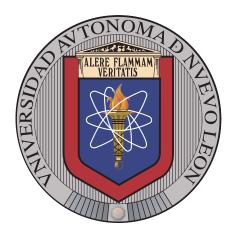
Universidad Autónoma de Nuevo León

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

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
M.C. José Anastas	sio Hernández Saldaña
A	Asesor
Nombre del revisor C	Nombre del revisor D
Revisor	Revisor
Ve	o. Bo.
Dra. Azucena Yolo	oxóchitl Ríos Mercado
Subdirector de E	Estudios de Posgrado

Aquí puedes poner tu dedicatoria si es que tienes una.

ÍNDICE GENERAL

Ag	grade	ecimier	atos		VIII
1.	INT	'RODU	JCCIÓN		1
2.	DEI	LIMIT	ACIÓN Y PLANTEAMIENTO DEL PROBLEMA		2
3.	JUS	TIFIC	CACIÓN		3
4.	FOF	RMUL	ACIÓN DE OBJETIVOS		5
	4.1.	Genera	ales		5
	4.2.	Especí	ficos		5
5 .	MA	RCO T	ΓΕÓRICO		6
	5.1. Geofísica y Geoeléctrica				
		5.1.1.	Definición de Geofísica		6
		5.1.2.	Resistividad de la Tierra		6
		5.1.3.	Sondeo Eléctrico Vertical (SEV)		8
	5.2.	Adquis	sición de Datos Geofísicos		12
		5.2.1.	Intervalo de Muestreo en SEV		13
		5.2.2.	Proceso de Adquisición In Situ		15

Índice	GENER	AL	V
5.3.	Machine Learning (ML) en la Geofísica		15
5.4.	. Random Forests		
	5.4.1.	Siembra del bosque	17
	5.4.2.	Predicciones del bosque	17
	5.4.3.	Margen y Error de Generalización	18
	5.4.4.	Robustez y Convergencia	18
	5.4.5.	Aplicaciones de Random Forests en Geofísica	18
6. RE	\mathbf{SULT}^{A}	ADOS Y CONCLUSIONES	19
A. Ap	éndice	I	20

ÍNDICE DE FIGURAS

5.1.	Estructura atómica del oro	7
5.2.	Configuración general de electrodos	9
5.3.	Esquema de la contribución de la respuesta eléctrica	11
5.4.	Esquema del arreglo Wenner	12
5.5.	Esquema del arreglo Schlumberger	12
5.6.	Esquema del arreglo Dipolo-dipolo	12

ÍNDICE DE TABLAS

AGRADECIMIENTOS

Aquí puedes poner tus agradecimientos. (No olvides agradecer a tu comité de tesis, a tus profesores, a la facultad).

Capítulo 1 INTRODUCCIÓN

que es la geofisica:

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

eL geoelectrica

para que sirve

oportunidades de la cuiencia 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 geoeléctricos in situ, no es posible conocer el resultado del trabajo hasta una vez realizado el procesado de los mismos, por lo que no se tiene certeza de si la adquisición realizada en campo representara con claridad el objeto de prospección, es decir, si el muestreo realizado logra cubrir el espectro de frecuencia, y por consiguiente, representar con claridad las unidades geológicas de un sitio en particular.

El muestreo propuesto en la etapa de planeación de adquisición, en muchas ocasiones incorpora un grado alto de ambigüedad, debido a que no es posible conocer con certeza la distribución y espesores de las unidades geoeléctrica y por lo tanto no es factible un muestreo completamente efectivo, siendo solo parcialmente evidente durante la exploración directa del medio, ya que dicho procedimiento solo permite aprecia una porción ínfima del terreno.

Se plantea como herramienta de análisis y mejora de muestreo la implementación de técnicas de Machine Learning, empleando esta técnica de aprendizaje mediante su entrenamiento con datos procesados y calibrados por sondeo directo, de manera que esto permita identificar oportunidades de mejora en el muestreo y la respuesta en general.

Capítulo 3 JUSTIFICACIÓN

Existen múltiples aplicaciones de ML y DL en el procesamiento, modelado e interpretación geofísica (Li et al., 2024; Liu et al., 2020; El-Qady y Ushijima, 2001; Wrona et al., 2018), algunas de estas aplicaciones corresponden a implementaciones académicas, al igual es posible identificar aplicaciones comerciales implementadas en software principalmente de exploración sísmica de hidrocarburos (Diaferia et al., 2024; Panebianco et al., 2024).

La aplicación de herramientas de ML durante la ejecutar de muestreo de Sondeo Eléctrico Vertical (SEV), en particular durante el proceso de adquisición in situ, presenta la posibilidad de evaluar y ampliar el intervalo de muestreo original, mejorando la adquisición tradicional, permitiendo evaluar el muestreo contra la regresiones y establecer nuevos intervalos de oportunidad.

Durante el muestreo tradicional los intervalos se diseñan considerando el objetivo de exploración (idealmente), este análisis previo es un factor determinante, en esta etapa se establece el intervalo de muestreo que permitiría identificar el objeto de exploración, este objeto de exploración pueden ser unidades geológicas, acuíferos, fallas, zonas de fracturas, estructuras antropogénicas, etc.

De manera general se busca 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, lo que implicaría conocer previamente la conformación, distribución y espesor de cada unidad, lo que resulta irrisorio.

De manera que un modelo entrenado permita generar múltiples regresiones e identificar regiones de interés no cubiertas por el muestreos y de esta manera generar puntos adicionales *in situ* optimizando la adquisición de datos e impactando en la calidad del muestreo , mejorando el acotamiento de las frecuencias identificadas.

Pese a existir aplicaciones de ML implementadas en geofísica, no se identifica alguna enfocada este esta problemática en particular, sin embargo hay aplicaciones en otros campos de estudio con un enfoque en el muestreo y clasificación de datos no paramétricos (Entezami et al., 2022; Bkassiny et al., 2013; Shi y Wang, 2021) empleando técnicas como Dirichlet Process Mixture Model (DPMM), radial basis function network (RBFN), Multiple Point Statistics (MPS) y Bayesian Compressive Sensing (BCS).

porque random forest y su copmaración con BCS

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 geoeléctricos con unidades geológicas de terrenos ya caracterizados mediante SEV's.
- Establecer un delo de clasificación y regresión utilizando Random Forests empleando modelos geoeléctricos.
- Generar un modelo predictivo que permita clasificar datos geoeléctricos 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 Cornwal, 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 especifica 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).

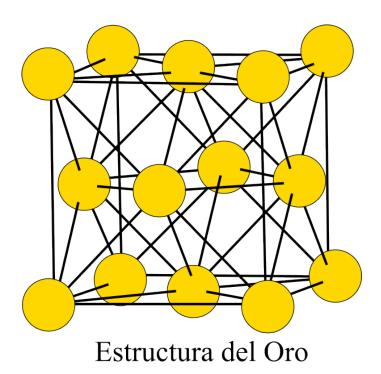


FIGURA 5.1: Esquema de la estructura atómica de oro que conforma la cristalización octahedral, modificado de Sorrell (1973)

.

Los métodos Geoeléctricos 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 geoeléctrica, 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 hidrogelogica 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.

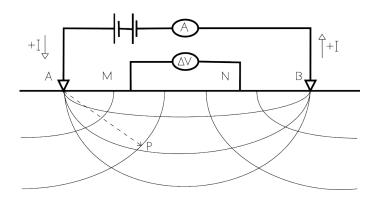


FIGURA 5.2: Configuración general de arreglo de electrodos, modificado de Reynolds (2011)

.

La resistividad del subsuelo se calcula a partir de la ley de Ohm, considerando el caso general en donde el medios es homogéneo y el arreglo de electrodos presenta una distribución convencional, donde se establece una relación directamente proporcional entre la 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 $et\ al.,\ 2023$).

$$R = \frac{\Delta V}{I} \tag{5.1}$$

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:

$$R = \rho \frac{L}{A} \to \rho = R \frac{A}{L} \to \rho = R \cdot k \tag{5.2}$$

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).

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

$$\rho_A = 2\pi \cdot \frac{\Delta U}{I} \cdot k \tag{5.4}$$

5.1.3.1 Arreglo de Electrodos y Factor Geométrico

Cada arreglo presenta ventas, 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).

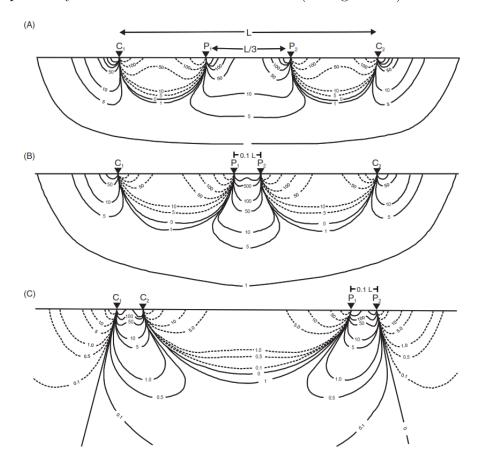


FIGURA 5.3: Esquema de la contribución de la respuesta de resistividad eléctrica, modificado de Reynolds (2011)

.

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 geoeléctricos son:

Wenner

$$\rho_A = 2\pi \cdot R \cdot a \tag{5.5}$$

$$C_1 \leftarrow \cdots = a \leftarrow \cdots \Rightarrow P_1 \leftarrow \cdots = a \leftarrow \cdots \Rightarrow P_2 \leftarrow \cdots = a \leftarrow \cdots \Rightarrow C_2$$

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

Schlumberger

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

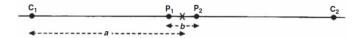


FIGURA 5.5: Esquema del arreglo Schlumberger, modificado de Reynolds (2011)

.

Dipolo-dipolo

$$\rho_A = \pi n(n+1)(n+2)a \cdot R \tag{5.7}$$

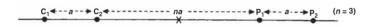


FIGURA 5.6: Esquema del arreglo Dipolo-dipolo, modificado de Reynolds (2011)

.

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 distintas condiciones, como son:

- Los espesores de cada unidad.
- La distribución de las distintas unidades.
- Profundidad de investigación
- Ruido en la señal.

Para establecer un intervalo de muestreo apropiado, se deben considerar el Teorema de Muestreo de Nyquist y El teorema de Shannon-Hartley (teorema de codificación de canal ruidoso)

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).

$$f_s \ge 2 \cdot f_{max} \tag{5.8}$$

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).

Considerando el medio geologico como una region con precencia coanstante de ruido electrico de fuentes tanto naturales como humanas, es impresindible considerar el teorema de Shannon-Hartley aplicando apilamiento de muestreo como metodo de reduccion de la relacion ruido señal, durante la adquisicion de datos; esto quiere decir calcular el promedio de muestreos cointinuos en un intervalo definido de aperturas entre electrodos.

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:

$$f_s = \frac{1}{\Delta x} \tag{5.9}$$

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

$$\Delta x \le \frac{\lambda_{min}}{2} \tag{5.10}$$

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 resistivimetro 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 analistas desde un enfoque no paramétrico, teniendo esto en cuanta, 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 técnica Random Forests emplea múltiples arboles de decisión independientes entre si, donde cada árbol realiza una votación de clases, donde se selecciona la más popular de la entrada de cada árbol realizando una combinación de salida, permitiendo realizar una clasificación de características complejas o realizar regresiones de datos complejos multivariables (Breiman, 2001; Lan *et al.*, 2020).

La herramienta de Random Forests, de acuerdo con Breiman (2001) emplea tres elementos clave en el proceso de entrenamiento, bagging, selección aleatoria de características y agregación por votación, resultando en la combinación del los resultados en una predicción o clasificación robusta y ajustada (Lan *et al.*, 2020).

5.4.1 Siembra del bosque

Breiman (2001) nos dice que Random Forests es un conjunto de clasificadores $H(x, \theta_k)$, x es un vector de entrada y θ_k corresponden a vectores aleatorios independientes.

A partir de los datos de entrada, se generan subconjuntos de datos de entrenamiento, estos se seleccionan con cierta aleatoriedad empleando la técnica bootstrap sampling, en cada nodo de los subconjuntos de entrenamiento se selecciona un subconjunto de características por votación de popularidad, dejando crecer cada árbol sin realizar poda hasta completar los criterios de finalización, es decir un numero de instancias preestablecido(Breiman, 2001).

5.4.2 Predicciones del bosque

La salida de un Random Forest para una entrada x se basa en las predicciones individuales de los árboles para cada clase $h_k(x)$, se realiza un conteo de cada clase, producto de la predicción de cada árbol, sumando las salidas $I(h_k(x) = c)$, y finalmente se selecciona clase con mayor numero de predicciones, obteniendo la predicción de clasificación H(x), donde x es una función indicadora que vale 1 si $h_k(x) = c$, y 0 en caso contrario (Breiman, 2001).

$$H(x) = \operatorname{argmax}_{c} \sum_{k=1}^{K} I(h_{k}(x) = c)$$
 (5.11)

El proceso de la regresión se obtiene a partir de la media aritmética de cada predicción individual, donde cada árbol produce un valor numérico $h_k(x)$ correspondiente a cada x, al corresponder con promedio de las predicciones se le otorga mas

estabilidad cuando tenemos un numero elevado de arboles y un conjunto de datos grande, entendiéndolo como un modelo central que incorpora información de cada árbol (Breiman, 2001).

$$H(x) = \frac{1}{K} \sum_{k=1}^{K} h_k(x)$$
 (5.12)

- 5.4.3 Margen y Error de Generalización
- 5.4.4 Robustez y Convergencia
- 5.4.5 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.
- BKASSINY, M., S. K. JAYAWEERA y Y. LI (2013), «Multidimensional dirichlet process-based non-parametric signal classification for autonomous self-learning cognitive radios», *IEEE Transactions on Wireless Communications*, **12**(11), págs. 5413–5423.
- Breiman, L. (2001), «Random forests», Machine learning, 45, págs. 5–32.
- DIAFERIA, G., L. VALOROSO, L. IMPROTA y D. PICCININI (2024), «A high-resolution seismic catalog for the Southern Apennines (Italy) built through template-matching», *Geochemistry, Geophysics, Geosystems*, **25**(3), pág. e2023GC011160.
- EL-Qady, G. y K. Ushijima (2001), «Inversion of DC resistivity data using neural networks», *Geophysical Prospecting*, **49**(4), págs. 417–430.
- Entezami, A., H. Shariatmadar y C. De Michele (2022), «Non-parametric empirical machine learning for short-term and long-term structural health monitoring», *Structural Health Monitoring*, **21**(6), págs. 2700–2718.
- 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.

Bibliografía 22

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.

- LAN, T., H. Hu, C. JIANG, G. YANG y Z. ZHAO (2020), «A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification», *Advances in Space Research*, **65**(8), págs. 2052–2061.
- 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.
- Panebianco, S., C. Satriano, G. Vivone, M. Picozzi, A. Strollo y T. A. Stabile (2024), «Automated detection and machine learning-based classification of seismic tremors associated with a non-volcanic gas emission (Mefite d'Ansanto, Southern Italy)», Geochemistry, Geophysics, Geosystems, 25(2), pág. e2023GC011286.
- Parasnis, D. S. (2012), *Principles of applied geophysics*, Springer Science & Business Media.
- Revil, 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.

Bibliografía 23

Shi, C. y Y. Wang (2021), «Non-parametric machine learning methods for interpolation of spatially varying non-stationary and non-Gaussian geotechnical properties», *Geoscience Frontiers*, **12**(1), págs. 339–350.

- 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.