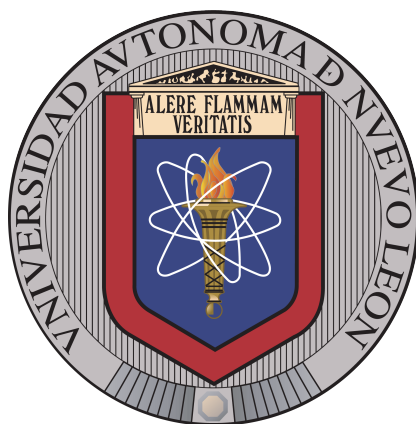


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICO



TÍTULO DE LA TESIS

POR

JUAN JESÚS TORRES SOLANO

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE

MAESTRÍA EN CIENCIA DE DATOS

MARZO 2025

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICO

Los miembros del Comité de Tesis recomendamos que la Tesis «Título de la tesis», realizada por el alumno Juan Jesús Torres Solano, con número de matrícula 2173262, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Ciencia de Datos.

El Comité de Tesis

Dr. José Anastasio Hernández Saldaña
Asesor

Nombre del revisor C
Revisor

Nombre del revisor D
Revisor

Vo. Bo.

Dra. Azucena Yolojóchitl Ríos Mercado
Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, marzo 2025

*Aquí puedes poner tu dedicatoria
si es que tienes una.*

ÍNDICE GENERAL

Agradecimientos	VIII
1. INTRODUCCIÓN	1
2. DELIMITACIÓN Y PLANTEAMIENTO DEL PROBLEMA	2
3. JUSTIFICACIÓN	3
4. FORMULACIÓN DE OBJETIVOS	4
4.1. Generales	4
4.2. Específicos	4
5. MARCO TEÓRICO	5
5.1. Geofísica y Geoelectrica	5
5.1.1. Definición de Geofísica	5
5.1.2. Resistividad de la Tierra	5
5.1.3. Sondeo Eléctrico Vertical (SEV)	7
5.2. Adquisición de Datos Geofísicos	11
5.2.1. Intervalo de Muestreo en SEV	12
5.2.2. Proceso de Adquisición In Situ	13

ÍNDICE GENERAL	v
5.3. Machine Learning (ML) en la Geofísica	13
5.4. Random Forests	15
5.4.1. Propiedades de Random Forests	15
5.4.2. Robustez y Precisión	15
5.4.3. Random Forests para Clasificación y Regresión	15
5.4.4. Aplicaciones de Random Forests en Geofísica	15
6. RESULTADOS Y CONCLUSIONES	16
A. Apéndice I	17

ÍNDICE DE FIGURAS

5.1. Estructura atómica del oro	6
5.2. Configuración general de electrodos	8
5.3. Esquema de la contribución de la respuesta eléctrica	10
5.4. Esquema del arreglo Wenner	11
5.5. Esquema del arreglo Schlumberger	11
5.6. Esquema del arreglo Dipolo-dipolo	11

ÍNDICE DE TABLAS

AGRADECIMIENTOS

Aquí puedes poner tus agradecimientos. (No olvides agradecer a tu comité de tesis, a tus profesores, a la facultad y a CONACyT en caso de que hallas sido beneficiado con una beca).

CAPÍTULO 1 INTRODUCCIÓN

que es la geofísica:

La exploración geofísica consiste en un conjunto de metodologías que a través de la medición de propiedades petrofísicas del subsuelo es

el geoelectrica

para que sirve

oportunidades de la ciencia de datos en el análisis en tiempo real de la respuesta

CAPÍTULO 2

DELIMITACIÓN Y PLANTEAMIENTO DEL PROBLEMA

Durante un trabajo de prospección geofísica, al realizar adquisición de datos geoelectricos *in situ*, no es posible conocer el resultado del trabajo hasta una vez realizado procesos de inversión de datos resultado de prospección, por lo que no se tiene certeza de si una lectura presenta inconsistencias o si forma parte de una respuesta esperada para determinado resultado.

El problema es no contar con un modelo que pueda anticipar el resultado de un modelo de inversión geológico, o al menos aproximar un resultado que pueda tener coherencia geológica en interpretación

Como primera aproximación a una solución se considerara un primer supuesto, en el cual se cuenta con respuestas de sondeos eléctricos verticales, más simple pero no menos complejo de interpretar, al contar con un solo segmento de señal, dentro de esta señal podemos encontrar distintas unidades geológicas, profundidad de acuífero, espesores de unidades y profundidad de exploración.

La propuesta de solución es mediante Aprendizaje automático (ML, por sus siglas en inglés), empleando una técnica de aprendizaje que permita identificar patrones que estén asociados a la respuesta de un modelo de inversión, de esta manera obtener un modelo pronóstico de la inversión, sirviendo de guía para incrementar la densidad de muestreo para mejorar el modelo y respuestas.

CAPÍTULO 3 JUSTIFICACIÓN

Existen múltiples aplicaciones de ML y DL en el procesamiento, modelado e interpretación geofísica (INSERTAR REFERENCIAS), así como aplicaciones directas por software en las disciplinas mas comerciales (Sísmica de exploración de hidrocarburos PETREL, Shlumberger, INSERTAR REFERENCIA).

Es de esperar que al implementar ciencias de datos en diversas áreas de la ciencia, se optimicen aspectos básicos, mejorando la analítica e interpretación de resultados, o bien, modelando predicciones de tendencias.

En este sentido, es oportuna la implementación de modelos de regresión y clasificación durante la adquisición de datos, en particular al ejecutar muestreo de Sondeo Eléctrico Vertical (SEV, VES por sus siglas en inglés), el proceso de adquisición *in situ* consta de un intervalo de muestreo predefinido, el cual está acotado de acuerdo al objetivo de exploración, este análisis previo es un factor determinante, debido a que establecer un intervalo de muestreo correcto permite identificar el objetivo, unidades geológicas, acuíferos, fallas, fracturas, estructuras antropogenicas, etc.. , es decir mantener un intervalo de muestreo menor a la frecuencia de ocurrencia del objetivo de estudio, por lo que el éxito de la exploración dependerá en su totalidad de la planeación previa de la adquisición.

De manera que un modelo que permita clasificar las lecturas y generar una regresión para proponer muestreos adicionales *in situ* tendría la ventaja de optimizar y mejorar la calidad de la adquisición y muestreo de las frecuencias deseadas.

CAPÍTULO 4

FORMULACIÓN DE OBJETIVOS

4.1 Generales

- Establecer un modelo de entrenamiento de regresión mediante Random Forests que genere intervalos de muestreo en SEV's al identificar contrastes en los datos.

4.2 Específicos

- Generar modelos geoelectricos con unidades geológicas de terrenos ya caracterizados mediante SEV's.
- Establecer un delo de clasificación y regresión utilizando Random Forests empleando modelos geoelectricos.
- Generar un modelo predictivo que permita clasificar datos geoelectricos durante la adquisición de SEV's *in situ*.

CAPÍTULO 5 MARCO TEÓRICO

5.1 Geofísica y Geoeléctrica

5.1.1 Definición de Geofísica

En términos generales la geofísica es la aplicación de los principios físicos de la materia en el estudio del planeta Tierra, o cual quier otro cuerpo celeste, desde el campo magnético, pasando por los fenómenos atmosféricos al medio solido del subsuelo, hasta las profundidades del núcleo interno planetario, ya sea que se empleé una fuente natural como la propagación de ondas elásticas generadas por sismicidad, ó bien, la inducción de campo electromagnético de fuente controlada (Parasnis, 2012; Reynolds, 2011; Lay y Wallace, 1995).

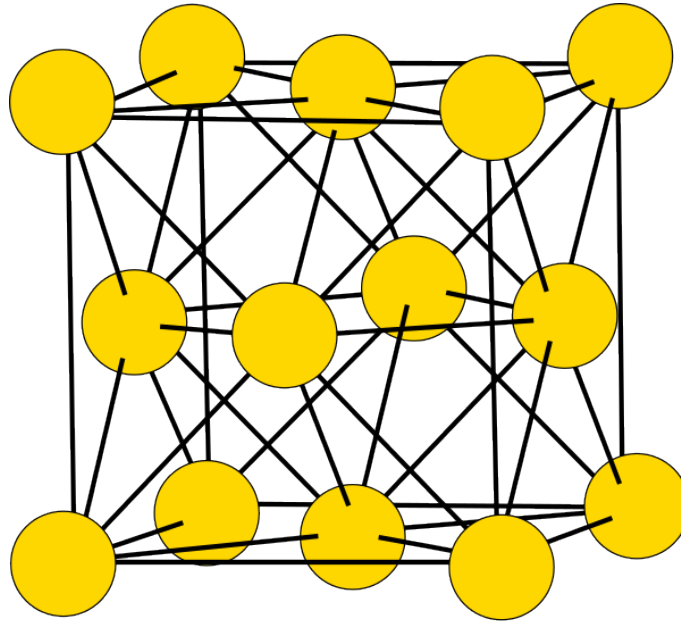
El nacimiento de la geofísica es relativamente reciente, la primera prospección geoeléctrica data de 1830 realizados por Fox (1830) en Cornwall, Reino Unido, donde aplico técnicas de Self-Potential en exploración de mineralización de sulfuro en vetas, la medición del potencial natural resulto altamente efectiva para la prospección de este tipo de mineralizaciones ya que su anomalía se caracterizaba por presentar una respuesta muy marcada con respecto al medio (Reynolds, 2011; Revil y Jardani, 2013).

5.1.2 Resistividad de la Tierra

De manera general la materia presenta características definidas a partir de los elementos que la integran, en primer orden la configuración atómica establece las

propiedades físicas corresponden a la estructura de electrones, protones y neutrones que presentan los átomos; a su vez, las moléculas pueden estar conformadas por una clase específica de átomos (moléculas homonucleares) o por conjuntos de diferentes tipos (compuestos), cuya conformación depende de factores físico-químicos (Tiab y Donaldson, 2024).

La configuración molecular inorgánica presente en la materia, definirá el tipo de estructura cristalina (mineral) que formarán, en conjunto; esta configuración cristalina es la que encontramos en el medio geológico conformando los minerales que componen la estructura mineral de una unidad geológica (ver figura 5.1) (Gandhi y Sarkar, 2016; Tiab y Donaldson, 2024).



Estructura del Oro

FIGURA 5.1: Esquema de la estructura atómica de oro que conforma la cristalización octahedral, modificado de Sorrell (1973)

Los métodos Geoelectrónicos se clasifican en dos grupos, métodos pasivos y de inducción, los primeros corresponden a aquellos en los que se mide el potencial eléctrico

natural, usualmente medido en mili volts, en donde se requiere de electrodos no polarizables para tener medidas lo mas claras posibles; mientras que los métodos de inducción emplean un arreglo de electrodos, o inductores de campo electromagnéticos, mediante los cuales se induce un campo eléctrico al subsuelo, calculando la diferencia de potencia eléctrica en el medio, o bien, el decaimiento de la polarización inducida (Revil y Jardani, 2013; Reynolds, 2011; Igboama *et al.*, 2023).

Los métodos de inducción, Sondeo Eléctrico verticales (VES, por sus siglas en inglés), Tomografía de Resistividad Eléctrica (ERT, por sus siglas en inglés), polarización inducida (IP, por sus siglas en inglés), presentan una gran ventaja ya que no dependen del medio para poder realizar una lectura, ademas de poder realizarlos en cualquier momento, manteniendo el equipo en condiciones de operación, y pode diseñar arreglos de adquisición que nos permitan tener un muestreo tan amplio o limitado como sea conveniente, solo limitados por el alcance y potencia de los equipos empleados. Por otro lado su interpretación presenta un alta ambigüedad, solo acotado por la cantidad de referencias que puedan cruzarse para robustecer el modelo geológico y de inversión, y así poder llegar a una interpretación satisfactoria (Reynolds, 2011; Igboama *et al.*, 2023).

El método de prospección geoelectrica, en especifico el SEV y la TRE, consiste en determinar la distribución de resistividades del subsuelo, de manera que se pueda establecer una correlación entre la resistividad y un modelo ajustado a la realidad geológica-estructural, geotécnica o geohidrológica del objeto de estudio.

5.1.3 Sondeo Eléctrico Vertical (SEV)

Los SEV corresponden al método de mas rápida ejecución y económicamente mas accesible, por lo que es ampliamente empleado para solucionar problemas de ingeniería, minería, geotecnia, monitoreo e impacto ambiental y abastecimiento de Aguas potable; siendo de gran utilidad en la exploración de hidrogeologica ya que la respuesta resistiva de un medio saturado permite establecer diferencias concisas y

discriminar entre agua dulce, salada, rocas fracturadas, arcillas, arenas, conglomerados, etc.

La resistividad es medida mediante la inyección de una corriente en el subsuelo y mientras que se monitorea y captura la diferencia de potencial eléctrico en la superficie, esta lectura corresponde al valor de la contribución resistiva de todas las capas por donde fluye la corriente.

La inyección de corriente y medición del potencial se realiza a través de un arreglo de dos pares de electrodos, $A, B(C_1, C_2)$ y $M, N(P_1, P_2)$ respectivamente, siendo el electrodo $A(C_1)$ el polo positivo y $B(C_2)$ el polo negativo de inyección, mientras que el electrodo $M(P_1)$ corresponde al polo positivo y $N(P_2)$ al polo negativo de los electrodos de potencial.

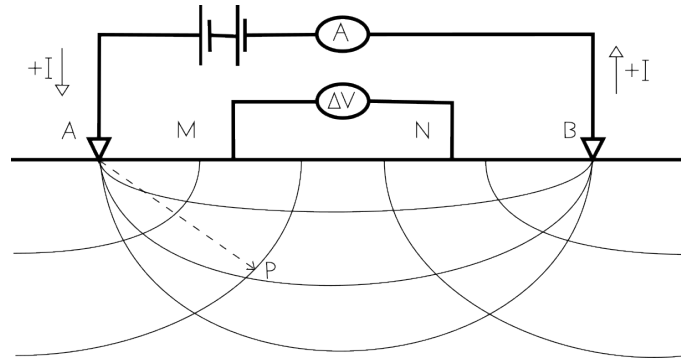


FIGURA 5.2: Configuración general de arreglo de electrodos, modificado de Reynolds (2011)

La resistividad del subsuelo se calcula a partir de la ley de Ohm, considerando el caso general en donde el medio es homogéneo y el arreglo de electrodos presenta una distribución convencional, donde se establece una relación directamente proporcional entre la resistencia R , medida en Ohm (Ω), y el cociente entre la diferencia de potencial ΔV y la corriente inducida I , para un valor puntual (Igboama *et al.*, 2023).

$$R = \frac{\Delta V}{I} \quad (5.1)$$

Sabiendo que se puede calcular R para una sección con longitud L y un área A , transversal del material, conociendo la resistividad (ρ) del material (Igboama *et al.*, 2023; Lowrie y Fichtner, 2020), podemos reescribir la ecuación como:

$$R = \rho \frac{L}{A} \rightarrow \rho = R \frac{A}{L} \rightarrow \rho = R \cdot k \quad (5.2)$$

Donde la resistividad (ρ) es una constante de proporcionalidad del medio y k es el factor geométrico de distribución del flujo de corriente en términos de la del arreglo de los electrodos de inducción y potencial (distancias entre los electrodos A-M-N-B) (Igboama *et al.*, 2023; Lowrie y Fichtner, 2020).

$$k = 2\pi \left(\frac{1}{AM} - \frac{1}{AN} - \frac{1}{BM} + \frac{1}{BN} \right) \quad (5.3)$$

Tenemos que la resistividad aparente (ρ_A) de una sección del subsuelo, corresponde a la contribución resistiva de las unidades geológicas en esa sección, en términos de las distancias entre electrodos, la diferencia de potencial y el flujo de corriente en el medio (Igboama *et al.*, 2023; Lowrie y Fichtner, 2020), esta dado por la siguiente ecuación:

$$\rho_A = 2\pi \cdot \frac{\Delta U}{I} \cdot k \quad (5.4)$$

5.1.3.1 Arreglo de Electrodos y Factor Geométrico

Cada arreglo presenta ventajas, desventajas, rango de sensibilidad y espacio de ejecución, debido a estas características y se tiene que evaluar e identificar que arreglo cumple con las condiciones adecuadas para ser ejecutado, considerando el

espacio disponible en el sitio de estudio, el nivel de ruido (motores, conexiones a tierra mal aterrizadas, antenas, postes metálicos, arboles), la profundidad de objeto de prospección y la resolución vertical alcanzable (ver figura 5.3).

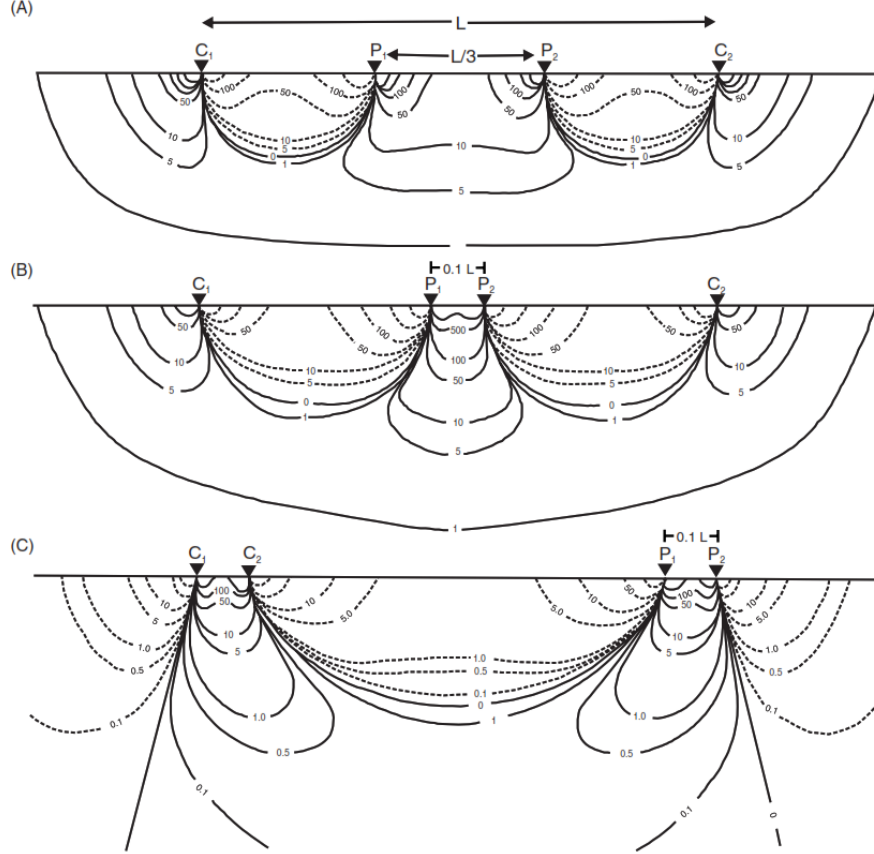


FIGURA 5.3: Esquema de la contribución de la respuesta de resistividad eléctrica, modificado de Reynolds (2011)

Como se observa en la sección anterior, la resistividad se determina empleando una configuración de los electrodos durante una medición, las distintas configuraciones de electrodos se encuentran ampliamente documentadas, cada una presenta un factor geométrico distinto (Igboama *et al.*, 2023; Lowrie y Fichtner, 2020), los principales arreglos geoelectricos son:

Wenner

$$\rho_A = 2\pi \cdot R \cdot a \quad (5.5)$$

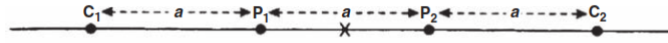


FIGURA 5.4: Esquema del arreglo Wenner, modificado de Reynolds (2011)

Schlumberger

$$\rho_A = \frac{\pi a^2}{b} \left[1 - \frac{b^2}{4a^2} \right] \cdot R, \quad a \geq 5b \quad (5.6)$$

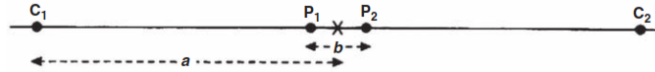


FIGURA 5.5: Esquema del arreglo Schlumberger, modificado de Reynolds (2011)

Dipolo-dipolo

$$\rho_A = \pi n(n+1)(n+2)a \cdot R \quad (5.7)$$

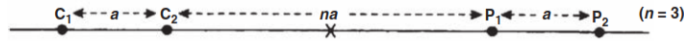


FIGURA 5.6: Esquema del arreglo Dipolo-dipolo, modificado de Reynolds (2011)

5.2 Adquisición de Datos Geofísicos

Previo al trabajo de adquisición se realiza un análisis de entorno, en el cual se verifica la viabilidad del arreglo dadas las condiciones del sitio, considerando lo siguiente: espacio disponible en el sitio de estudio, profundidad de exploración, nivel de ruido eléctrico, interferencias con la estabilidad del potencial natural del subsuelo, profundidad del objeto de exploración y dimensiones aproximadas del mismo.

5.2.1 Intervalo de Muestreo en SEV

El intervalo de muestreo empleado durante la adquisición de un SEV es un parámetro crítico que influye en la calidad y precisión de los datos geofísicos adquiridos, ya que esta estrechamente relacionado con la resolución vertical que deseamos de acuerdo al objeto de estudio. Para lograr este objetivo es necesario considerar el Teorema de Muestreo de Nyquist, el cual, es un principio fundamental en el procesamiento de señales analógicas y digitales, donde establece las condiciones mínimas necesarias para una reconstrucción una señal analógica a partir de muestras discretas (Alvarado Reyes y Stern Forgach, 2010).

El teorema de de muestreo de nyquist nos garantiza las condiciones necesarias y suficientes para llevar a cabo una adquisición exitosa de muestreo de una señal, llámese distribución de resistividad en un medio heterogéneo y discontinuo (Alvarado Reyes y Stern Forgach, 2010).

$$f_s \geq 2 \cdot f_{max} \quad (5.8)$$

Donde la frecuencia de muestreo f_s es por lo menos dos veces mayor a la frecuencia máxima f_{max} conocida, cuando el teorema no se cumple se genera una distorsión en la señal, sumando las frecuencias altas incompletas a la señal natural de baja frecuencia, generando ruido, y problemas de interpretación, se conoce como aliasing (Alvarado Reyes y Stern Forgach, 2010).

5.2.1.1 Factores que Determinan el Intervalo de Muestreo

En el contexto de la adquisición de datos mediante SEV, el intervalo de muestreo es equivalente al espaciado entre puntos donde se realizan mediciones de resistividad del subsuelo. Este intervalo de muestreo debe ser lo mas pequeño posible, de modo que permita obtener muestras de resistividad (Telford *et al.*, 1990), esta

relación se define de la siguiente manera:

$$f_s = \frac{1}{\Delta x} \quad (5.9)$$

donde el intervalo de muestreo Δx debe ser menor a la mitad de la longitud de onda (λ_{min} , espesor) asociado al objetivo de exploración

$$\Delta x \leq \frac{\lambda_{min}}{2} \quad (5.10)$$

5.2.2 Proceso de Adquisición In Situ

La adquisición de datos se realiza mediante la lectura directa en campo, al inducir corriente continua empleando un resistivímetro mediante de los electrodos de corriente A (C_1) y B (C_2), mientras se realiza la lectura de potencia en los electrodos M (P_1) y N (P_2), la lectura se realiza en intervalos regulares en instantes de inyección de corriente distintos (Telford *et al.*, 1990).

Durante la toma de datos es importante considerar los modelos previos realizados durante el análisis preliminar, ya que las resistividades esperadas para las unidades, permiten tener control en la dispersión de datos, identificando tomas erróneas y corrigiendo al momento con una nueva lectura (Telford *et al.*, 1990).

5.3 Machine Learning (ML) en la Geofísica

La aplicación de ML y el DL en la geofísica es ampliamente utilizado en exploración sísmica, abarcando los procesos de adquisición, procesado e interpretación, mejorando los tiempos de procesamiento, clasificación e interpretación, ya que es en este método donde se cuenta con la mayor cantidad de datos para entrenamiento (Wrona *et al.*, 2018); en menor medida se implementan técnicas de ML en la exploración y prospección geoeléctrica, hay algunos ejemplos destacables como son Liu

et al. (2020); El-Qady y Ushijima (2001); Li *et al.* (2024), sin embargo no es un estándar en la industria, pese a las ventajas que puede tener su aplicación, como es el caso de este estudio

El aprendizaje automático o machine learning, son un conjunto de técnicas que utilizan algoritmos con los cuales permite a un sistema aprender y generar predicciones, para lo que requiere un conjunto de datos para poder realizar el entrenamiento. Podemos clasificar los algoritmos de ML de dos maneras, por el tipo de aprendizaje, correspondiendo a Aprendizaje supervisado, no supervisado y por refuerzo, y por la relación que establecen con los parámetros del conjunto de datos de entrenamiento, es decir, modelos paramétricos y no paramétricos (Li *et al.*, 2024).

De los modelos no paramétricos destacan por su adaptabilidad a la estructura subyacente de los datos, por lo que pueden realizar aprendizaje de relaciones complejas entre datos, así como ausentes de linealidad, teniendo un costo en volumen de datos, requiriendo un número mayor para su entrenamiento, destacan los algoritmos siguientes.

- Árboles de decisión
- Random Forests
- K-Nearest Neighbors (KNN)
- Máquinas de soporte vectorial (kernelizados)

Dada la naturaleza de los datos de SEV's, heterogéneos, discontinuos y no lineales, es conveniente abordar su análisis desde un enfoque no paramétrico, teniendo esto en cuenta, la técnica Random Forests destaca siendo eficaz en la tarea de clasificación y regresión, teniendo algunos beneficios como son la reducción del sobre ajuste, interpretación de variables, resistencia al aliasing.

5.4 Random Forests

La herramienta de Random Forests, de acuerdo con Breiman (2001) es un clasificador que emplea conjuntos de arboles de decisión, previamente entrenados, estos son combinados por votación

5.4.1 Propiedades de Random Forests

5.4.2 Robustez y Precisión

5.4.2.1 Manejo de Datos Complejos

5.4.3 Random Forests para Clasificación y Regresión

5.4.4 Aplicaciones de Random Forests en Geofísica

CAPÍTULO 6

RESULTADOS Y CONCLUSIONES

APÉNDICE A
APÉNDICE I

BIBLIOGRAFÍA

- ALVARADO REYES, J. y C. STERN FORGACH (2010), «Un complemento al teorema de Nyquist», *Revista mexicana de física E*, **56**(2), págs. 165–171.
- BREIMAN, L. (2001), «Random forests», *Machine learning*, **45**, págs. 5–32.
- EL-QADY, G. y K. USHIJIMA (2001), «Inversion of DC resistivity data using neural networks», *Geophysical Prospecting*, **49**(4), págs. 417–430.
- FOX, R. W. (1830), «On the electro-magnetic properties of metalliferous veins in the mines of Cornwall», *Philosophical Transactions of the Royal Society of London*, págs. 399–414.
- GANDHI, S. y B. SARKAR (2016), *Essentials of mineral exploration and evaluation*, Elsevier.
- IGBOAMA, W. N., M. AROYEHUN, J. AMOSUN, O. AYANDA, O. HAMMED y J. OLOWOFELA (2023), «Review of geoelectrical methods in geophysical exploration», *Nigerian Journal of Physics*, **32**(3), págs. 141–158.
- LAY, T. y T. C. WALLACE (1995), *Modern global seismology*, Elsevier.
- LI, M., S. YIN, Z. LIU y H. ZHANG (2024), «Machine learning enables electrical resistivity modeling of printed lines in aerosol jet 3D printing», *Scientific Reports*, **14**(1), pág. 14614.
- LIU, B., Q. GUO, S. LI, B. LIU, Y. REN, Y. PANG, X. GUO, L. LIU y P. JIANG (2020), «Deep learning inversion of electrical resistivity data», *IEEE Transactions on Geoscience and Remote Sensing*, **58**(8), págs. 5715–5728.

- LOWRIE, W. y A. FICHTNER (2020), *Fundamentals of geophysics*, Cambridge university press.
- PARASNIS, D. S. (2012), *Principles of applied geophysics*, Springer Science & Business Media.
- REVL, A. y A. JARDANI (2013), *The self-potential method: Theory and applications in environmental geosciences*, Cambridge University Press.
- REYNOLDS, J. M. (2011), *An introduction to applied and environmental geophysics*, John Wiley & Sons.
- SORRELL, C. A. (1973), *Rocks and minerals: A guide to field identification*, Macmillan.
- TELFORD, W. M., L. P. GELDART y R. E. SHERIFF (1990), *Applied geophysics*, Cambridge university press.
- TIAB, D. y E. C. DONALDSON (2024), *Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties*, Elsevier.
- WRONA, T., I. PAN, R. L. GAWTHORPE y H. FOSSEN (2018), «Seismic facies analysis using machine learning», *Geophysics*, **83**(5), págs. O83–O95.