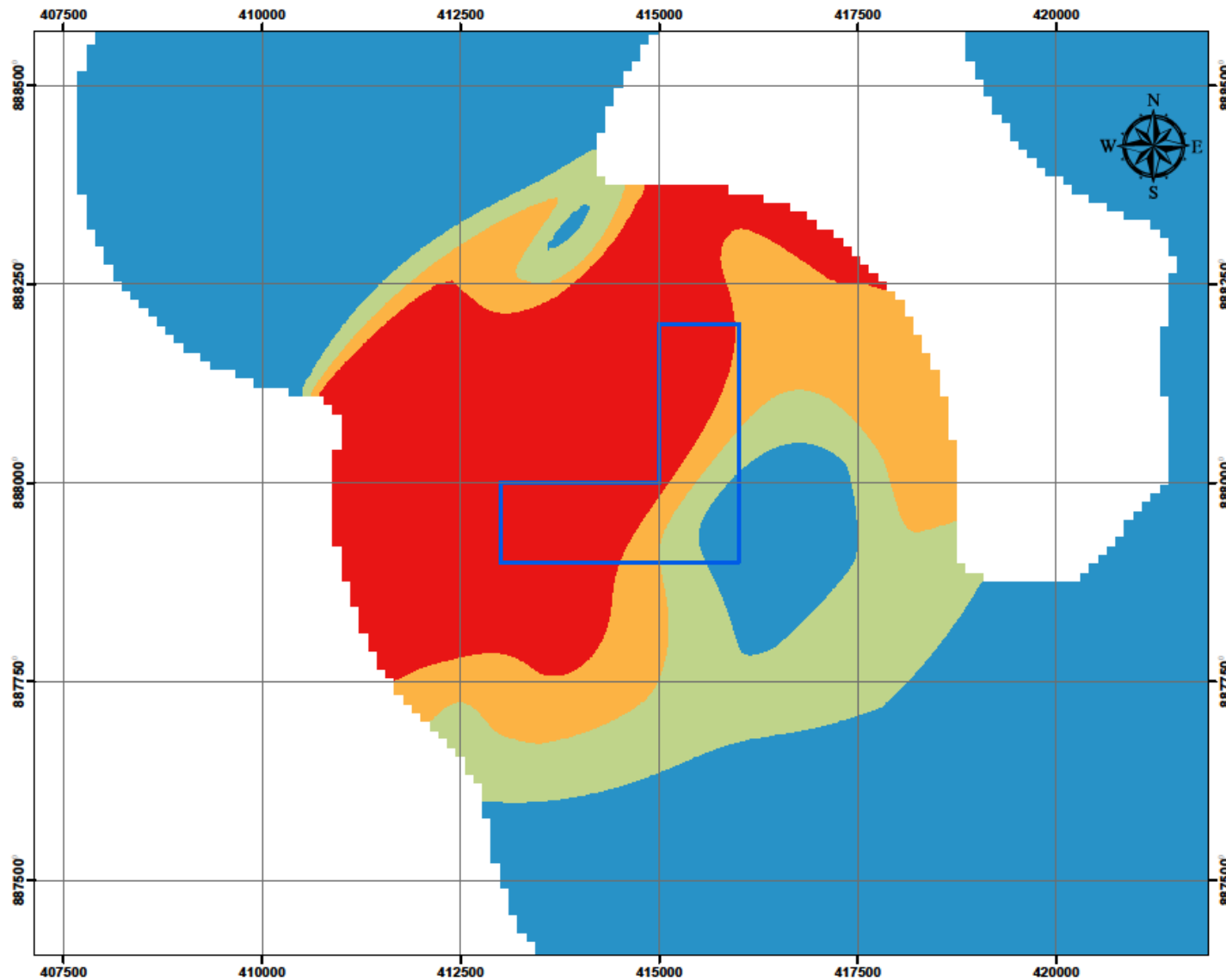
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DEL ZN	PROYECCIÓN: UTM	1/50000	08
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO GEOQUIMICO DE CU



Background	264.000
Treshold 1	1233.000
Treshold 2	2202.000
Treshold 3	3171.000

LEYENDA

Área de estudio

Concentración de Cu

	-818.6484985 - 264
	264.0000001 - 1,233
	1,233.0000001 - 2,202
	2,202.0000001 - 3,171

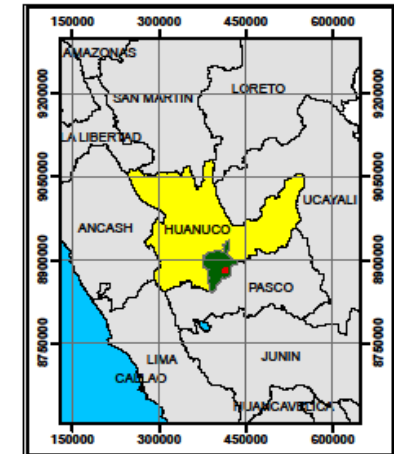
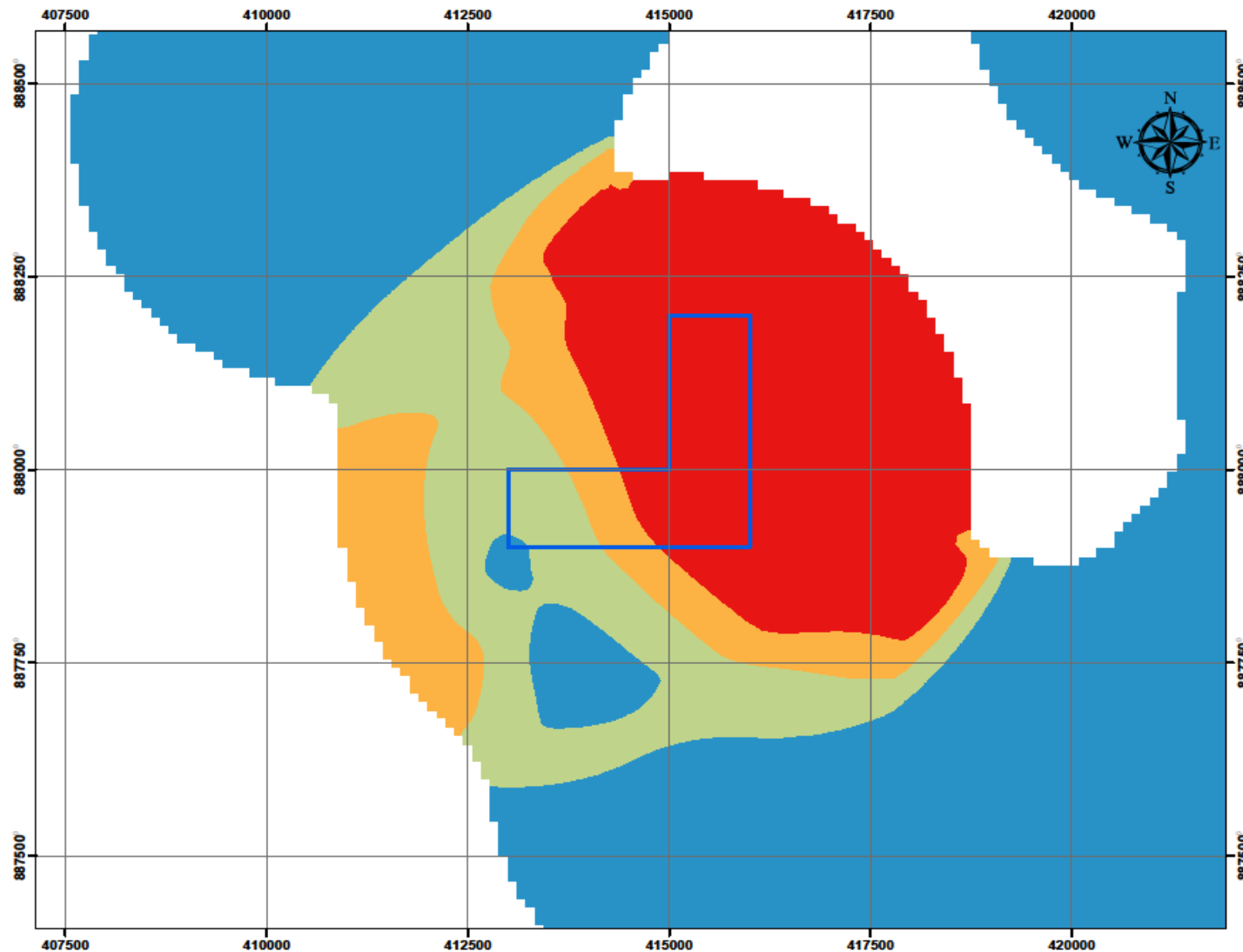
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERÍA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DEL CU	PROYECCIÓN: UTM	1/50000	09
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO GEOQUIMICO DE MO



Background	4.7
Treshold 1	9.568
Treshold 2	14.436
Treshold 3	19.304

LEYENDA



Área de estudio

Concentración de Mo

0.882457256 - 4.7
4.700000001 - 9.568
9.568000001 - 14.436
14.436000001 - 19.304

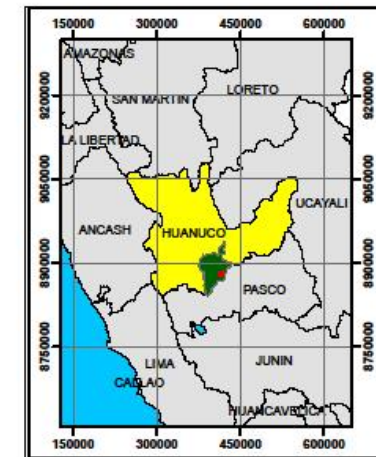
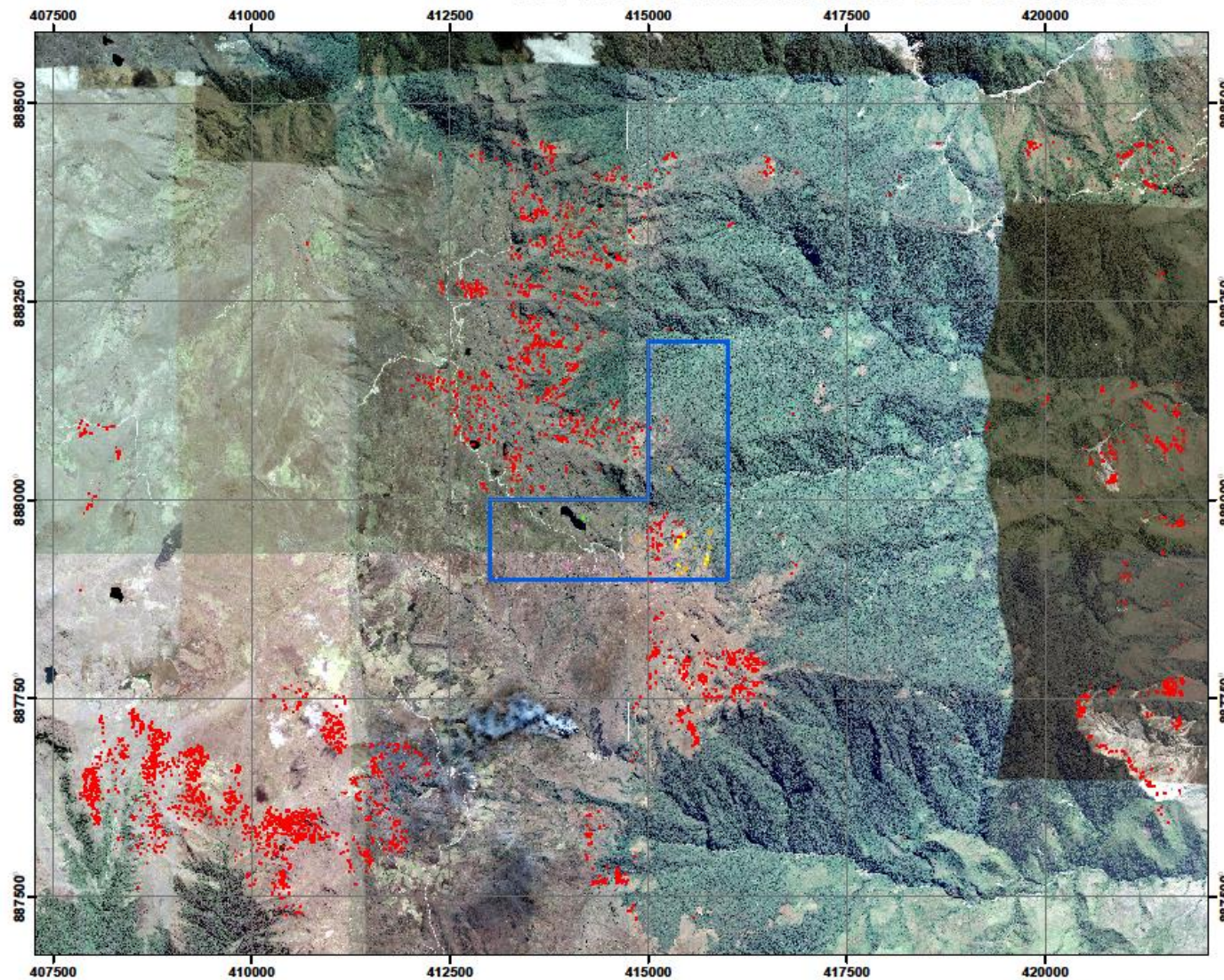
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO GEOQUIMICO DE MO	PROYECCIÓN: UTM	1/50000	10
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

PLANO DE ANOMALIAS ESPECTRALES



LEYENDA

- Oxidos
- Alunita
- Pirofilita
- Caolinita
- Dickita
- Clorita
- Epidota
- Jarosita
- Area_de_estudio

0 1,250 2,500 5,000 m

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE ANOMALIAS ESPECTRALES	PROYECCIÓN: UTM	1/50000	11
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	1/07/2024	

ANEXO II. ESTADITICA GEOQUIMICA

CORRELACIÓN DE SPEARMAN

	Au	Al	Ca	Cr	Cu	Fe	K	Mg	Mn	Na	P	Sr	Ti	V	Zn	Zr	Ag	As	Be	Co	La	Mo	Ni	Pb	Sb	Sc
Al	0.049																									
Ca	0.269	0.096																								
Cr	0.117	-0.051	0.123																							
Cu	0.328	0.527	0.342	0.066																						
Fe	0.383	0.41	0.598	0.179	0.676																					
K	0.203	0.538	-0.004	0.37	0.471	0.319																				
Mg	-0.073	0.64	0.131	-0.487	0.363	0.25	-0.007																			
Mn	0.075	0.569	0.296	0.087	0.37	0.617	0.289	0.353																		
Na	0.265	-0.124	0.719	0.364	0.113	0.371	0.079	-0.138	0.091																	
P	0.255	0.149	0.708	0.394	0.376	0.645	0.022	0.081	0.333	0.487																
Sr	0.096	0.27	0.762	-0.046	0.124	0.487	-0.047	0.303	0.583	0.523	0.407															
Ti	-0.078	0.496	0.372	-0.235	0.035	0.4	-0.055	0.447	0.535	0.092	0.283	0.574														
V	-0.146	0.607	0.25	-0.175	0.143	0.54	0.171	0.531	0.573	0.043	0.189	0.514	0.772													
Zn	0.371	0.511	0.376	0.091	0.877	0.726	0.393	0.357	0.417	0.12	0.405	0.24	0.113	0.203												
Zr	0.349	0.201	0.74	-0.013	0.557	0.798	0.136	0.254	0.41	0.445	0.593	0.58	0.397	0.339	0.61											
Ag	0.424	0.188	0.418	-0.054	0.649	0.524	0.096	0.204	0.22	0.177	0.355	0.204	-0.027	-0.094	0.602	0.572										
As	0.237	-0.115	0.22	0.512	0.233	0.235	0.256	-0.33	0.027	0.371	0.335	0.004	-0.277	-0.17	0.173	0.142	0.259									
Be	0.125	0.641	-0.176	0.079	0.404	0.144	0.749	0.154	0.211	-0.285	-0.043	-0.126	0.003	0.165	0.305	0.006	0.078	-0.012								
Co	0.184	0.726	0.188	-0.157	0.64	0.498	0.232	0.632	0.584	-0.108	0.343	0.279	0.444	0.39	0.582	0.325	0.468	-0.022	0.394							
La	-0.217	-0.072	0.011	0.447	-0.375	-0.033	0.169	-0.296	0.226	0.27	-0.037	0.317	0.047	0.234	-0.322	-0.074	-0.46	0.154	0.022	-0.367						
Mo	0.099	-0.186	0.28	0.614	0.157	0.115	0.191	-0.411	0.085	0.459	0.3	0.096	-0.316	-0.324	0.144	0.148	-0.008	0.452	-0.071	-0.303	0.381					
Ni	-0.222	0.348	-0.123	0.538	-0.004	-0.032	0.404	0.019	0.366	0.035	0.026	0.121	0.047	0.155	0.005	-0.111	-0.2	0.283	0.33	0.075	0.537	0.369				
Pb	0.411	0.041	0.049	0.38	0.386	0.258	0.46	-0.213	0.04	0.176	0.137	-0.078	-0.383	-0.234	0.457	0.198	0.217	0.385	0.343	-0.012	0.157	0.486	0.242			
Sb	0.371	-0.015	0.382	-0.071	0.603	0.458	0.014	0.085	0.065	0.274	0.345	0.154	-0.087	-0.197	0.511	0.574	0.752	0.058	-0.005	0.301	-0.379	0.092	-0.35	0.232		
Sc	-0.061	0.856	0.262	-0.193	0.323	0.478	0.322	0.705	0.65	-0.024	0.218	0.481	0.727	0.808	0.355	0.328	0.086	-0.281	0.378	0.65	-0.019	-0.334	0.184	-0.223	-0.119	
Y	-0.017	0.145	0.401	0.206	-0.098	0.3	0.011	0.089	0.577	0.471	0.25	0.72	0.389	0.437	0.002	0.285	-0.094	0.097	-0.087	0.014	0.713	0.36	0.456	0.058	-0.168	0.292

>0.9	Muy buena
0.75 - 0.90	Buena
0.50 - 0.75	Moderada
0.25 - 0.50	Baja
0 - 0.25	Muy baja

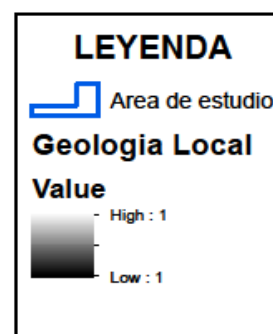
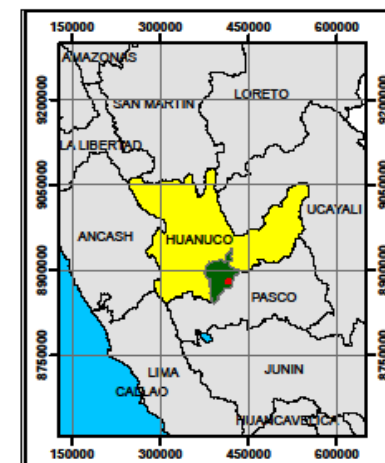
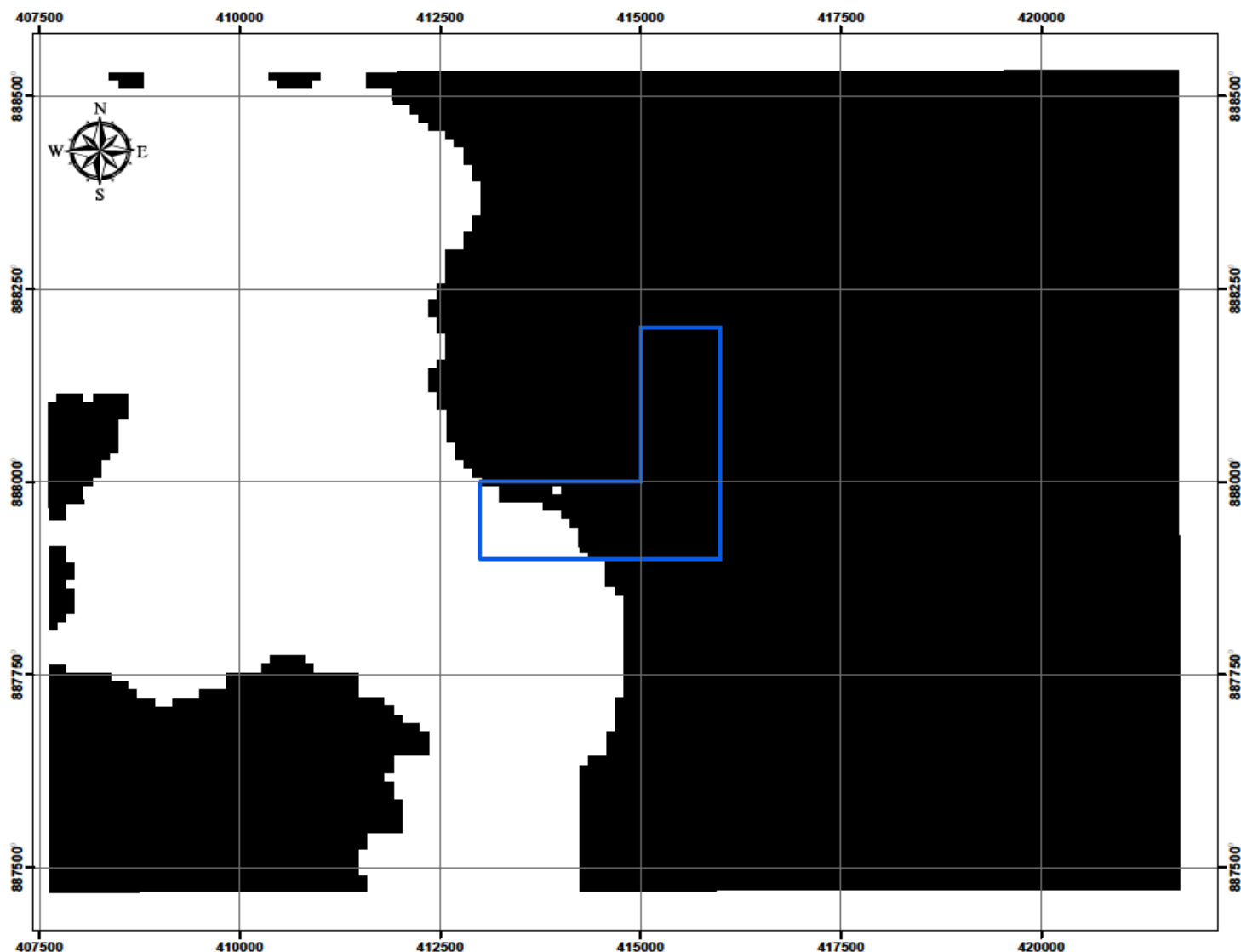
CORRELACIÓN DE PEARSON

	Au	Al	Ca	Cr	Cu	Fe	K	Mg	Mn	Na	P	Sr	Ti	V	Zn	Zr	Ag	As	Be	Co	La	Mo	Ni	Pb	Sb	Sc
Al	-0.016																									
Ca	0.905	0.072																								
Cr	0.464	-0.011	0.5																							
Cu	0.895	0.072	0.923	0.442																						
Fe	0.69	0.219	0.75	0.353	0.796																					
K	0.407	0.658	0.447	0.426	0.462	0.391																				
Mg	-0.463	0.6	-0.427	-0.511	-0.396	-0.168	-0.048																			
Mn	0.351	0.304	0.408	0.169	0.398	0.485	0.333	0.201																		
Na	0.723	0.071	0.858	0.471	0.687	0.61	0.36	-0.3	0.357																	
P	0.91	0.036	0.985	0.557	0.942	0.76	0.452	-0.494	0.338	0.806																
Sr	0.229	0.302	0.411	0.002	0.223	0.426	0.115	0.297	0.628	0.571	0.279															
Ti	-0.433	0.539	-0.296	-0.345	-0.441	-0.041	-0.044	0.706	0.175	-0.106	-0.374	0.464														
V	-0.385	0.586	-0.291	-0.229	-0.328	0.208	0.127	0.576	0.11	-0.12	-0.336	0.387	0.786													
Zn	0.722	0.155	0.725	0.293	0.873	0.807	0.337	-0.184	0.322	0.515	0.75	0.23	-0.296	-0.117												
Zr	0.723	0.18	0.816	0.203	0.797	0.877	0.369	-0.131	0.387	0.696	0.793	0.518	-0.025	0.087	0.741											
Ag	0.857	0.113	0.905	0.333	0.925	0.738	0.402	-0.301	0.349	0.67	0.905	0.274	-0.338	-0.296	0.788	0.803										
As	0.733	0.007	0.784	0.586	0.748	0.599	0.447	-0.497	0.069	0.685	0.822	0.103	-0.393	-0.241	0.595	0.637	0.733									
Be	0.115	0.673	0.103	0.14	0.176	0.113	0.781	0.111	0.124	-0.073	0.124	-0.13	0.01	0.167	0.107	0.09	0.131	0.094								
Co	0.186	0.73	0.268	-0.082	0.296	0.357	0.4	0.58	0.599	0.189	0.222	0.377	0.413	0.306	0.311	0.314	0.349	0.052	0.405							
La	-0.185	-0.083	-0.125	0.35	-0.284	-0.088	0.143	-0.178	0.181	0.12	-0.156	0.228	0.098	0.219	-0.384	-0.163	-0.353	-0.019	-0.002	-0.319						
Mo	0.626	-0.166	0.676	0.684	0.625	0.402	0.325	-0.514	0.149	0.609	0.71	0.102	-0.471	-0.448	0.454	0.452	0.514	0.687	0.001	-0.134	0.236					
Ni	-0.131	0.37	-0.082	0.504	-0.116	-0.08	0.452	0.171	0.268	0.015	-0.082	0.12	0.141	0.175	-0.126	-0.131	-0.171	0.107	0.349	0.165	0.479	0.252				
Pb	0.802	0.003	0.766	0.525	0.802	0.603	0.524	-0.482	0.188	0.595	0.802	0.07	-0.525	-0.356	0.688	0.623	0.718	0.762	0.244	0.082	-0.018	0.725	0.083			
Sb	0.815	-0.07	0.862	0.277	0.911	0.695	0.277	-0.334	0.452	0.653	0.857	0.276	-0.429	-0.404	0.758	0.746	0.883	0.582	0.026	0.295	-0.273	0.541	-0.228	0.678		
Sc	-0.342	0.805	-0.249	-0.285	-0.275	0.074	0.272	0.784	0.376	-0.093	-0.324	0.462	0.789	0.785	-0.12	0.007	-0.238	-0.396	0.327	0.605	0.066	-0.475	0.248	-0.408	-0.333	
Y	0.004	0.104	0.121	0.2	-0.069	0.222	0.05	0.145	0.472	0.395	0.037	0.706	0.405	0.395	-0.102	0.208	-0.076	0.05	-0.178	0.034	0.728	0.25	0.385	-0.015	-0.072	0.343

>0.9	Muy buena
0.75 - 0.90	Buena
0.50 - 0.75	Moderada
0.25 - 0.50	Baja
0 - 0.25	Muy baja

ANEXO III. ANALISIS ESPACIAL EN ARCGIS

PLANO DE PROXIMIDAD DE ROCAS IGNEAS



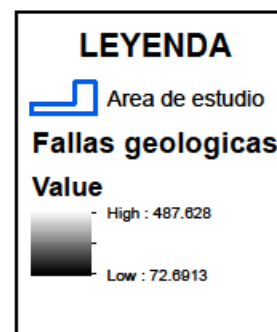
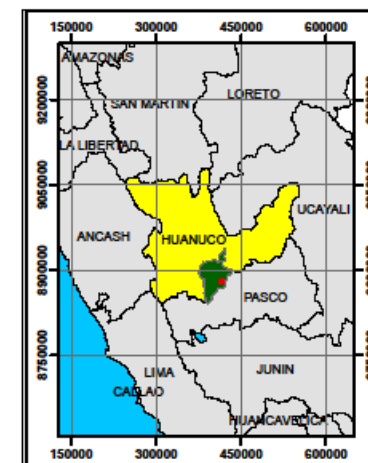
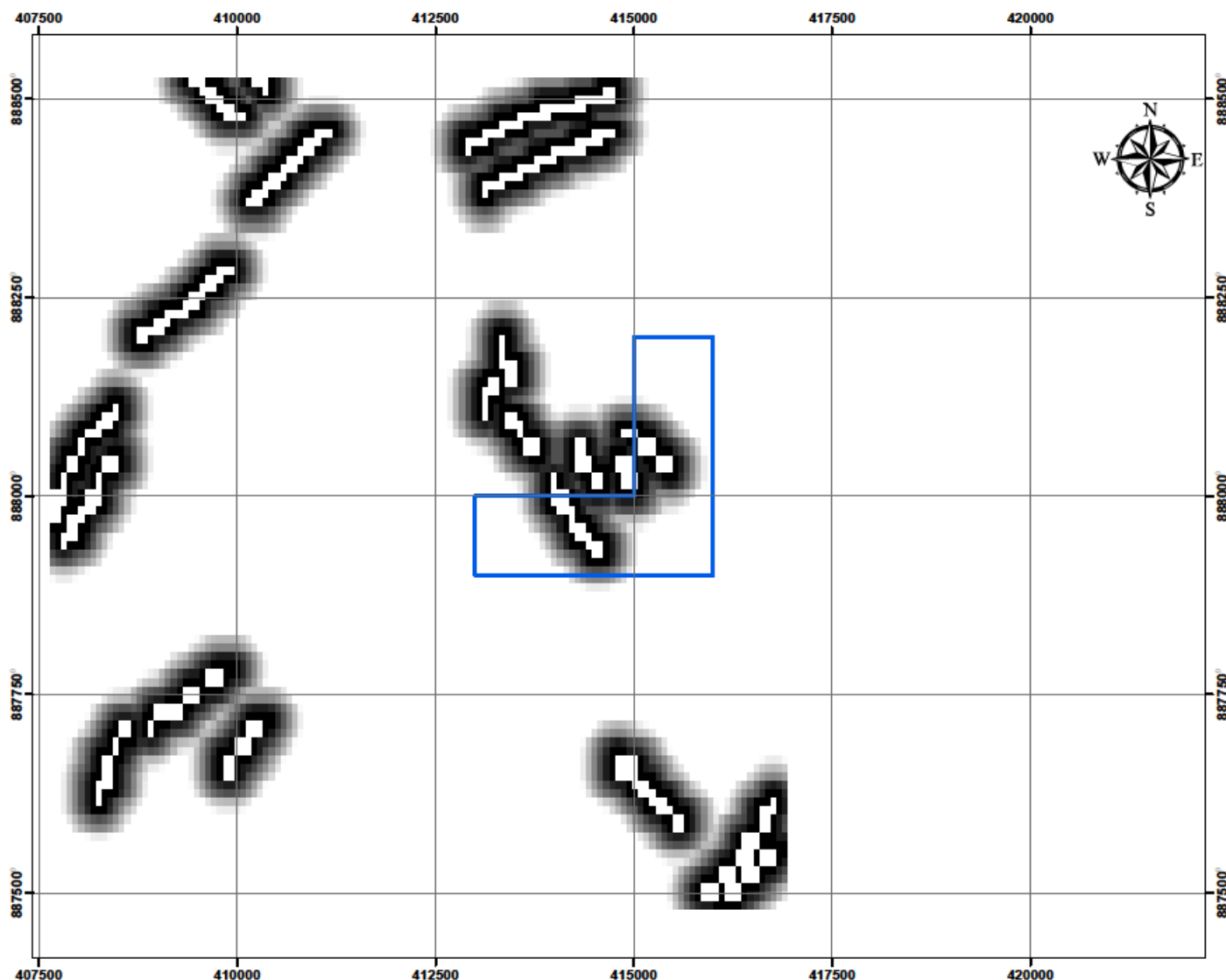
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE PROXIMIDAD DE ROCAS IGNEAS	PROYECCIÓN: UTM	1/50000	12
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE PROXIMIDAD DE FALLAS GEOLOGICAS



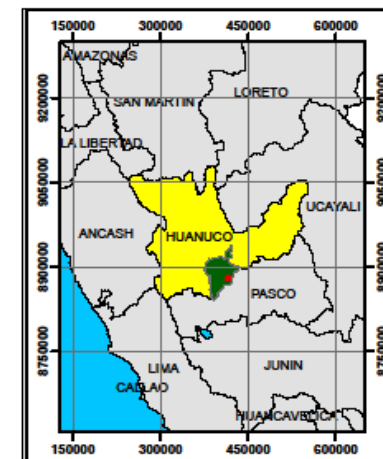
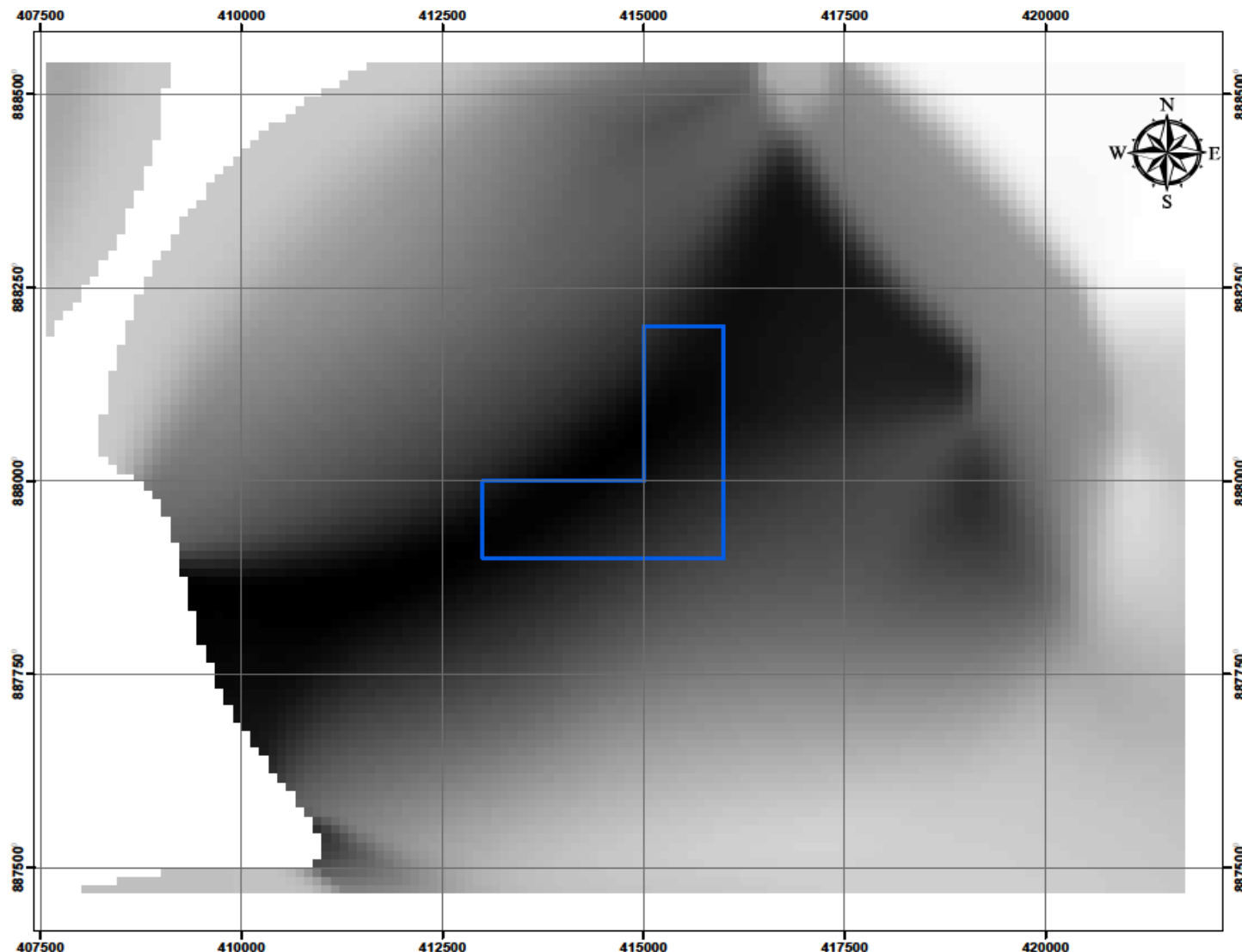
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE PROXIMIDAD DE FALLAS GEOLOGICAS	PROYECCIÓN: UTM	1/50000	13
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE AU



LEYENDA

 Area de estudio

Concentración de Au

Value

High : 0.471179
Low : -1.57803

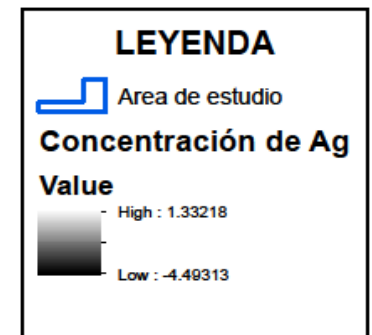
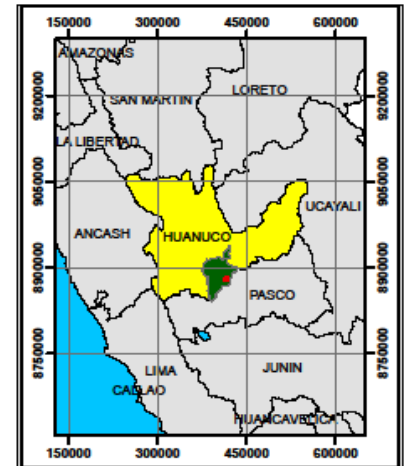
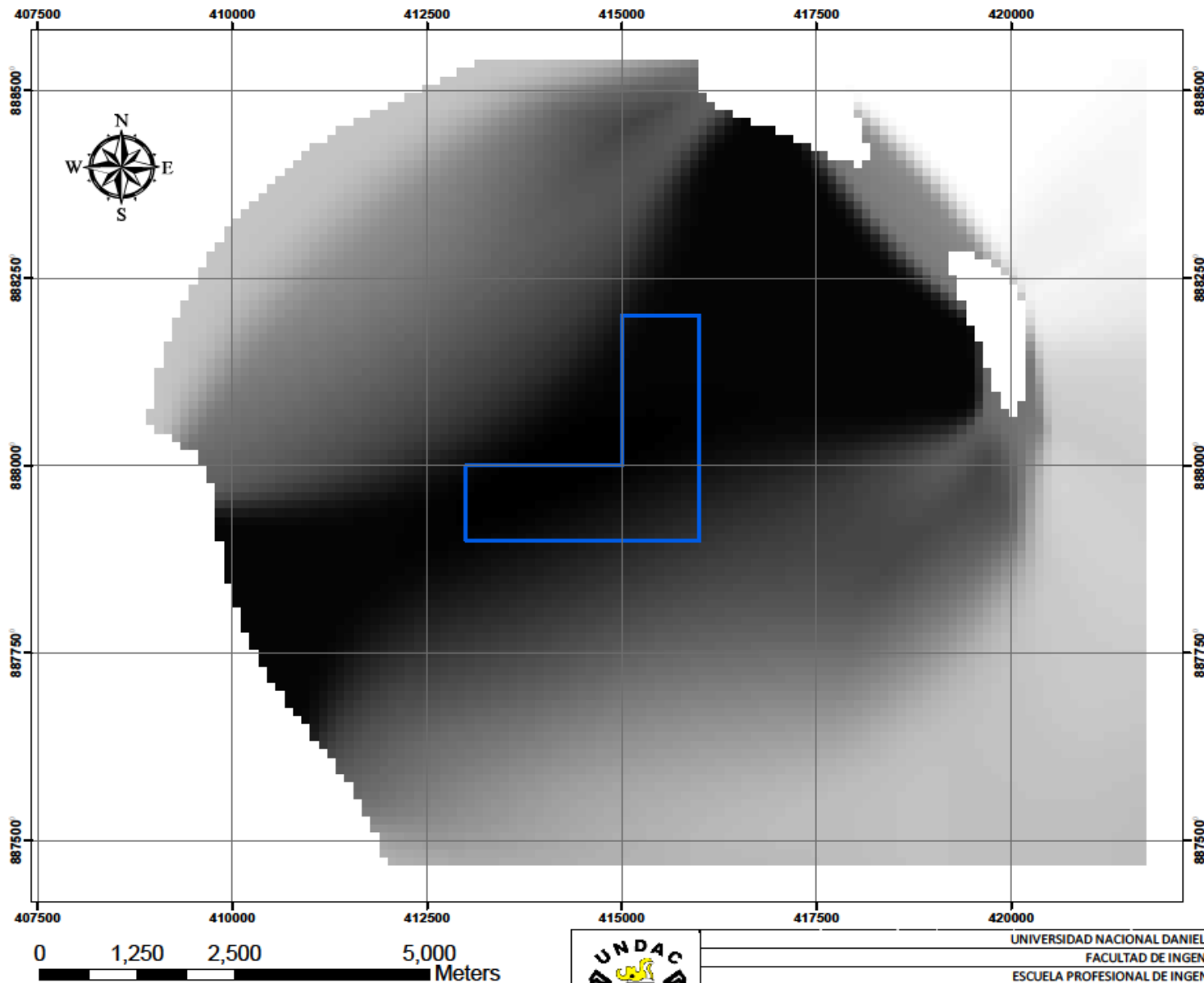
0 1,250 2,500 5,000 Meters

1:50,000



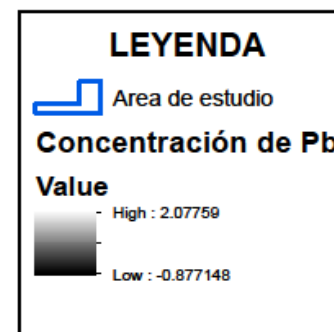
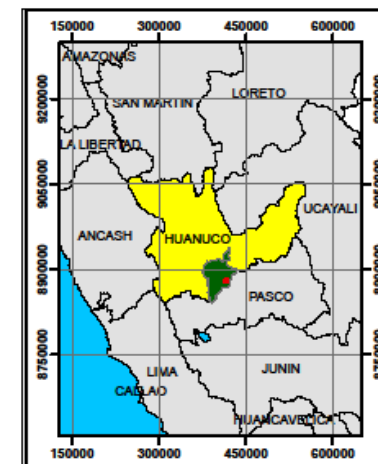
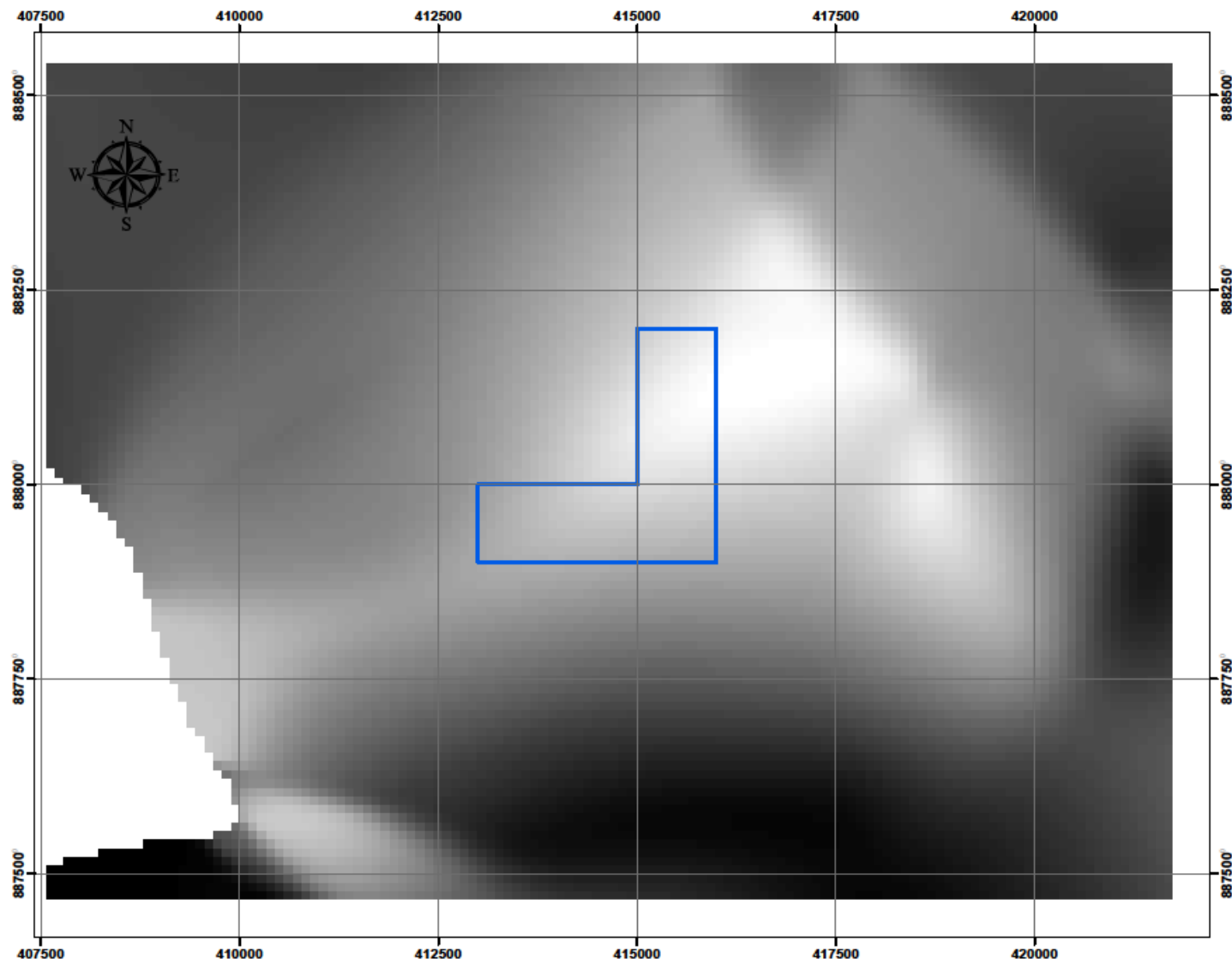
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE CONCENTRACIÓN GEOQUIMICA DE AU	PROYECCIÓN: UTM	1/50000	14
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE AG



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERÍA					
ESCUELA PROFESIONAL DE INGENIERÍA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE CONCENTRACIÓN GEOQUIMICA DE AG	PROYECCIÓN: UTM	1/50000	15
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE PB



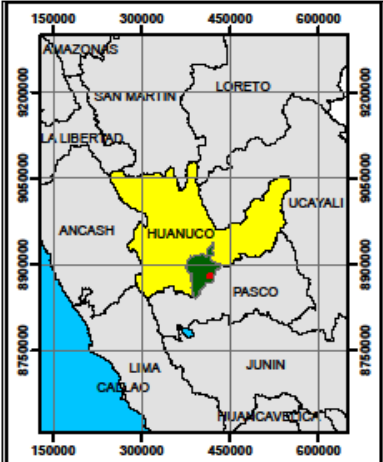
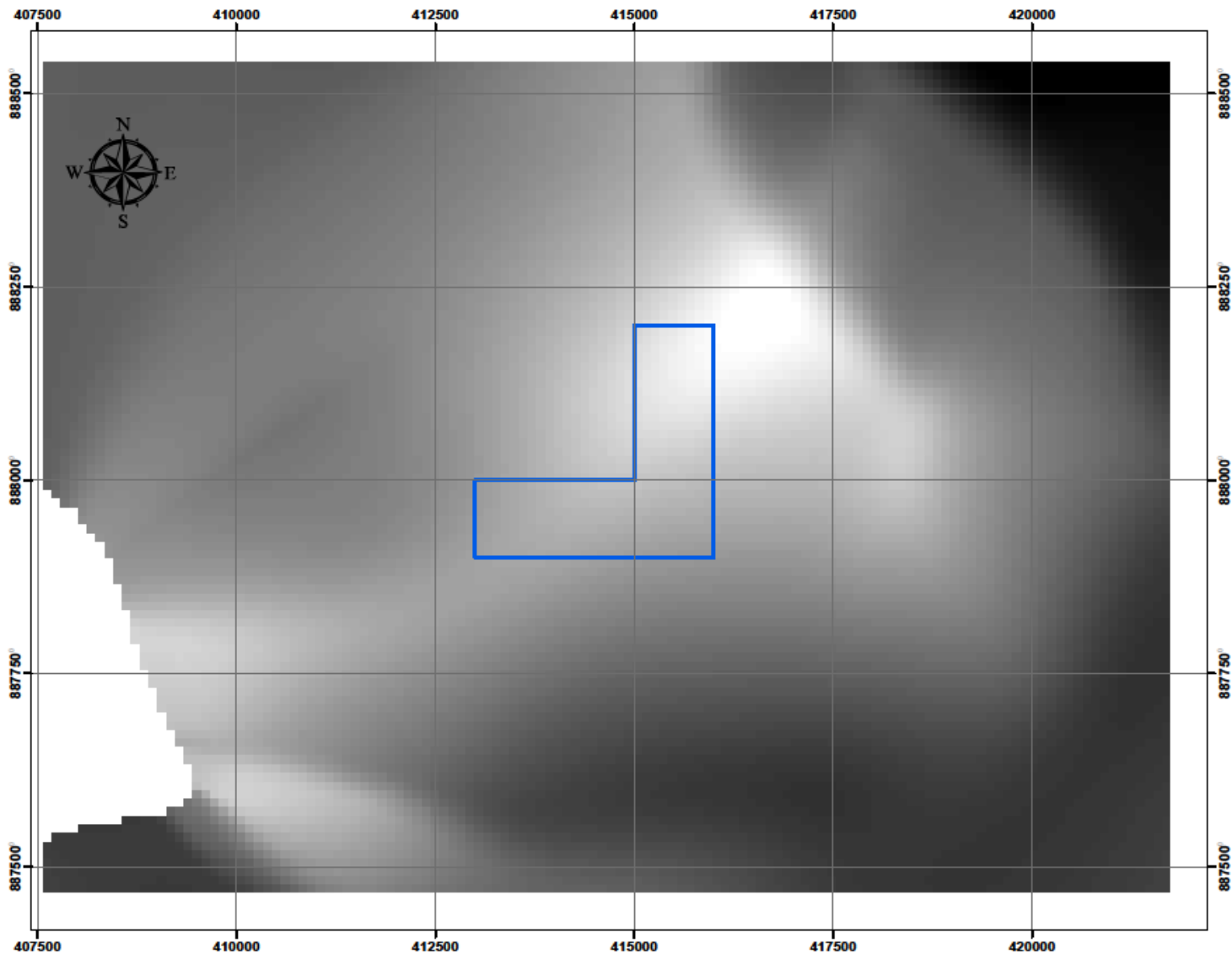
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE CONCENTRACIÓN GEOQUIMICA DE PB	PROYECCIÓN: UTM	1/50000	16
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE ZN



LEYENDA

 Area de estudio

Concentración de Zn

Value

 High : 0.315077
Low : -0.18567

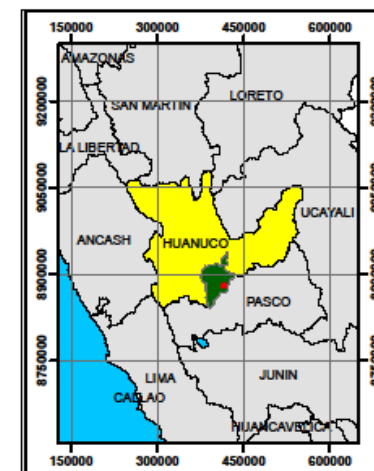
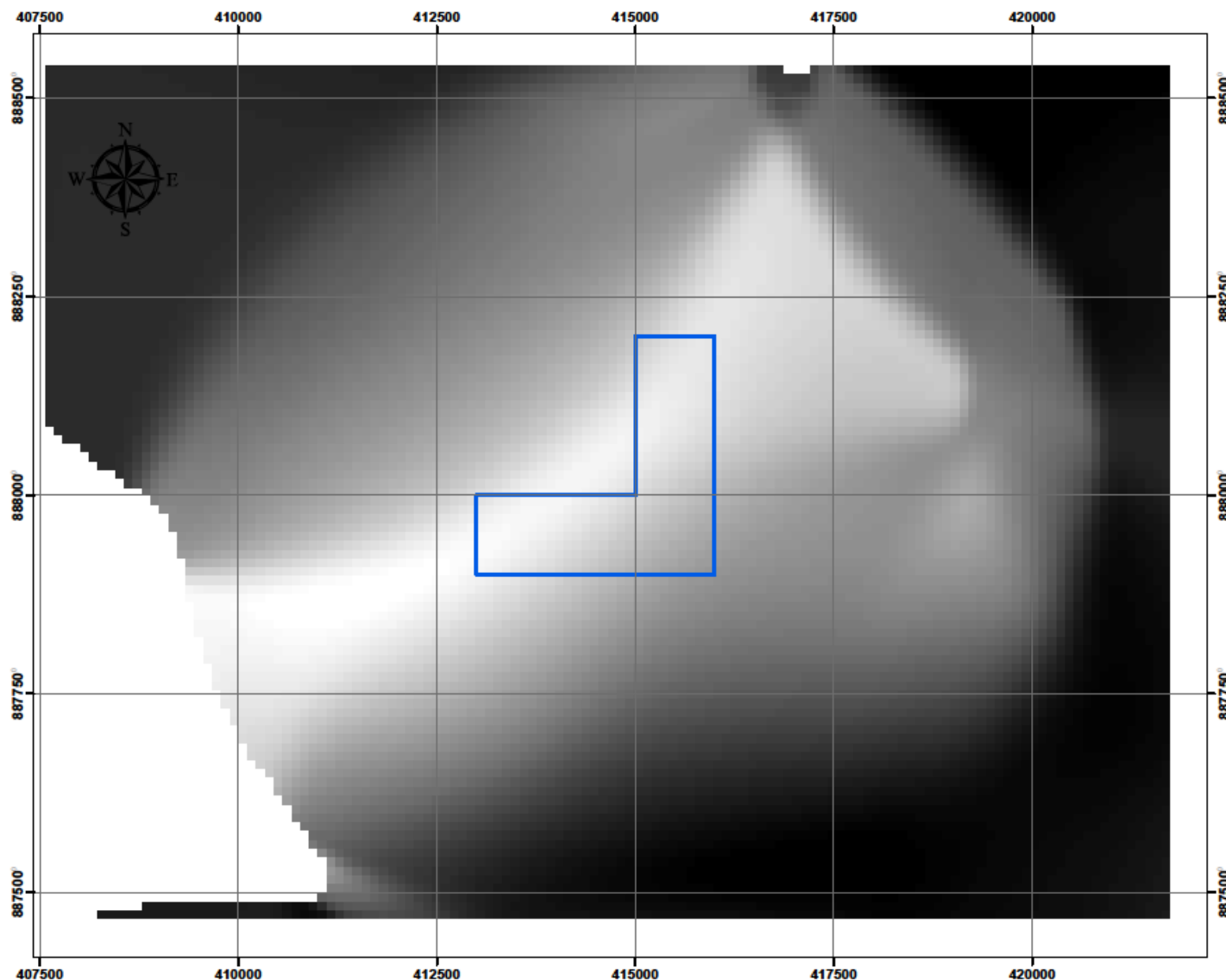
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE CONCENTRACIÓN GEOQUIMICA DE ZN	PROYECCIÓN: UTM	1/50000	17
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE CU



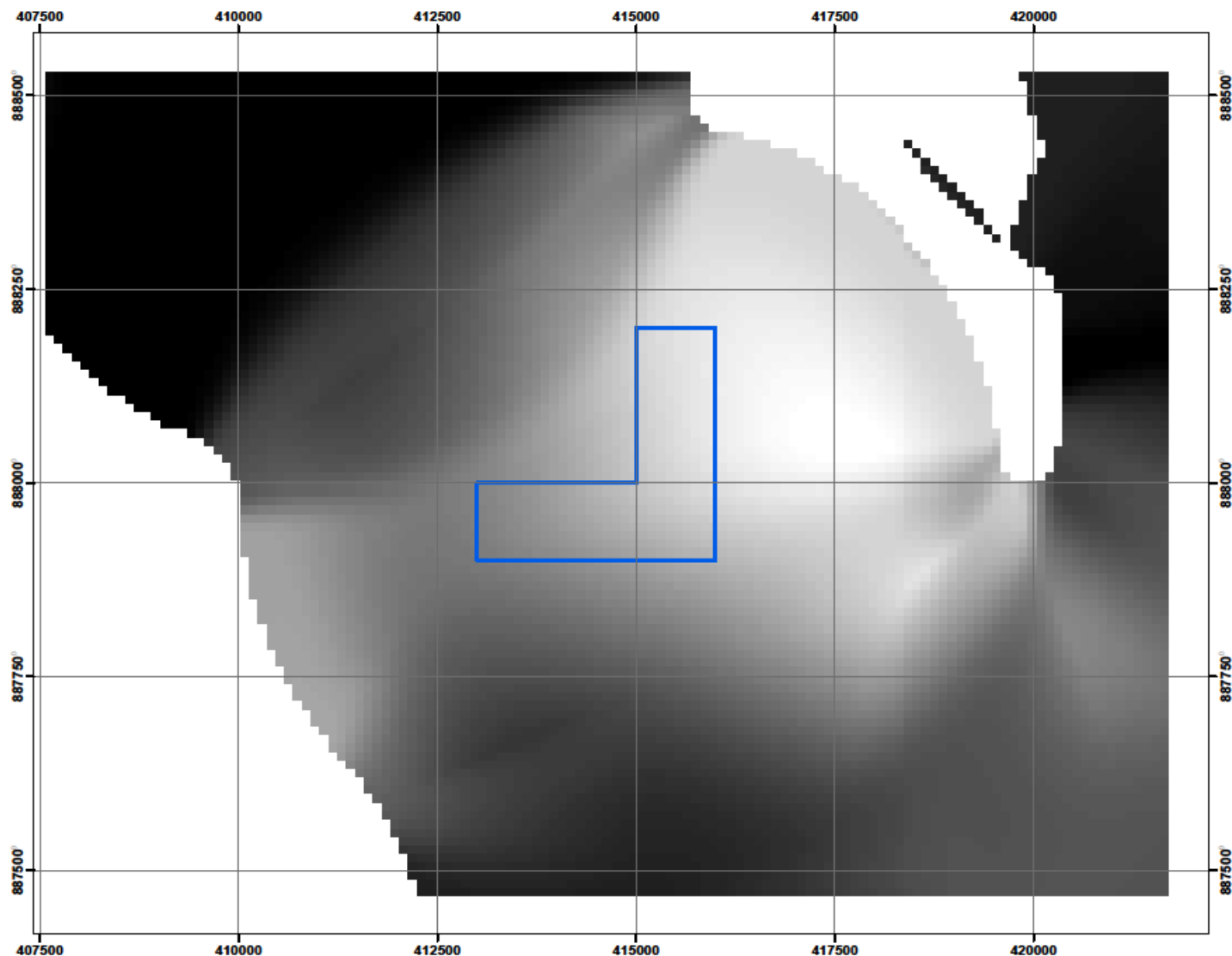
0 1,250 2,500 5,000 Meters

1:50,000



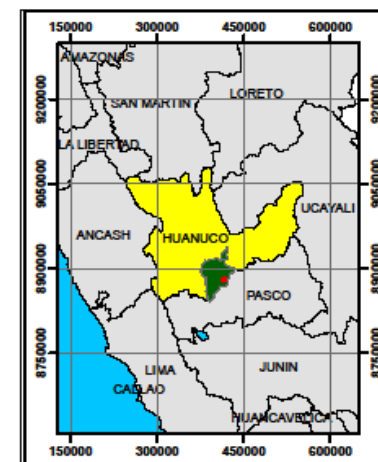
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE CONCENTRACIÓN GEOQUIMICA DE CU	PROYECCIÓN: UTM	1/50000	18
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DE CONCENTRACIÓN GEOQUIMICA DE MO



0 1,250 2,500 5,000 Meters

1:50,000



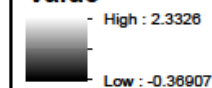
LEYENDA



Area de estudio

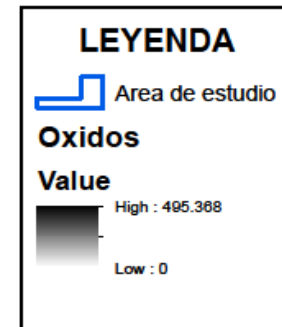
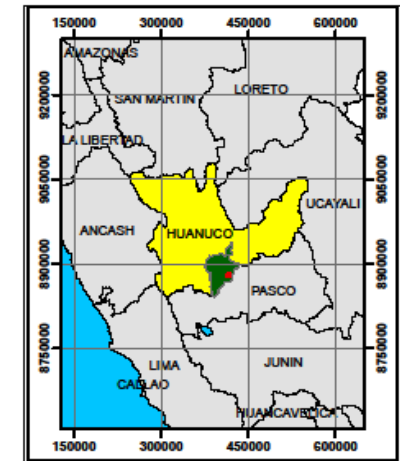
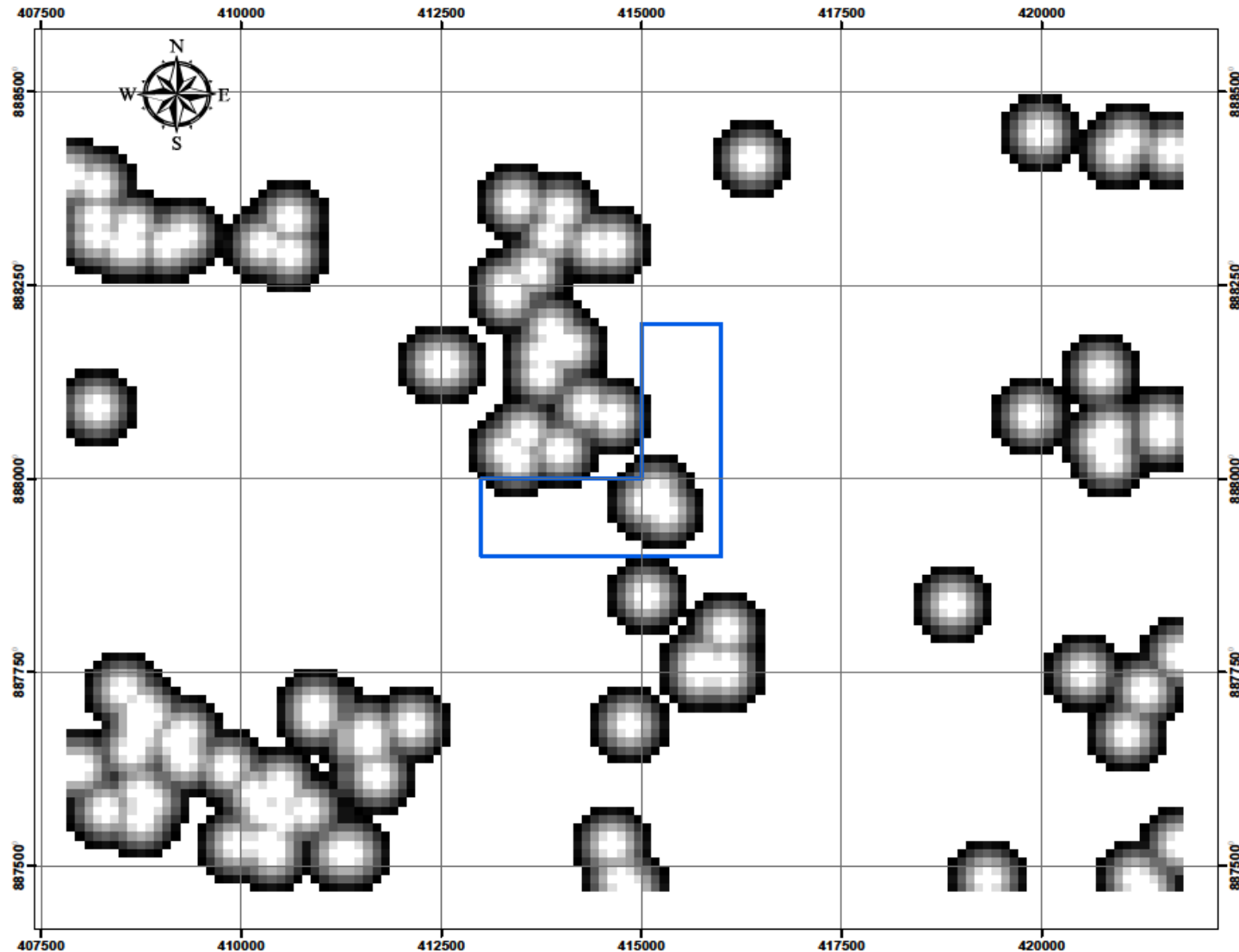
Concentración de Mo

Value



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE	PROYECCIÓN: UTM	1/50000	19
	REVISADO POR:	CONCENTRACIÓN	ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES	GEOQUIMICA DE MO	COORDENADAS: WGS - 84	2/08/2024	

PLANO DE PROXIMIDAD DE LAS ANOMALIAS ESPECTRALES

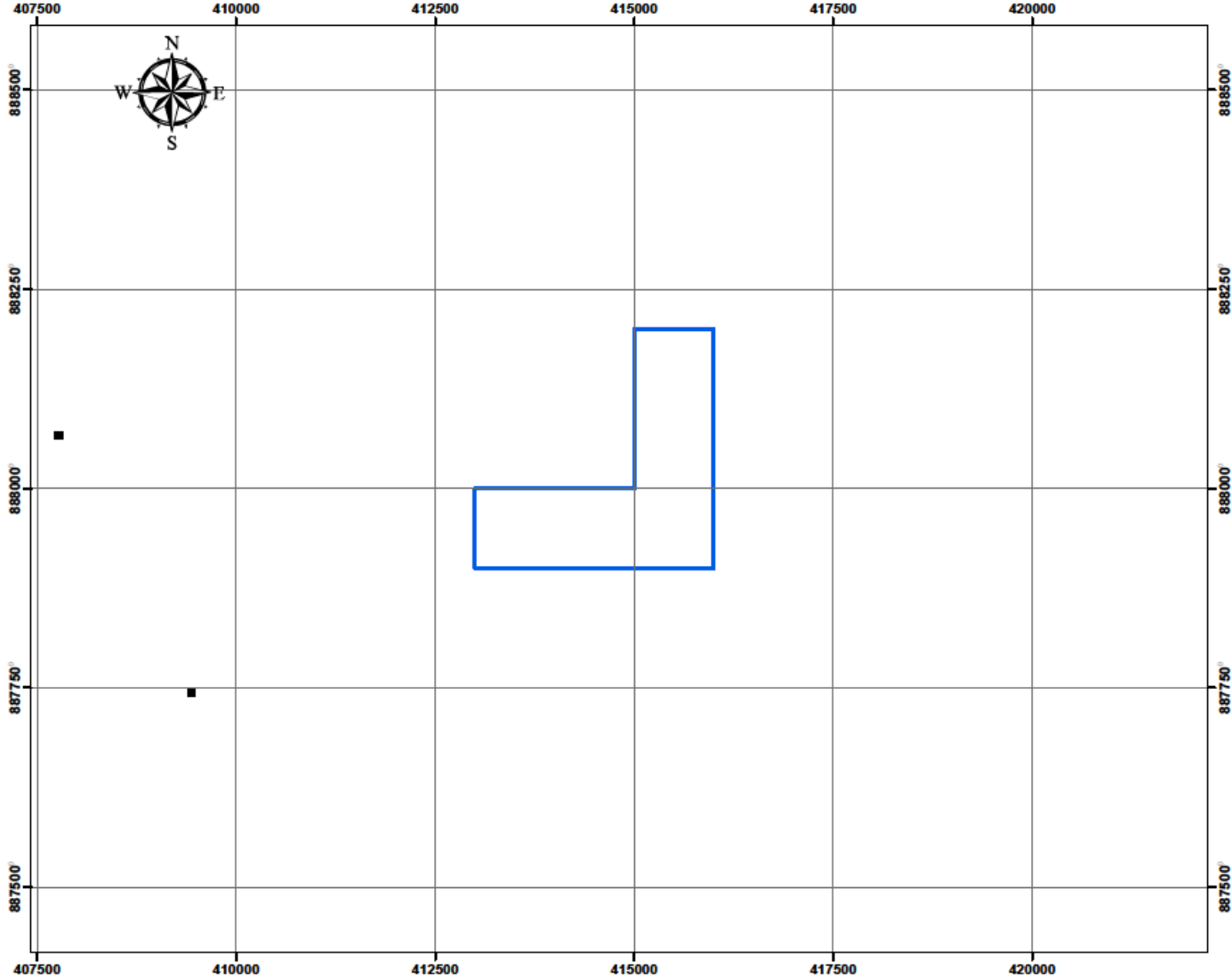


0 1,250 2,500 5,000 Meters

1:50,000

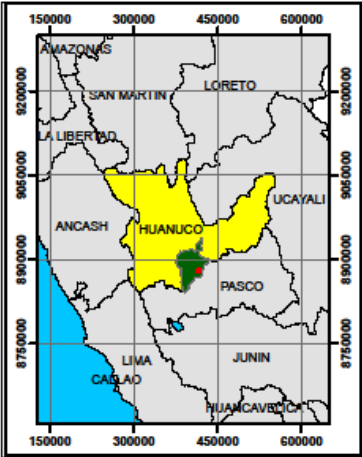
UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERÍA					
ESCUELA PROFESIONAL DE INGENIERÍA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DE PROXIMIDAD DE LAS ANOMALIAS ESPECTRALES	PROYECCIÓN: UTM	1/50000	20
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

PLANO DEL AREA DE INTERES DE LA OCURENCIA MINERAL



0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERIA					
ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	PLANO DEL AREA DE INTERES DE LA OCURENCIA MINERAL	PROYECCIÓN: UTM	1/50000	21
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

ANEXO IV. ESTRUCTURA DE ALGORITMOS (CÓDIGOS) EN PYTHON

1. Importación de librerías.

Mapas de predicción mineral con Python y Sklearn

Importemos las librerías principales

Dentro de las librerías que se van a utilizar está Rasterio que permite la lectura de los metadatos de la información de entrada a fin de obtener la georeferencia de las mismas.

```
In [104]: #Importing the initial few python packages as their usual aliases
import pandas as pd
import numpy as np
from skimage import io
import matplotlib.pyplot as plt
from IPython.display import Markdown as md
```

```
In [105]: import rasterio as rio
```

2. Importación de rasters de entrada.

Importemos los rasters de entrada

Para este proyecto se seleccionaron la geoquímica, Geología, Fallas, Anomalías espectrales
Comencemos con la geoquímica:

*****_NOTA: dentro del paréntesis escribir los elementos a utilizar, siempre y cuando hayamos nombrado nuestros archivos solo con el nombre del elemento._*****

```
In [106]: Au = io.imread('./Python/Au.tif')
Ag = io.imread('./Python/Ag.tif')
Pb = io.imread('./Python/Pb.tif')
Zn = io.imread('./Python/Zn.tif')
Cu = io.imread('./Python/Cu.tif')
Mo = io.imread('./Python/Mo.tif')
```

```
In [107]: #Revisemos qué se importó para saber bien con qué estamos trabajando
plt.figure(figsize=(12, 20))
plt.imshow(Cu)
```

```
In [108]: #Loading up georeferenced tiff images for Geology
geo = io.imread('./Python/Geologia_local_f.tif')
```

```
In [109]: #Revisemos qué se importó para saber bien con qué estamos trabajando
plt.figure(figsize=(12, 20))
plt.imshow(geo)
```

```
In [110]: #Loading up georeferenced tiff images for rest of info
Falla = io.imread('./Python/Falla_f.tif')
Oxi = io.imread('./Python/Oxidos_f.tif')
```

```
In [111]: plt.figure(figsize=(12, 20))
plt.imshow(Falla)
```

Ahora importamos la variable de interés a predecir:

```
In [112]: #Loading up georeferenced tiff image for the deposits
AOI = io.imread('./Python/AOI_f.tif')
```

3. Verificación del tamaño de pixel

Antes de proceder debemos asegurarnos que todos los datos de entrada tengan el mismo tamaño (columnas y filas). De lo contrario, se deben volver a exportar las imágenes con la misma resolución. En este caso todas deben ser de 97x128 (filasxcols)

```
In [113]: #Analysing the shape of each numpy array,
print('Au shape is',Au.shape)
print('Ag shape is',Ag.shape)
print('Pb shape is',Pb.shape)
print('Zn shape is',Zn.shape)
print('Cu shape is',Cu.shape)
print('Mo shape is',Mo.shape)
print('Geo is',geo.shape)
print('Falla shape is',Falla.shape)
print('Oxi shape is',Oxi.shape)
print('Area of Interest shape is',AOI.shape)
```

```
Au shape is (97, 128)
Ag shape is (97, 128)
Pb shape is (97, 128)
Zn shape is (97, 128)
Cu shape is (97, 128)
Mo shape is (97, 128)
Geo is (97, 128)
Falla shape is (97, 128)
Oxi shape is (97, 128)
Area of Interest shape is (97, 128)
```

```
In [114]: plt.figure(figsize=(12, 20),dpi=120)
plt.imshow(AOI)
# io.imshow(AOI)
```

4. Cambio de matrices de 2D A 1D (Sklearn).

Sklearn solo trabaja con matrices 1D, por lo que debemos convertir nuestros rasters 2D en una sola columna. La opción "flatten" permite hacer esto manteniendo la estructura de la información.

```
In [115]: #Flattening the arrays to become a single 1D numpy array or vector List of pixel values
```

```
Au1 = Au.flatten()
Ag1 = Ag.flatten()
Pb1 = Pb.flatten()
Zn1 = Zn.flatten()
Cu1 = Cu.flatten()
Mo1 = Mo.flatten()
geo1 = geo.flatten()
Falla1 = Falla.flatten()
Oxi1 = Oxi.flatten()
AOI1 = AOI.flatten()
```

```
#Checking the shape or array length Lines up for all arrays
```

```
print(Au1.shape)
print(Ag1.shape)
print(Pb1.shape)
print(Zn1.shape)
print(Cu1.shape)
print(Mo1.shape)
print(geo1.shape)
print(Falla1.shape)
print(Oxi1.shape)
print(AOI1.shape)
```

```
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
(12416,)
```

5. Uso de librerías Panda

Ahora utilizemos la librería Pandas

Esta funciona como una base datos y es muchísimo más eficiente que utilizar solo matrices. Además, mantenemos la relación de toda la información de entrada para cada fila. Creemos nuestras columnas con nuestros valores de entrada. T se utiliza para transponer la información ingresada y poderla guardar en columnas.

```
In [116]: df = pd.DataFrame(data=[Au1, Ag1, Pb1, Zn1, Cu1, Mo1, geo1, Falla1, Oxi1, AOI1]).T
#df = pd.DataFrame(data=[Au1, Ag1, Pb1, Zn1, Cu1, Mo1, geo1, Falla1, Oxi1, AOI1]).T
```

```
In [117]: #Asignar Los títulos apropiados a cada columna
df.columns=['Au','Ag','Pb','Zn','Cu','Mo','Geologia','Fallas','Anomalias espectrales', 'AOI']
#df.columns=['Au','Ag','Pb','Zn','Cu','Mo','Geologia','Fallas','Anomalias espectrales', 'AOI']

#Verificación
print(df.head(5))
```

	Au	Ag	Pb	Zn	Cu	Mo	Geologia	Fallas	\
0	-0.281144	0.0	0.017937	0.000426	-0.009735	-0.230426	0.0	0.0	
1	-0.283690	0.0	0.016987	-0.000034	-0.009338	-0.269861	0.0	0.0	
2	-0.278217	0.0	0.016219	-0.000522	-0.008858	-0.298409	0.0	0.0	
3	-0.258834	0.0	0.015693	-0.001032	-0.008582	-0.312731	0.0	0.0	
4	-0.225928	0.0	0.015257	-0.001553	-0.008696	-0.317053	0.0	0.0	
	Anomalias espectrales AOI								
0			0.0	1.0					
1			0.0	1.0					
2			0.0	1.0					
3			0.0	1.0					
4			0.0	1.0					

```
In [118]: #Verificación
print(df.tail(5))
```

	Au	Ag	Pb	Zn	Cu	Mo	Geologia	\
12411	0.001487	-0.133122	-0.105132	-0.061835	-0.144675	0.538584	0.0	
12412	-0.002388	-0.139988	-0.074368	-0.060242	-0.137109	0.535752	0.0	
12413	-0.006407	-0.146881	-0.042434	-0.058533	-0.129015	0.532956	0.0	
12414	-0.010528	-0.153811	-0.009256	-0.056699	-0.120455	0.530202	0.0	
12415	-0.014700	-0.160797	0.025216	-0.054731	-0.111515	0.527492	0.0	
	Fallas Anomalias espectrales AOI							
12411	325.085327		0.0	1.0				
12412	363.456451		0.0	1.0				
12413	411.204041		0.0	1.0				
12414	465.451355		0.0	1.0				
12415	0.000000		0.0	1.0				

Vemos que tenemos los valores anteriormente marcados como "No data" dentro de nuestra base de datos. Guardaremos la ifnromación de su índice para luego asignarlo al exportar el raster. Como referencia podemos tomar cualquier elemento de la geoquímica que marcamos como "0". En este caso utilizamos "Au" como referencia.

6. Verificación de no Data y tratamiento estadístico

```
In [119]: n = df.loc[0].at["Ag"]
print(n)

0.0
```

```
In [120]: noData = [df['Ag'].values == n]
```

Ahora, creemos una nueva versión de la base de datos para continuar con nuestro pre procesamiento que será sucedido por la transformación de los mismos. Como funcion adicional en Pandas, se puede obtener un resumen estadístico de nuestras variables:

```
In [121]: df_filter = df
print(df_filter.describe())
```

	Au	Ag	Pb	Zn	Cu \
count	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04
mean	-2.696825e+37	-4.683815e+37	-1.636183e+37	-1.164787e+37	-2.899635e+37
std	9.192504e+37	1.172412e+38	7.280357e+37	6.187245e+37	9.500991e+37
min	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38
25%	-1.140588e+00	-4.309719e+00	-8.100091e-03	-1.849221e-02	-1.228156e-01
50%	-6.413518e-01	-2.244232e+00	5.033697e-01	4.498120e-02	3.509227e-01
75%	-1.206105e-01	-1.088037e-01	9.708137e-01	1.090636e-01	7.238575e-01
max	4.711786e-01	1.332183e+00	2.077592e+00	3.150771e-01	1.427978e+00

	Mo	Geologia	Fallas	Anomalías espectrales \
count	1.241600e+04	12416.000000	12416.000000	12416.000000
mean	-6.084300e+37	2.569353	67.030631	42.452667
std	1.303967e+38	16.673995	131.316497	122.725045
min	-3.402823e+38	0.000000	0.000000	0.000000
25%	-3.173938e-01	0.000000	0.000000	0.000000
50%	4.396279e-01	0.000000	0.000000	0.000000
75%	1.068026e+00	0.000000	72.691292	0.000000
max	2.332602e+00	110.767677	487.627991	397.205963

	AOI
count	1.241600e+04
mean	-inf
std	inf
min	-1.797693e+308
25%	-1.797693e+308
50%	-1.797693e+308
75%	-1.797693e+308
max	1.000000e+00

Sin embargo, podemos notar que nuestra información de entrada tiene un problema. Nuestros depósitos NO son una variable continua, sino un una variable nominal.

7. Cambio de una variable continua a una variable nominal.

```
In [122]: print(df_filter.dtypes)
```

```
Au          float64
Ag          float64
Pb          float64
Zn          float64
Cu          float64
Mo          float64
Geologia    float64
Fallas      float64
Anomalias espectrales float64
AOI         float64
dtype: object
```

```
In [123]: #Cambiamos los AOI por una variable nominal
df.AOI=df_filter['AOI'].astype(str)
```

```
print(df_filter.dtypes)
```

```
Au          float64
Ag          float64
Pb          float64
Zn          float64
Cu          float64
Mo          float64
Geologia    float64
Fallas      float64
Anomalias espectrales float64
AOI         object
dtype: object
```

8. Verificación de las variables nominales a la librería Pandas.

```
In [124]: #Si se hizo correctamente, Pandas ya no puede calcular las estadísticas de una variable nominal
#y no se debería mostrar a continuación:
print(df_filter.describe())
```

	Au	Ag	Pb	Zn	Cu \
count	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04
mean	-2.696825e+37	-4.683815e+37	-1.636183e+37	-1.164787e+37	-2.899635e+37
std	9.192504e+37	1.172412e+38	7.280357e+37	6.187245e+37	9.500991e+37
min	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38
25%	-1.140588e+00	-4.309719e+00	-8.100091e-03	-1.849221e-02	-1.228156e-01
50%	-6.413518e-01	-2.244232e+00	5.033697e-01	4.498120e-02	3.509227e-01
75%	-1.206105e-01	-1.088037e-01	9.708137e-01	1.090636e-01	7.238575e-01
max	4.711786e-01	1.332183e+00	2.077592e+00	3.150771e-01	1.427978e+00

	Mo	Geologia	Fallas	Anomalias espectrales
count	1.241600e+04	12416.000000	12416.000000	12416.000000
mean	-6.084300e+37	2.569353	67.030631	42.452667
std	1.303967e+38	16.673995	131.316497	122.725045
min	-3.402823e+38	0.000000	0.000000	0.000000
25%	-3.173938e-01	0.000000	0.000000	0.000000
50%	4.396279e-01	0.000000	0.000000	0.000000
75%	1.068026e+00	0.000000	72.691292	0.000000
max	2.332602e+00	110.767677	487.627991	397.205963

```
In [125]: print(df_filter.head(5))
```

	Au	Ag	Pb	Zn	Cu	Mo	Geologia	Fallas	\
0	-0.281144	0.0	0.017937	0.000426	-0.009735	-0.230426	0.0	0.0	
1	-0.283690	0.0	0.016987	-0.000034	-0.009338	-0.269861	0.0	0.0	
2	-0.278217	0.0	0.016219	-0.000522	-0.008858	-0.298409	0.0	0.0	
3	-0.258834	0.0	0.015693	-0.001032	-0.008582	-0.312731	0.0	0.0	
4	-0.225928	0.0	0.015257	-0.001553	-0.008696	-0.317053	0.0	0.0	

	Anomalias espectrales	AOI
0	0.0	1.0
1	0.0	1.0
2	0.0	1.0
3	0.0	1.0
4	0.0	1.0

Ahora, todos los valores marcados como 0.0 en la columna AOI son lugares no conocidos "No data", por lo que debemos retirarlos de la tabla de entrenamiento. Luego, podemos utilizar el modelo entrenado para para clasificar estos puntos.

9. Entrenamiento de información.

```
In [126]: # df_filter_all=df_filter.drop(columns=['Felsic Intrusions', 'Intermediate Intrusions', 'Interm/Felsic Intrusions']) #All intrusions
# df_filter_separated=df_filter.drop(columns=['Intrusions']) #Intrusions Separated
```

```
In [127]: #Creemos un subgrupo llamado modelo donde solo incluyamos los lugares donde tenemos conocimiento
df_model=df_filter[df_filter['AOI']!='(0.0)']

#Checking our new data
print(df_model.describe())
```

	Au	Ag	Pb	Zn	Cu \
count	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04	1.241600e+04
mean	-2.696825e+37	-4.683815e+37	-1.636183e+37	-1.164787e+37	-2.899635e+37
std	9.192504e+37	1.172412e+38	7.280357e+37	6.187245e+37	9.500991e+37
min	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38	-3.402823e+38
25%	-1.140588e+00	-4.309719e+00	-8.100091e-03	-1.849221e-02	-1.228156e-01
50%	-6.413518e-01	-2.244232e+00	5.033697e-01	4.498120e-02	3.509227e-01
75%	-1.206105e-01	-1.088037e-01	9.708137e-01	1.090636e-01	7.238575e-01
max	4.711786e-01	1.332183e+00	2.077592e+00	3.150771e-01	1.427978e+00

	Mo	Geologia	Fallas	Anomalias espectrales
count	1.241600e+04	12416.000000	12416.000000	12416.000000
mean	-6.084300e+37	2.569353	67.030631	42.452667
std	1.303967e+38	16.673995	131.316497	122.725045
min	-3.402823e+38	0.000000	0.000000	0.000000
25%	-3.173938e-01	0.000000	0.000000	0.000000
50%	4.396279e-01	0.000000	0.000000	0.000000
75%	1.068026e+00	0.000000	72.691292	0.000000
max	2.332602e+00	110.767677	487.627991	397.205963

de 1.24 millones de pixeles que teníamos inicialmente, solo 12416 tienen información. Ahora, dentro de los puntos conocidos tenemos los puntos 1 (Depositos minerales) y 0 (No data) con los que queremos entrenar el modelo. Creamos las matrices que contienen solo la información de los puntos de entrenamiento. Nuestra variable "y" será nuestro objetivo (Depositos minerales), los valores de "x" serán todos nuestros atributos de entrada. A fin de cuentas, tendremos una especie de problema matricial donde G será definido por nuestro modelo.

10. Elaboración de matriz de correlación.

```
In [128]: import seaborn as sn
plt.figure(figsize=(15,10))
sn.heatmap(df_model.corr(), annot=True)
plt.show()
```


11. Uso del PCA (Opcional)

```
In [129]: #Creating our target variable and feature variables from the dataframe in preperation for machine Learning
y = np.array(df_model.pop('AOI').values)
X = np.array(df_model)
```

PCA

A veces nos encontramos con casos donde tenemos una gran cantidad de atributos. Como ya sabemos que entre más atributos mayor es la complejidad de nuestros modelos (y mayor esfuerzo computacional necesitamos) resulta útil utilizar el tipo de aprendizaje automático no supervisado que comunmente se llama PCA. Acá veremos como aplicarlo de forma rápida a nuestros datos para después utilizar el resultado dentro de nuestros algoritmos de ML.

```
In [130]: import pandas as pd
from sklearn.decomposition import PCA

# Suponiendo que df_model es tu DataFrame
pca = PCA(0.95) # Manteniendo el 95% de la varianza
principalComponents = pca.fit_transform(df_model)

# Determina el número de componentes
num_componentes = principalComponents.shape[1]

# Genera nombres de columnas basados en el número de componentes
nombres_columnas = [f'PCA{i+1}' for i in range(num_componentes)]

# Crea el DataFrame con el número correcto de columnas
principalDf = pd.DataFrame(data=principalComponents, columns=nombres_columnas)

# Ahora principalDf debería crearse correctamente
print(principalDf.head())
```

	PCA1	PCA2	PCA3
0	-8.696107e+37	-1.687728e+37	3.120700e+36
1	-8.696107e+37	-1.687728e+37	3.120700e+36
2	-8.696107e+37	-1.687728e+37	3.120700e+36
3	-8.696107e+37	-1.687728e+37	3.120700e+36
4	-8.696107e+37	-1.687728e+37	3.120700e+36

12. Transformación de datos

Transformación de datos

Primero, debemos hacer una separación de un set de entnamiento y un set de comprobación. En este caso definimos que 75% de los depósitos conocidos serán usados para entrenar el algoritmo, el 25% para validación. Esta separación es aleatoria.

Luego, antes de proceder con el modelo, debemos escalar todas las variables de entrada para poderlas comparar. Esto significa aplicar una transformación que nos permita tener la misma escala para todas las variables.

```
In [131]: from sklearn.preprocessing import MinMaxScaler
#Scaling our features
scaler = MinMaxScaler()
X = scaler.fit_transform(X)
```

Y ahora si podemos seprar nuestros datos en entrenamiento y validación:

```
In [132]: #Importing the train/test and scaling modules
from sklearn.model_selection import train_test_split
# from sklearn.preprocessing import MinMaxScaler

#Splitting the data 75/25% into training and test data and then scaling the features
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, stratify=y)
```

```
In [133]: X_train
```

```
Out[133]: array([[0., 0., 0., ..., 1., 0., 1.],
 [1., 1., 1., ..., 0., 0., 0.],
 [1., 1., 1., ..., 0., 0., 0.],
 ...,
 [1., 1., 1., ..., 0., 0., 0.],
 [1., 1., 1., ..., 0., 0., 0.],
 [1., 1., 1., ..., 0., 0., 0.]])
```

13. Importación de librerías de Machine Learning para Random Forests.

Machine Learning

Como siempre debemos comenzar con la importación de las librerías de interés (tengamos en cuenta que nuestro ambiente virtual ya tiene instalado sklearn).

```
In [134]: from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
```

El primer modelo que utilizaremos sera el bosque aleatorio (random forests). En principio, no existe una ley que establezca qué modelo es el que mejor resultado tendrá y el usuario debe ser el que defina qué modelo se utilizará. Sin embargo, las relaciones en geología no suelen ser ni lineales ni excluyentes, lo que significa que en principio los atributos de entrada tienen relación entre si. Con esta idea, los bosques aleatorios deben entregar resultados menos precisos que las redes neuronales. La comparación entre modelos la haremos al final para poder verificar esta hipótesis.

14. Combinación de atributos.

```
In [135]: #Importing the grid search module
from sklearn.model_selection import GridSearchCV

#Creating our initial test classifier
RF = RandomForestClassifier()

#Selecting our algorithm variables that will be tested on our dataset and running them through a grid search
grid_param_RF = { 'n_estimators': [100,200,300,500,800,1000], 'criterion': ['gini', 'entropy'], 'bootstrap': [True, False]}
gd_sr_RF = GridSearchCV(estimator=RF, param_grid=grid_param_RF, scoring='accuracy', cv=10, n_jobs=-1)
gd_sr_RF.fit(X_train, y_train)

#Assigning a variable to the best paramaters and the classification accuracy
best_parameters_RF = gd_sr_RF.best_params_
best_result_RF = gd_sr_RF.best_score_

print(best_parameters_RF)
print(best_result_RF)

{'bootstrap': True, 'criterion': 'gini', 'n_estimators': 100}
0.9961341121042953
```

Un valor total de 99.61% para el uso de toda la información que tenemos, no es bajo y tampoco muy alto. ¿Por qué? Suponiendo en principio que obtuvimos un buen resultado, ya tenemos un modelo de ML para predecir nuestros resultados:

```
In [136]: #Creating a new model with the desired parameters
model = RandomForestClassifier(bootstrap=best_parameters_RF["bootstrap"], criterion= best_parameters_RF["criterion"], n_estimator

#Fitting the model to the training data
clf = model.fit(X_train, y_train)
```

15. Predicción de mapa para Random Forests.

Ahora podemos clasificar todos los puntos de nuestro mapa! Nuevamente haciendo la apropiada transformación al igual que la predicción.

```
In [137]: #Obtain the predictions of the model in the training data
predictions = model.predict(X_test)
```

```
In [138]: print('Mean Absolute Error:', mean_absolute_error(y_test, predictions))
print('Mean Squared Error:', mean_squared_error(y_test, predictions))
print('Root Mean Squared Error:', np.sqrt(mean_squared_error(y_test, predictions)))
```

16. Verificación del modelo a través del MSE (Mean Square Error).

MSE

Para determinar que tan bueno es nuestro modelo utilizando la solución escogida (en este caso "bootstrap" y el criterio "entropy") podemos graficar el promedio del error cuadrado mientras aumentamos las iteraciones. Tener en cuenta que cada ciclo de este for se entrena un nuevo modelo, por lo que no se va a obtener una curva homogénea. Para mirar la estabilidad del modelo, se debe construir el modelo de RF desde cero dentro de cada iteración... Sin embargo, aún podemos ver en general si el modelo puede mejorarse con mayores iteraciones a las propuestas anteriormente.

```
In [139]: iterations = np.arange(1, 10, 5)
result = []

for i in iterations:
    clf1 = RandomForestClassifier(bootstrap=best_parameters_RF["bootstrap"], criterion=best_parameters_RF["criterion"], n_estimators=best_parameters_RF["n_estimators"])
    clf1.fit(X_train, y_train)
    predict = clf1.predict(X_test)
    result.append(mean_squared_error(y_test, predict))
```

17. Elaboración de la matriz de confusión para RF.

```
In [141]: from sklearn.metrics import confusion_matrix
from sklearn import metrics

disp = metrics.plot_confusion_matrix(model, X_test, y_test)
disp.figure_.suptitle("Confusion Matrix RFC")
print(accuracy_score(y_test, predictions))
print(confusion_matrix(y_test, predictions))
print(classification_report(y_test, predictions))
```

18. Verificación de las variables más importantes del modelo.

```
In [144]: importance = model.feature_importances_
k = plt.figure(figsize=(7,7))
plt.bar([x for x in range(len(importance))], importance)
plt.title('Decision tree classifier Feature Importance Scores')
plt.show()
list(enumerate(df_model.columns))
```

19. Desbalance de datos.

```
In [145]: from collections import Counter
from imblearn.under_sampling import RandomUnderSampler
from imblearn.over_sampling import SMOTE
from sklearn.metrics import roc_auc_score

# summarize class distribution
clf_before = model.fit(X_train, y_train)
print("Before undersampling: ", Counter(y_train))
print("ROC AUC score for untampered data: ", roc_auc_score(y_test, predictions))

# define undersampling strategy
undersample = RandomUnderSampler(sampling_strategy='majority')
X_train_under, y_train_under = undersample.fit_resample(X_train, y_train)
print("After undersampling: ", Counter(y_train_under))
clf_under = model.fit(X_train_under, y_train_under)
pred_under = clf_under.predict_proba(X_test)
y_prob_split1 = np.split(pred_under, 2, axis=1)
print("ROC AUC score for undersampled data Undersampling: ", roc_auc_score(y_test, y_prob_split1[1]))

# define oversampling strategy
SMOTE = SMOTE()
X_train_SMOTE, y_train_SMOTE = SMOTE.fit_resample(X_train, y_train)
print("After oversampling: ", Counter(y_train_SMOTE))
clf_SMOTE = model.fit(X_train_SMOTE, y_train_SMOTE)
pred_SMOTE = clf_SMOTE.predict_proba(X_test)
y_prob_split2 = np.split(pred_SMOTE, 2, axis=1)
print("ROC AUC score for oversampled SMOTE data: ", roc_auc_score(y_test, y_prob_split2[1]))

Before undersampling: Counter({'-1.797693e+308': 9276, '1.0': 36})
ROC AUC score for untampered data: 0.5
After undersampling: Counter({'-1.797693e+308': 36, '1.0': 36})
ROC AUC score for undersampled data Undersampling: 0.7780697498921948
After oversampling: Counter({'-1.797693e+308': 9276, '1.0': 9276})
```

20. Elaboración de la curva ROC (Receiver Operating Characteristic) para Random Forests.

```
In [146]: from sklearn.metrics import roc_curve
def roc_curve_plot(y_test, pred_proba_c1, pred_under, pred_SMOTE):
    fprs, tprs, thresholds = roc_curve(y_test, pred_proba_c1, pos_label='2.0')
    fprs1, tprs1, thresholds1 = roc_curve(y_test, pred_under, pos_label='2.0')
    fprs2, tprs2, thresholds2 = roc_curve(y_test, pred_SMOTE, pos_label='2.0')

    plt.plot(fprs, tprs, label='ROC_origin')
    plt.plot(fprs1, tprs1, label='ROC_under')
    plt.plot(fprs2, tprs2, label='ROC_SMOTE')
    plt.plot([0,1], [0,1], 'k--', label='Random')

    start, end = plt.xlim()
    plt.xticks(np.round(np.arange(start, end, 0.1), 2))
    plt.xlim(0,1); plt.ylim(0,1)
    plt.xlabel('FPR (1 - Sensitivity)'); plt.ylabel('TPR (Recall)')
    plt.legend(); plt.grid()

roc_curve_plot(y_test, predictionsRF[:,1], pred_under[:,1], pred_SMOTE[:,1])
```

21. Importación de librerías de Machine Learning para Artificial neural networks, combinación de atributos, predicción de mapa y verificación.

ANN

```
In [147]: from sklearn.neural_network import MLPClassifier

#Creating our initial test classifier
ANN = MLPClassifier()

#Selecting our algorithm variables that will be tested on our dataset and running them through a grid search
grid_param_ANN = { 'solver': ['lbfgs', 'sgd', 'adam'], 'max_iter': [50,100,125,150,175,200,250,500,1000], 'activation': ['identity', 'tanh', 'relu'] }
gd_sr_ANN = GridSearchCV(estimator=ANN, param_grid=grid_param_ANN, scoring='accuracy', cv=10, n_jobs=-1)
gd_sr_ANN.fit(X_train, y_train)

#Assigning a variable to the best paramaters and the classification accuracy
best_parameters_ANN = gd_sr_ANN.best_params_
best_result_ANN = gd_sr_ANN.best_score_

print(best_parameters_ANN)
print(best_result_ANN)

{'activation': 'identity', 'max_iter': 50, 'solver': 'lbfgs'}
0.9961341121042953

In [148]: clfann = MLPClassifier(solver=best_parameters_ANN["solver"], random_state=12, activation= best_parameters_ANN["activation"], max_iter=best_parameters_ANN["max_iter"])
clfann.fit(X_train, y_train)

Out[148]: MLPClassifier(activation='identity', max_iter=50, random_state=12, solver='lbfgs')

In [149]: #Obtain the predictions of the model in the training data
predictionsann = clfann.predict(X_test)
```

22. Elaboración de matriz de confusión para ANN.

```
In [150]: disp = metrics.plot_confusion_matrix(clfann, X_test, y_test)
disp.figure_.suptitle("Confusion Matrix ANN")
print(accuracy_score(y_test, predictionsann))
print(confusion_matrix(y_test, predictionsann))
print(classification_report(y_test, predictionsann))
```

23. Verificación de variables y desbalances de datos.

```
In [152]: %%capture --no-display
iterations1 = np.arange(1, 500, 5)
result1 = []

for i in iterations1:
    clf2 = MLPClassifier(solver=best_parameters_ANN["solver"], activation= best_parameters_ANN["activation"], max_iter = i)
    clf2.fit(X_train, y_train)
    predict = clf2.predict(X_test)
    result1.append(metrics.mean_squared_error(y_test, predict))

#Setting the image size
plt.figure(num=None, figsize=(16, 12))

plt.plot(iterations1, result1)
```

24. Elaboración de la curva ROC (Receiver Operating Characteristic) para ANN

```
In [153]: def roc_curve_plot1(y_test, pred_proba_c1, pred_under):
fprs, tprs, thresholds = roc_curve(y_test, pred_proba_c1, pos_label='2.0')
fprs1, tprs1, thresholds1 = roc_curve(y_test, pred_under, pos_label='2.0')

plt.plot(fprs, tprs, label='ROC_RFC')
plt.plot(fprs1, tprs1, label='ROC_ANN')
plt.plot([0,1], [0,1], 'k--', label='Random')

start, end = plt.xlim()
plt.xticks(np.round(np.arange(start, end, 0.1),2))
plt.xlim(0,1); plt.ylim(0,1)
plt.xlabel('FPR (1 - Sensitivity)'); plt.ylabel('TPR (Recall)')
plt.legend(); plt.grid()

roc_curve_plot1(y_test, predictionsRF[:,1], predictionsANN[:,1])
```

25. Comparación de ROC entre RF y ANN.

```
In [154]: print("Área bajo la curva para el bosque aleatorio: ", roc_auc_score(y_test, predictions))
print("Área bajo la curva para la red neuronal: ", roc_auc_score(y_test, predictionsann))
```

26. Aplicación de los modelos a toda la zona de interés.

Aplicar los modelos a toda la zona de interés

```
In [155]: #Splitting the initial dataframe
y_new = np.array(df.pop('AOI').values)
X_new = np.array(df)

#Scaling it to match the model input
X_new = scaler.fit_transform(X_new)
```

```
In [156]: #Aquí creamos las matrices de predicción usando nuestro modelo escogido tanto para RF como ANN
y_pred = model.predict(X_new)
y_prob = model.predict_proba(X_new)

predictionsann = clfann.predict(X_new)
probability = clfann.predict_proba(X_new)

print(y_pred.shape, y_prob.shape)

(12416,) (12416, 2)
```


27. Reformación de matrices para el cambio de valores.

Reformación de las matrices

```
In [157]: #Splitting the y_prob array into two separate 1 dimensional arrays RFC
y_prob_split = np.split(y_prob, 2, axis=1)

y_prob_1 = y_prob_split[0]
y_prob_2 = y_prob_split[1]

#Splitting the y_prob array into two separate 1 dimensional arrays ANN
y_prob_splitann = np.split(probability, 2, axis=1)

y_prob_1ann = y_prob_splitann[0]
y_prob_2ann = y_prob_splitann[1]

print(y_prob_1.shape)
print(y_prob_1ann.shape)

(12416, 1)
(12416, 1)
```

Pero primero, necesitaremos asignar un valor sin datos.

```
In [158]: y_prob_2[tuple(noData)] = -1
y_prob_2ann[tuple(noData)] = -1
```

```
In [159]: #Reshaping the 1d arrays to be representable as an image RFC
y_pred = y_pred.reshape(128, 97).astype(float)
y_prob_1 = y_prob_1.reshape(128, 97)
y_prob_2 = y_prob_2.reshape(128, 97)

#Reshaping the 1d arrays to be representable as an image ANN
predictionsann = predictionsann.reshape(128, 97).astype(float)
y_prob_1ann = y_prob_1ann.reshape(128, 97)
y_prob_2ann = y_prob_2ann.reshape(128, 97)
```

28. Georreferenciación de metadatos.

Georeferenciamiento

Ahora extraigamos los metadatos de la geología. Tener en cuenta que toda la información de entrada debe tener la misma proyección para asegurar que la información sea consistente.

```
In [160]: #Accessing the metadata of the existing 'Geology' image
with rio.open('./Python/Geologia_local_f.tif') as src:
    study_area_ras = src.read()
    study_area_meta = src.profile

print(study_area_meta)

{'driver': 'GTiff', 'dtype': 'float32', 'nodata': 0.0, 'width': 128, 'height': 97, 'count': 1, 'crs': CRS.from_epsg(32718), 'transform': Affine(111.33864621649491, 0.0, 407568.217714001, 0.0, -109.6257439670105, 8885320.9892596), 'blockysize': 128, 'tiled': False, 'interleave': 'band'}
```

```
In [161]: #Viewing the transform and crs
study_area_transform = study_area_meta["transform"]
study_area_crs = study_area_meta["crs"]

print(study_area_transform)
print(study_area_crs)

| 111.34, 0.00, 407568.22|
| 0.00,-109.63, 8885320.99|
| 0.00, 0.00, 1.00|
EPSG:32718
```

```
In [162]: #Ensuring the count value is 1 (1 band of pixel data) and assigning the data as float to be compatible with the
#new data
study_area_meta['count'] = 1
study_area_meta['dtype'] = "float64"

print(study_area_meta)

{'driver': 'GTiff', 'dtype': 'float64', 'nodata': 0.0, 'width': 128, 'height': 97, 'count': 1, 'crs': CRS.from_epsg(32718), 'transform': Affine(111.33864621649491, 0.0, 407568.217714001, 0.0, -109.6257439670105, 8885320.9892596), 'blockysize': 128, 'tiled': False, 'interleave': 'band'}
```

29. Guardando datos de los modelos tratados.

```
In [163]: #Guardemos Los datos de los bosques aleatorios

with rio.open('./Python/yy_pred.tif', 'w', **study_area_meta) as dst1:
    dst1.write(y_pred, 1)

with rio.open('./Python/yy_prob_1.tif', 'w', **study_area_meta) as dst2:
    dst2.write(y_prob_1, 1)

with rio.open('./Python/yy_prob_2.tif', 'w', **study_area_meta) as dst3:
    dst3.write(y_prob_2, 1)

In [164]: #Guardemos Los resultados de ANN

with rio.open('./Python/yy_predann.tif', 'w', **study_area_meta) as dst1:
    dst1.write(predictionsann, 1)

with rio.open('./Python/yy_probann_1.tif', 'w', **study_area_meta) as dst2:
    dst2.write(y_prob_1ann, 1)

with rio.open('./Python/yy_probann_2.tif', 'w', **study_area_meta) as dst3:
    dst3.write(y_prob_2ann, 1)
```

30. Elaboración del mapa predictivo RF

```
In [165]: #Importing required packages
from mpl_toolkits.axes_grid1 import make_axes_locatable
from pylab import *

#Setting the image size
plt.figure(num=None, figsize=(16, 12), dpi=80, facecolor='w', edgecolor='k')

#Setting the eastings and northings boundary
bounds = [529000, 550000, 6712000, 6733000]

#Loading the image data and setting extents
plt.title('Area of Interest Probability RFC')
ax = plt.gca()
im = ax.imshow(y_prob_2, interpolation='nearest', extent=bounds, vmin=0, vmax=1)

#Plotting colorbar
divider = make_axes_locatable(ax)
cax = divider.append_axes("right", size="10%", pad=0.05)
plt.colorbar(im, cax=cax)

#Visualising the image
plt.show()
```

31. Elaboración del mapa

Mapa predictivo ANN ¶

```
In [166]: #Setting the image size
plt.figure(num=None, figsize=(16, 12), dpi=80, facecolor='w', edgecolor='k')

#Setting the eastings and northings boundary
bounds = [529000, 550000, 6712000, 6733000]

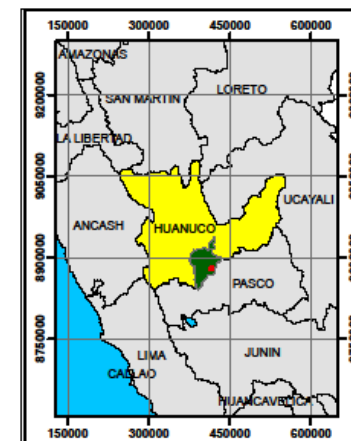
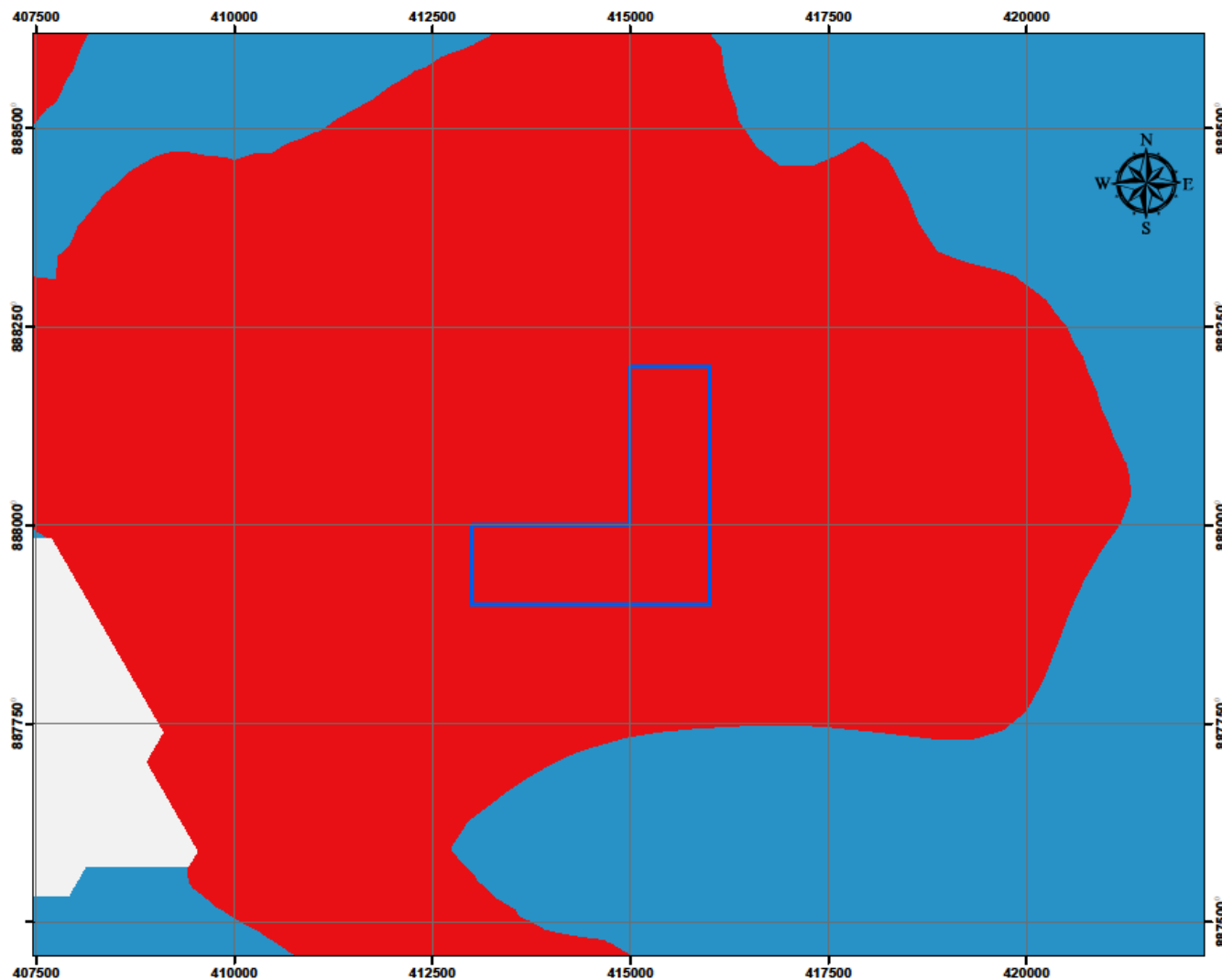
#Loading the image data and setting extents
plt.title('Area of Interest Probability ANN')
ax = plt.gca()
im = ax.imshow(y_prob_2ann, interpolation='nearest', extent=bounds, vmin=0, vmax=1)

#Plotting colorbar
divider = make_axes_locatable(ax)
cax = divider.append_axes("right", size="10%", pad=0.05)
plt.colorbar(im, cax=cax)

#Visualising the image
plt.show()
```


ANEXO V. MAPA DE PREDICCIÓN MINERAL

MAPA DE PREDICCIÓN MINERAL RF



LEYENDA

 Área de estudio

Prediction Map

 0 (Zona no favorable)

 1 (Zona favorable)

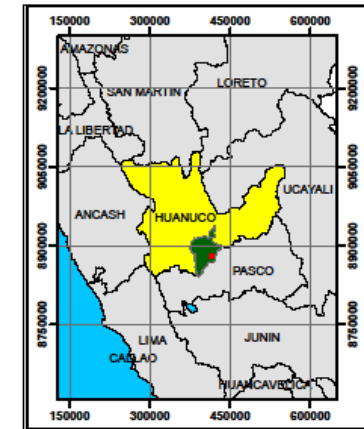
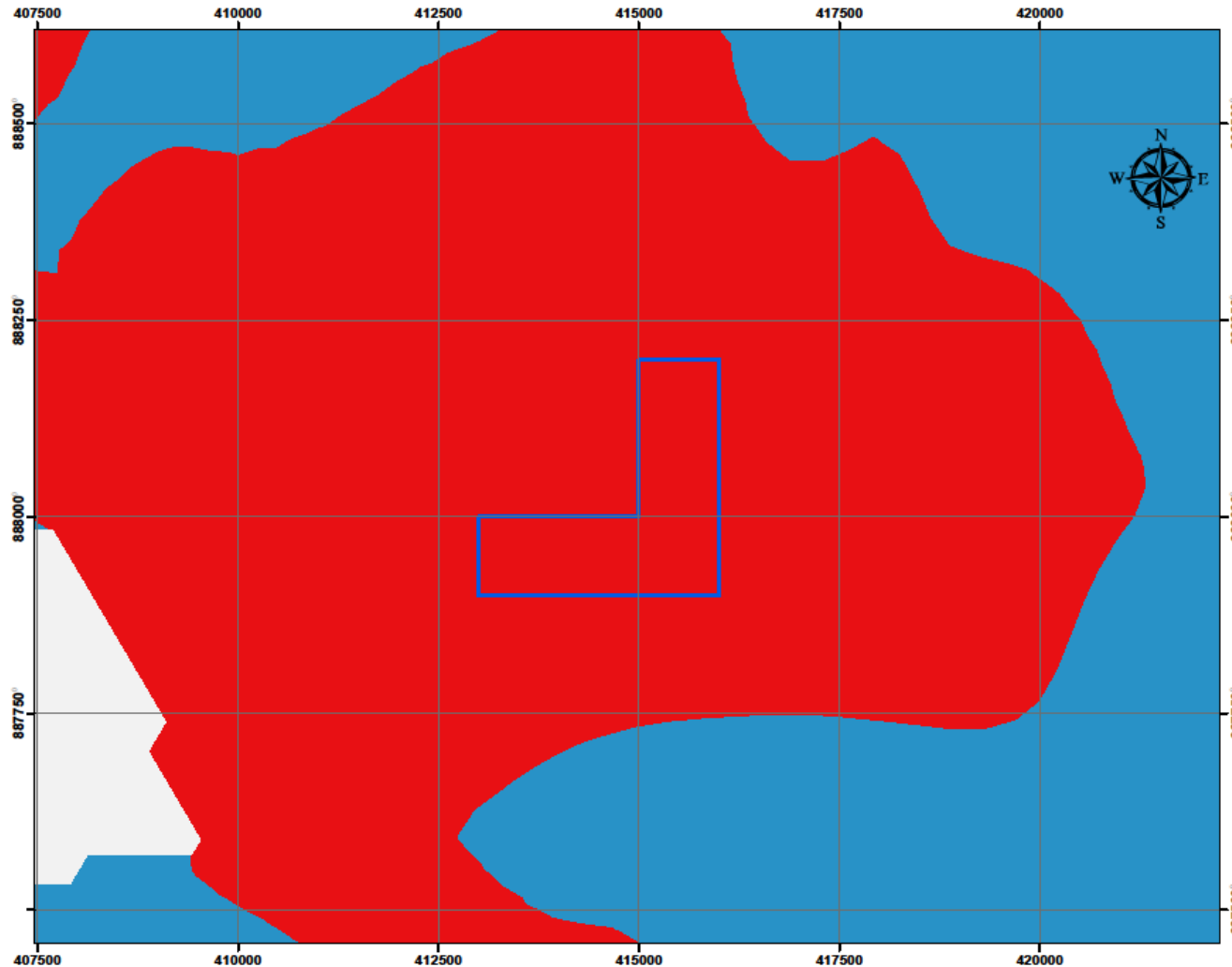
0 1,250 2,500 5,000 Meters

1:50,000



UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN					
FACULTAD DE INGENIERÍA					
ESCUELA PROFESIONAL DE INGENIERÍA GEOLOGICA					
NOMBRE DEL PROYECTO:	REALIZADO POR:	PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023	Bach. Gianmarco Castro Melendez	MAPA DE PREDICCIÓN MINERAL RF	PROYECCIÓN: UTM	1/50000	22
	REVISADO POR:		ZONA: 18 SUR	FECHA:	
	Mg. Eder Guido ROBLES MORALES		COORDENADAS: WGS - 84	2/08/2024	

MAPA DE PREDICCIÓN MINERAL ANN



LEYENDA

 Área de estudio

Prediction Map

 0 (Zona no favorable)

 1 (Zona favorable)

0 1,250 2,500 5,000 Meters

1:50,000

	UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN							
	FACULTAD DE INGENIERIA							
	ESCUELA PROFESIONAL DE INGENIERIA GEOLOGICA							
	NOMBRE DEL PROYECTO:		REALIZADO POR:		PLANO:	SISTEMA DE COORDENADAS	ESCALA:	LAMINA:
	Aplicación del Machine Learning y el pronóstico de mineralización en la concesión minera Gus del Aire, Huánuco 2023		Bach. Gianmarco Castro Melendez		MAPA DE PREDICCIÓN MINERAL ANN	PROYECCION: UTM	1/50000	23
REVISADO POR:		Mg. Eder Guido ROBLES MORALES		ZONA: 18 SUR		FECHA:		
				COORDENADAS: WGS - 84		2/08/2024		