
PREDICCIÓN DE LITOLOGÍA BASADA EN JERARQUÍA DE ATRIBUTOS SÍSMICOS CON MACHINE LEARNING

A PREPRINT

María Elena Véliz

Roberto Andrade

Joan Torres

Soledad Berrones

Christian Mejia-Escobar

Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental (FIGEMPA)
Universidad Central del Ecuador
Quito, Ecuador
{meveliz, rsandrade, cimejia}@uce.edu.ec

19 de octubre de 2025

ABSTRACT

Este estudio presenta una metodología para la predicción de litologías en el contexto de la exploración de hidrocarburos, utilizando redes neuronales supervisadas y atributos sísmicos cuantitativos derivados de datos sísmicos tridimensionales. El análisis se enfoca en la jerarquización de diversos atributos sísmicos, con el fin de mejorar la clasificación y caracterización de formaciones geológicas. A través del uso de redes neuronales artificiales, se desarrolló un modelo multiclase con una arquitectura optimizada que alcanzó una precisión de hasta el 95 %. Los resultados obtenidos demostraron que la combinación de atributos sísmicos y machine learning puede superar las limitaciones de los métodos tradicionales en la interpretación de litologías, mejorando significativamente la identificación de formaciones geológicas con potencial para contener hidrocarburos. Además, se proporcionaron visualizaciones en 2D y 3D de los volúmenes sísmicos procesados, destacando la precisión del modelo en la predicción de zonas con alta probabilidad de hidrocarburos. Este enfoque proporciona una herramienta eficiente para la toma de decisiones en la exploración petrolera, optimizando los recursos y reduciendo el tiempo necesario para el análisis sísmico.

Palabras clave: Predicción de litologías, Redes neuronales supervisadas, Atributos sísmicos, Datos sísmicos 3D, Machine learning

1. Introducción

La búsqueda de fuentes de energía más eficientes y sostenibles ha llevado a la industria petrolera a innovar en sus métodos de exploración; la energía basada en el petróleo sigue siendo un pilar fundamental en el desarrollo global, al impulsar industrias, sistemas de transporte y diversas actividades económicas. En este contexto, la caracterización precisa de las litologías del subsuelo que son almacenadoras de este recurso juega un papel clave en la exploración y explotación de hidrocarburos, ya que permite comprender la distribución y calidad de los yacimientos.

Sin embargo, los métodos tradicionales para analizar las litologías suelen ser manuales, lentos y propensos a errores, lo que puede afectar la toma de decisiones estratégicas en la industria petrolera. Ante esta limitación, el uso de herramientas avanzadas como Python y técnicas de Machine Learning representa una revolución, al ofrecer una alternativa más rápida, precisa y automatizada para el análisis del subsuelo. Esto no solo optimiza los procesos, sino que también contribuye a reducir costos en las perforaciones petroleras.

Entre las aplicaciones prácticas de estas tecnologías se destacan la identificación y clasificación de litologías en yacimientos de petróleo y gas, optimizando la interpretación de datos. No obstante, el problema principal radica en la predicción de litologías en volúmenes sísmicos tridimensionales, una tarea que los métodos comunes abordan mediante el análisis manual de registros de pozos y su extrapolación basada en tendencias.

El enfoque tradicional presenta debilidades significativas, como la dependencia de la experiencia del especialista, la subjetividad en los resultados y el tiempo requerido para procesar grandes volúmenes de datos. Estas limitaciones resaltan la necesidad de adoptar herramientas que mejoren la precisión y reduzcan el tiempo de análisis. La propuesta de este trabajo plantea el uso de Machine Learning y redes neuronales para desarrollar un modelo de predicción de litologías basado en atributos sísmicos. Este enfoque busca determinar zonas prospectivas con mayor precisión y en menor tiempo, utilizando técnicas avanzadas para clasificar litologías de manera supervisada, evaluar la importancia de las variables y propagar las predicciones a través de la sísmica 3D.

La metodología consiste en estructurar un dataset con atributos sísmicos cuantitativos, normalizar los datos, entrenar un modelo supervisado y aplicar las predicciones a volúmenes sísmicos completos. Este proceso integra los datos de los pozos con información sísmica. Los beneficios de esta propuesta incluyen la reducción de la subjetividad en el análisis, la aceleración del proceso de caracterización y la mejora en la precisión de los resultados. Esto es importante en el contexto donde se dispone de información limitada de pozos y se prioriza el uso de datos sísmicos.

Finalmente, la estructura del paper se divide en tres secciones: una introducción donde se presentará el marco teórico y los recursos utilizados, una descripción metodológica detallada del desarrollo del modelo, y un análisis de los resultados, destacando las fortalezas y limitaciones del enfoque, junto con recomendaciones para futuros trabajos.

2. Estado del arte

La caracterización de reservorios de hidrocarburos es un desafío en la exploración y producción de petróleo y gas. Tradicionalmente, este proceso ha dependido del análisis manual de registros de pozos y datos sísmicos, lo que puede ser lento y propenso a errores. En respuesta a estas limitaciones, diversas investigaciones han explorado la aplicación de redes neuronales artificiales (ANN) y atributos sísmicos para mejorar la precisión y eficiencia en la predicción de propiedades del subsuelo, cabe recalcar que la mayoría de los estudios son recientes entre los años 2019 y 2024 y no existe amplia información del tema.

El estudio de Wang et al., (2021) tiene como objetivo predecir la litología de formaciones a partir de datos de registro para optimizar la perforación en el campo de gas Kela-2 utilizando una red neuronal de retro propagación Mejorada (BP) con datos de registros sísmicos y de rayos gamma en formaciones de arenisca y lutita del campo mencionado dando como resultado que el modelo BP mejorado mostró una buena estabilidad y precisión en la predicción de litología con valores esperados cercanos a los valores reales.

Ruzic et al., (2021) ayuda en la planificación de pozos futuros y en la distinción de los datos geofísicos en áreas de pozos secos de aquellos de un pozo económicamente viable utilizando redes neuronales artificiales (ANN) con datos de atributos sísmicos en volúmenes sísmicos reprocesados y reinterpretados, dando como resultado que las redes neuronales no ayudaron en el procesos de distinguir la posible acumulación de gas, pero puede señalar los elementos sedimentológicos y estructurales dentro del volumen sísmico.

Ntibahanana et al., (2024) realizan una inversión de propiedades de rocas a partir de atributos sísmicos utilizando redes neuronales con un enfoque de ensamble ponderado con datos sísmicos y registros de pozo dando como resultado valores $R=0.993$ y un MAE de 0.00112 con una porosidad media estimada al modelo de referencia.

Nguyen Duy Muoi et al., (2023) en su estudio predicen la distribución de areniscas del Oligoceno tardío en la cuenca de Cuu Long utilizando redes neuronales artificiales y atributos sísmicos; las redes utilizadas son no supervisadas y además realizan un Análisis de Componentes Principales (PCA), entre los datos utilizados son datos sísmicos 3D (300 km²), atributos sísmicos, dando como resultado que la combinación de UNN con PCA mejora la reducción de ruido en la sísmica y permite identificar la distribución de reservorios en la cuenca.

Oumarou et al., (2021) El estudio revisa la literatura sobre el uso de atributos sísmicos en la caracterización de yacimientos, identificando sus limitaciones y explorando atributos poco utilizados en este campo. Se clasifican los atributos según su nivel de intervención en la identificación de zonas de acumulación de hidrocarburos. Para ello, se emplea un análisis multiatributo basado en co-kriging, redes neuronales y la matriz de covarianza, permitiendo mejorar las predicciones al combinar múltiples atributos. Los resultados destacan la importancia de estos atributos en la estimación de propiedades del yacimiento, como porosidad y permeabilidad, y en la detección de hidrocarburos, resaltando la necesidad de combinarlos con otras metodologías para optimizar su precisión.

La metodología utilizada en los distintos estudios depende en gran medida de datos sísmicos 3D, mientras que la disponibilidad limitada de datos de pozos puede afectar la precisión de la predicción. Nguyen et al. (2023) en sus estudios realiza una combinación de PCA con redes neuronales no supervisadas (UNN) que contribuye a la reducción de ruido en los datos sísmicos, la efectividad del método sigue estando condicionada por la calidad y cantidad de los datos sísmicos disponibles.

Por otro lado, Oumarou et al. (2021) resaltan que los atributos sísmicos por sí solos pueden no reflejar con precisión todas las propiedades del subsuelo, requiriendo su integración con otros métodos como análisis de pozos y redes neuronales. Además, los autores señalan la necesidad de validar los resultados con nuevos datos de perforación, lo que indica que la precisión del modelo aún no ha sido completamente comprobada.

La mayoría de los estudios de los diferentes autores utilizan redes neuronales no supervisadas (UNN) y no detallan los atributos sísmicos utilizados en el modelo, es por esto que nuestro estudio va enfocado en complementar dichas debilidades como es el uso de redes neuronales supervisadas, se da a conocer en detalle los diferentes atributos sísmicos utilizados para construir el modelo, adicional a esto se realiza una jerarquización de variables basadas en predicciones, cuya predicción se utiliza para 3 tipos de litologías y para concluir se realiza una visualización en 2D y 3D.

3. Metodología

Los avances recientes en los métodos sísmicos han mejorado la eficiencia de los geólogos en el análisis de estructuras geológicas y la determinación de parámetros de reservorios. El presente estudio tiene como objetivo predecir litología en volúmenes sísmicos con Machine Learning y jerarquización de variables sísmicas.

Los pozos ubicados dentro del volumen sísmico cuentan con una caracterización litológica realizada manualmente la cual será la variable dependiente del entrenamiento para el algoritmo elegido, en este caso redes neuronales artificiales. La extracción de valores de los atributos sísmicos realizada en las profundidades correspondientes de la litología actúa como variables independientes de este entrenamiento. Una vez entrenado el modelo permitirá generar predicciones precisas y útiles en áreas donde no se cuenta con información detallada de pozos, pero sí de datos sísmicos.

En la industria del petróleo las ANN están siendo ampliamente utilizadas para resolver problemas específicos como para completar registros de pozo discontinuos, detectar anomalías en la sísmica por efecto AVO y la selección de electrofacies a partir de registros eléctricos. Sin embargo, la técnica de multiatributos como entrenamiento para clasificación litológica es aún un tópico reciente. Algunos atributos sísmicos pueden reflejar cambios en la litología y facies sedimentarias, mientras que otros pueden afectar a la detección o generar redundancia. En este trabajo se propone una metodología innovadora para identificar los atributos relevantes en una caracterización litológica gracias a la importancia de variables y pruebas con Análisis del Componente Principal (PCA), eficiente para la exploración de hidrocarburos en escenarios con datos limitados de pozos.

La Figura 1 presenta las actividades realizadas para alcanzar los objetivos tanto general como específicos de este estudio.

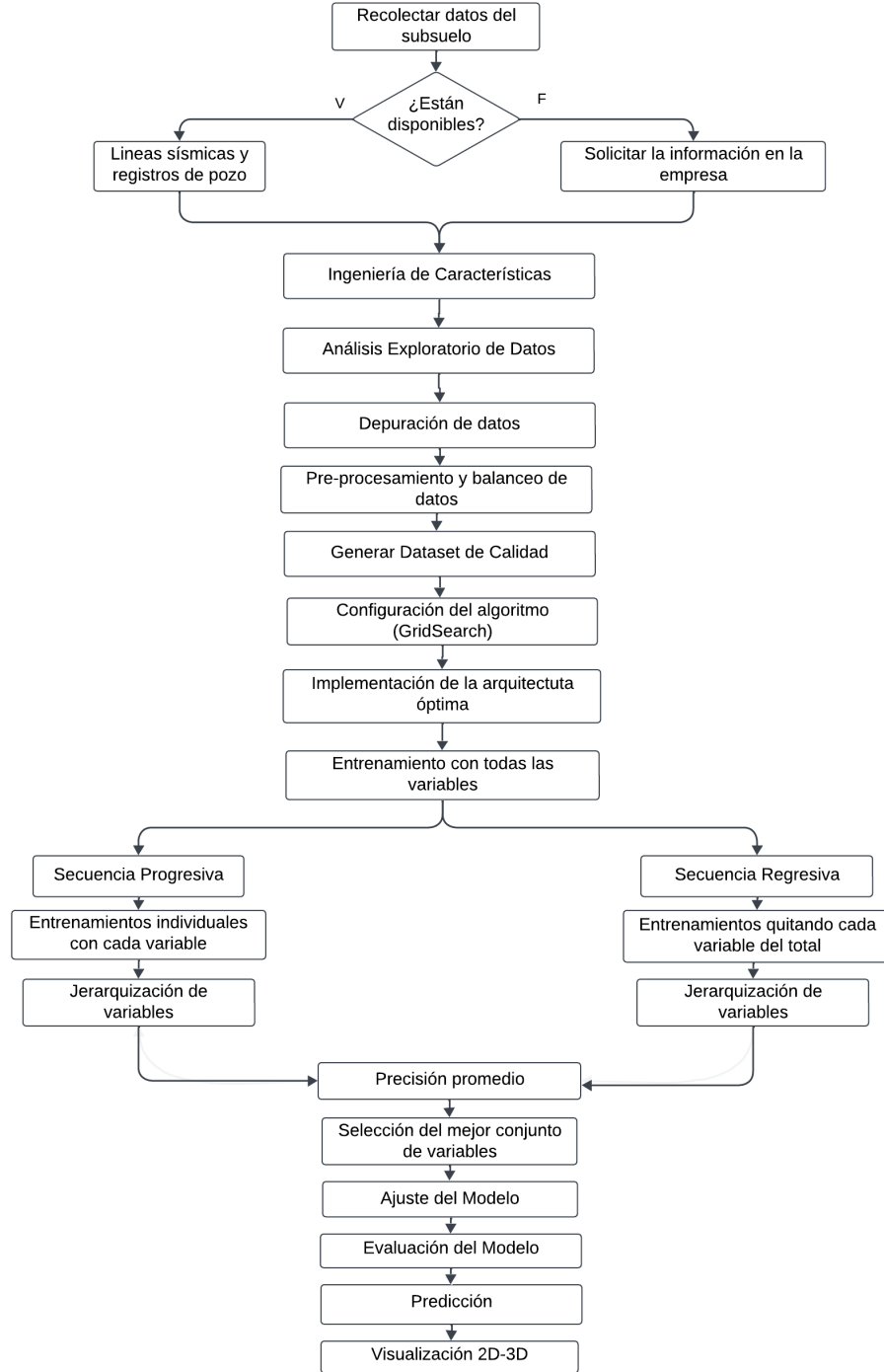


Figura 1. Diagrama de flujo de la metodología utilizada en la predicción de litologías mediante Machine Learning. Este flujograma muestra las etapas principales del proceso, desde la estructuración del dataset hasta la predicción y visualización de litologías en volúmenes sísmicos, destacando el uso de atributos sísmicos y redes neuronales supervisadas

Etapas de la metodología

1. Estructuración del Dataset. Se recopilan variables cuantitativas derivadas de la ubicación de los pozos, de atributos sísmicos como densidad, porosidad, curvatura, coeficiente de Poisson, Similarity, Envelope, y RMS. Además, cada

registro del dataset se le asigna una litología clasificada en tres tipos con base en datos reales de pozos. 2. Preprocesamiento de datos. Se realiza la normalización de las variables para garantizar que los valores estén dentro de rangos comparables. Además de aplicar una función para equilibrar la variable dependiente, añadiendo mayor peso a aquellas con menor cantidad de datos y viceversa para el valor que más veces se repite. Se determina la relación de cada variable independiente con la variable dependiente, aquella que presente mayor precisión será agregada al modelo y así sucesivamente con la siguiente variable de mayor precisión. Hasta encontrar el modelo óptimo para la clasificación litológica. Estos modelos se registrarán en una tabla constando las variables que participan y la precisión del mismo Tabla 1.

MODELO	COORDENADA X	COORDENADA Y	COORDENADA Z	IP	IS	DENSIDAD	POISSON	VPVS	DECOMP	SIMILARIDAD	ENVOLVENTE	CURVATURA	RMS	LITOLOGIA	PRECISION

Tabla 1. Tabla de registro de los modelos para saber la precisión de las variables.

3. División de los datos. El conjunto de datos se separa en dos subconjuntos: entrenamiento (para ajustar el modelo) y prueba (para evaluar su desempeño). 4. Configuración y ajuste del modelo. Se configura un algoritmo de Machine Learning adecuado para clasificación supervisada. Se establecen los hiperparámetros del modelo, como el número de neuronas, capas y tasa de aprendizaje. 5. Evaluación. Se evalúa el modelo previamente entrenado, utilizando el subconjunto de los datos destinados para prueba. 6. Predicción. El modelo se utiliza para para la predicción litológica en un cubo sísmico.

Datos

El conjunto de datos tutorial fue proporcionado por Beicip-Franlab, el cual incluye 3 pozos en formato. las, un cubo sísmico, 5 cubos correspondientes a los resultados de una inversión (IP, IS, densidad, Poisson ratio y VPVS) en formato .segy y 3 horizontes delimitando el reservorio. El cubo sísmico 3D tiene una superficie de 12 km² correspondiente a rangos de 600 INLine y 500 XLine y una profundidad de 1500 ms. Datos sísmicos 3D de alta resolución proporcionan rasgos estructurales y estratigráficos que aseguran una mejor interpretación y atributos más confiables. Junto con los registros de pozos se incluyen los resultados de una caracterización petrofísica manual, identificando tres facies: lutitas (codificación 1), areniscas saturadas en agua (codificación 2) y areniscas con hidrocarburo (codificación 3) (Figura 2).

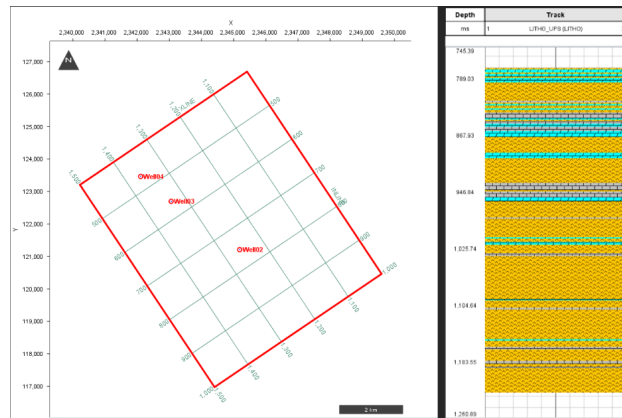


Figura 2. A la izquierda se presenta la ubicación de los 3 pozos utilizados y a la derecha la litología que corresponde a Lutitas, Areniscas y Areniscas con hidrocarburo.

Con el software InterWell v.2024 se calcularon un total de cinco atributos de interés (Descomposición espectral a 50hz, similaridad, envolvente, curvatura y RMS) a partir de la sísmica (Tabla 2). Dentro del mismo programa, después de realizar un procesamiento y escalamiento adecuado a las frecuencias de la sísmica, utilizando la opción “Crear tabla a partir de registros litológicos y cubos sísmicos” se generó una matriz de 14 columnas (coordenadas XY, profundidad en tiempo, valores de los atributos y clase de litología) y 5913 muestras a profundidad definida en tiempo.

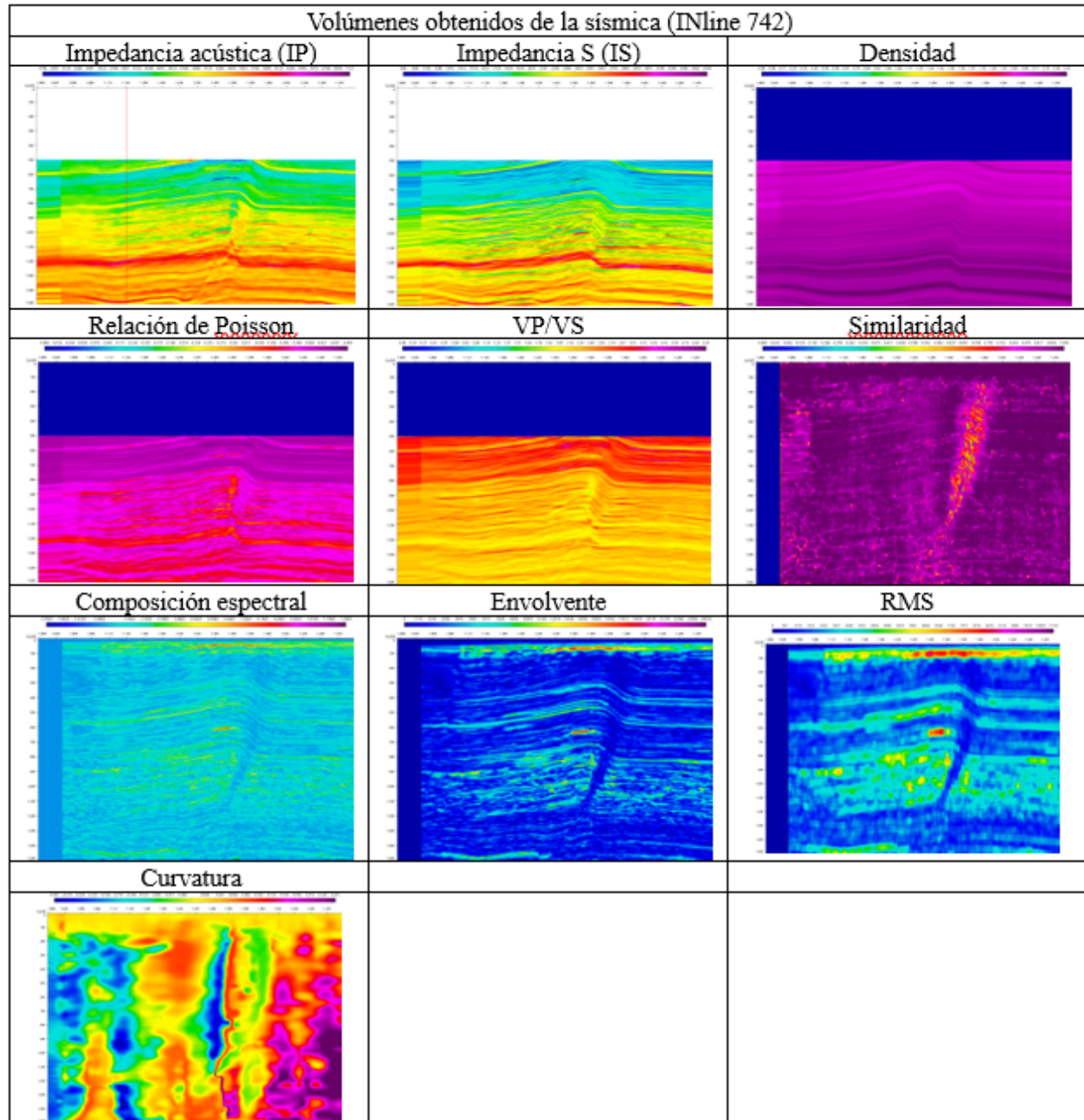


Tabla 2. Propiedades sísmicas utilizadas en la caracterización de formaciones geológicas

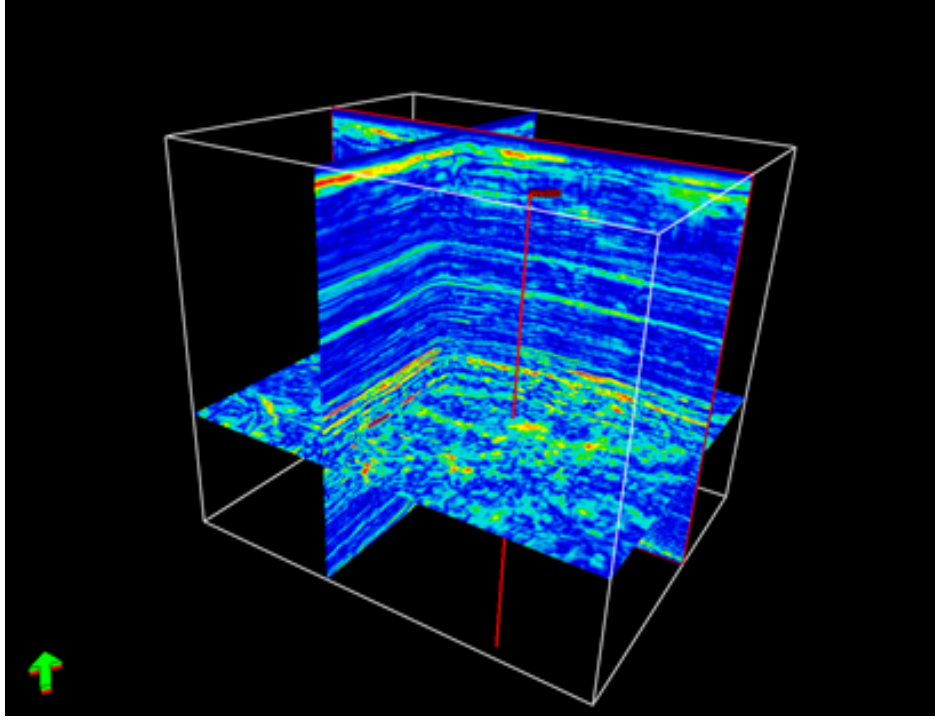


Figura 3. Volumen sísmico 3D del atributo envolvente y la ubicación del pozo 02.

El cubo sísmico representado en la Figura 3 ilustra un modelo tridimensional de uno de los atributos sísmicos generados y la ubicación de un pozo, en cuya ubicación se realizó la extracción de valores correspondiente a esa variable.

Arquitecturas y modelos

Se diseñó una red neuronal artificial (ANN) para clasificación multiclase, compuesta por una capa de entrada adaptada al número de atributos del conjunto de datos, varias capas ocultas con funciones de activación ReLU y una capa de salida con función softmax para generar probabilidades normalizadas. Las capas ocultas se optimizaron ajustando el número de neuronas. Además, se empleó el algoritmo Adam para la optimización debido a su capacidad de ajuste dinámico de tasas de aprendizaje.

Para prevenir el sobreajuste y mejorar la generalización, se integraron técnicas como Batch Normalization para normalizar activaciones y Dropout para desactivar aleatoriamente neuronas durante el entrenamiento. La configuración final de la arquitectura se obtuvo evaluando métricas de desempeño como precisión, pérdida y la matriz de confusión.

Análisis Exploratorio y Transformación de Datos

En esta fase, se realizó un análisis exploratorio de los datos con el objetivo de caracterizar la distribución y las propiedades generales de las variables. Este proceso permitió identificar posibles sesgos y definir las estrategias adecuadas para su procesamiento y transformación.

Los resultados mostraron que algunas variables, como ENVOLVENTE, IS, RMS y SIMILARIDAD, presentaban una notable asimetría. Para corregir esto, se aplicó PowerTransformer utilizando el método Yeo-Johnson, el cual fue elegido por su capacidad para manejar valores negativos y reducir sesgos sin requerir ajustes manuales. La transformación se aplicó sistemáticamente a estas variables, lo que resultó en una distribución más simétrica. La efectividad de la transformación se verificó mediante la comparación de histogramas antes y después del procesamiento, observando mejoras claras en la normalidad de los datos.

Creación del Dataset de Calidad

En el preprocesamiento, se detectó que la variable litología, compuesta por tres clases, presentaba un desbalance significativo, lo cual podría afectar negativamente el rendimiento del modelo. Para mitigar este problema, se incrementó la resolución en las regiones donde predominaban las clases 2 y 3, mejorando así su representación en el conjunto de datos.

Además, se identificaron valores de 0 en las variables DENSIDAD, POISSON y VPVS, los cuales no correspondían a mediciones reales, sino que eran asignados automáticamente por el software en casos de datos faltantes. Estos valores fueron eliminados, asegurando que el conjunto de datos estuviera libre de entradas no válidas.

Con estos ajustes, se obtuvo un dataset más equilibrado y representativo, adecuado para su posterior análisis.

Preprocesamiento de los Datos

Con el dataset de calidad listo, se procedió al preprocesamiento para estructurar adecuadamente los datos para el modelo de machine learning. Este proceso incluyó la selección de variables, el escalado de los datos, la codificación de variables categóricas y la división del conjunto de datos.

Selección de Variables: Se identificaron las variables independientes y la variable dependiente. Las variables seleccionadas fueron: 'COORD_X', 'COORD_Y', 'COORD_Z', 'IP', 'IS', 'DENSIDAD', 'POISSON', 'SIMILARIDAD', 'ENVOLVENTE' y 'RMS', que contienen información relevante para predecir la variable dependiente litología. La selección se realizó mediante un proceso de jerarquización de variables, priorizando aquellas con mayor capacidad explicativa y eliminando las que podrían introducir redundancia y dificultar el aprendizaje.

Escalado de los Datos: Se probó inicialmente un escalador estándar, basado en la media y la desviación estándar, pero este enfoque no funcionó debido a la presencia de valores atípicos que distorsionaban la media. Por ello, se optó por RobustScaler, que utiliza la mediana y los cuartiles, lo que lo hace menos sensible a los valores extremos y asegura un escalado más homogéneo, mejorando así la convergencia del modelo durante el entrenamiento.

Codificación de la Variable Litología: Como litología es una variable categórica, se transformó utilizando One-Hot Encoding, lo que permitió convertir sus tres clases en variables binarias, facilitando su interpretación por el modelo.

División del Dataset: El conjunto de datos se dividió en tres subconjuntos: entrenamiento, validación y evaluación. Esta división es crucial para garantizar que el modelo pueda aprender de los datos de entrenamiento, evitar el sobreajuste mediante la validación, y ser evaluado objetivamente con el conjunto de evaluación, lo que asegura su capacidad de generalización en datos no vistos.

Configuración del Algoritmo

En esta fase, se definió la arquitectura de la red neuronal y se optimizaron los hiperparámetros para maximizar su rendimiento.

Optimización de Hiperparámetros: Inicialmente, se intentó la optimización manual de hiperparámetros, pero debido a la gran cantidad de combinaciones posibles (288), el proceso resultó ser muy lento. En lugar de seguir con este enfoque, se utilizaron herramientas automáticas como GridSearchCV y Keras Tuner. Aunque GridSearchCV es eficaz para la validación cruzada, presentó problemas de compatibilidad con Google Colab y Python, por lo que se optó por Keras Tuner, que resultó ser más eficiente.

Arquitectura de la Red Neuronal: Se diseñó una red neuronal artificial (ANN) para clasificación multiclase, compuesta por una capa de entrada con 10 neuronas, que reciben las 10 variables seleccionadas. Este número de neuronas se eligió para reflejar la dimensionalidad de las características de entrada.

Capas Ocultas: Se incluyeron dos capas ocultas: la primera con 64 neuronas y la segunda con 32 neuronas, para permitir que la red neuronal modelara patrones no lineales y complejos en los datos.

Capa de Salida: La capa de salida está formada por 3 neuronas, que corresponden a las tres clases de litología. Este diseño permite clasificar las muestras en las tres categorías de litología (clases 1, 2 y 3).

Optimización y Técnicas de Regularización: Para mejorar el rendimiento y prevenir el sobreajuste, se emplearon diversas técnicas. Se utilizó el algoritmo Adam para la optimización, aprovechando su capacidad de ajuste dinámico de tasas de aprendizaje. Además, se integraron técnicas como Batch Normalization para normalizar las activaciones y Dropout para desactivar aleatoriamente neuronas durante el entrenamiento. Estas técnicas ayudaron a mejorar la generalización del modelo.

Con esta configuración, se buscó un modelo robusto y preciso que aprenda eficazmente las relaciones entre las variables y genere predicciones de alta calidad para la clasificación de litologías. La configuración final se evaluó mediante métricas como precisión, pérdida y la matriz de confusión.

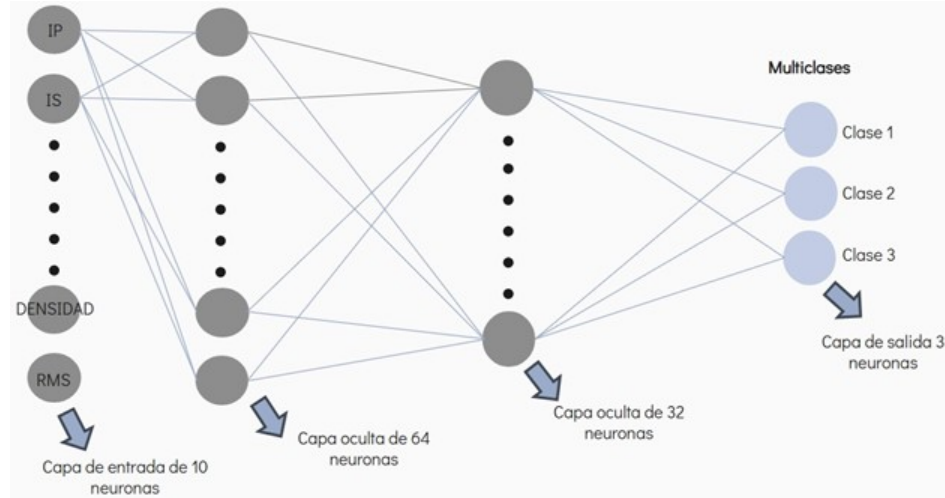


Figura 4. Esquema de la arquitectura del modelo final utilizado para la clasificación de litologías. Se implementaron técnicas de regularización como Batch Normalization y Dropout para mejorar la generalización del modelo.

Entrenamiento del Modelo

El primer modelo se entrenó considerando las 13 variables Tabla 3 independientes, con el objetivo de establecer un valor de precisión de referencia (0.9). Posteriormente, se evaluaron dos enfoques: el método de selección progresiva y el regresivo.

MODELO	VARIABLES INDEPENDIENTES													VARIABLE DEPENDIENTE	PRECISION
	COORDENADA X	COORDENADA Y	COORDENADA Z	IP	IS	DENSIDAD	POISSON	VIPVS	DECOMP	SIMILARIDAD	ENVOLVENTE	CURVATURA	RMS	LITOLOGIA	
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0.9

Tabla 3. Primer modelo utilizado con la presencia de 13 variables.

Secuencia Progresiva

Se evaluó cada variable independiente por separado para determinar su precisión individual Tabla 4. Con base en estos resultados, se combinaron las dos variables con mayor precisión y, si esta mejoraba, se añadía la siguiente variable con mejor desempeño, repitiendo el proceso de forma sucesiva. Sin embargo, dado que la precisión no mostró mejoras, se optó por la secuencia regresiva.

MODELO	VARIABLES INDEPENDIENTES													VARIABLE DEPENDIENTE	PRECISION
	COORDENADA X	COORDENADA Y	COORDENADA Z	IP	IS	DENSIDAD	POISSON	VIPVS	DECOMP	SIMILARIDAD	ENVOLVENTE	CURVATURA	RMS	LITOLOGIA	
2	X													X	0.67
3		X												X	0.68
4			X											X	0.68
5				X										X	0.76
6					X									X	0.76
7						X								X	0.74
8							X							X	0.7
9								X						X	0.62
10									X					X	0.65
11										X				X	0.67
12											X			X	0.71
13												X		X	0.63
14													X	X	0.71
15				X	X									X	0.73
16				X		X								X	0.74
17					X	X								X	0.7

Tabla 4. Evaluación de las variables independientes

Secuencia Regresiva

Se entrenó el modelo utilizando todas las variables y eliminando una a la vez para evaluar el impacto en la precisión como se muestra en la Tabla 5. A partir de estos resultados, se identificaron y descartaron las dos variables cuya omisión mejoraba el rendimiento del modelo, repitiendo el proceso de omitir variables por su impacto en la mejora de la precisión. Siguiendo esta metodología, se alcanzó una precisión máxima del 95 % empleando 10 variables independientes, omitiendo las variables de decomp, curvatura y similitud.

MODELO	VARIABLES INDEPENDIENTES														VARIABLE DEPENDIENTE	PRECISION
	COORDENADA X	COORDENADA Y	COORDENADA Z	IP	IS	DENSIDAD	POISSON	VPVS	DECOMP	SIMILARIDAD	ENVOLVENTE	CURVATURA	RMS	LITOLOGIA		
18	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0.88	
19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	0.92	
20	X	X	X	X	X	X	X	X	X	X	-	X	X	X	0.87	
21	X	X	X	X	X	X	X	X	X	-	X	X	X	X	0.93	
22	X	X	X	X	X	X	X	X	-	X	X	X	X	X	0.93	
23	X	X	X	X	X	X	X	-	X	X	X	X	X	X	0.92	
24	X	X	X	X	X	X	-	X	X	X	X	X	X	X	0.89	
25	X	X	X	X	X	X	-	X	X	X	X	X	X	X	0.87	
26	X	X	X	X	-	X	X	X	X	X	X	X	X	X	0.86	
27	X	X	X	-	X	X	X	X	X	X	X	X	X	X	0.85	
28	X	X	-	-	X	X	X	X	X	X	X	X	X	X	0.9	
29	X	-	X	X	X	X	X	X	X	X	X	X	X	X	0.9	
30	-	X	X	X	X	X	X	X	X	X	X	X	X	X	0.91	
31	X	X	X	X	X	X	X	-	-	X	X	X	X	X	0.94	
32	X	X	X	X	X	X	X	-	-	X	X	-	X	X	0.95	
33	X	X	X	X	X	X	X	-	-	-	X	-	X	X	0.92	

Tabla 5. Evaluación de una variable a la vez para determinar su precisión

Jerarquización de Variables

Se realizó la jerarquización de las 10 variables independientes utilizando el impacto que provocan al omitirlas del modelo. Se aplicó la función Softmax a la jerarquía de variables para ponderar la influencia de cada una en el modelo.

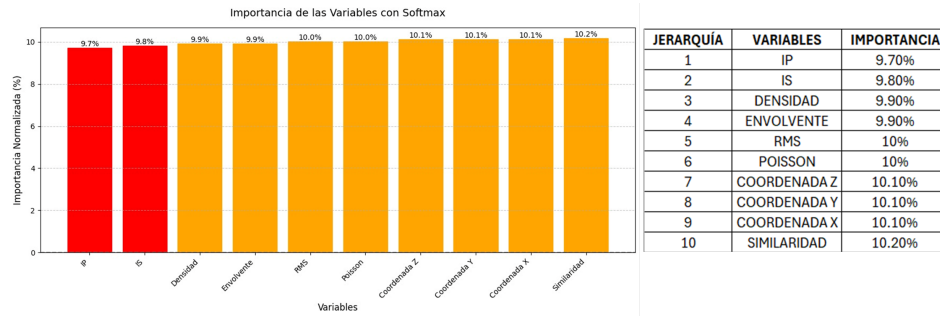


Figura 5. Importancia de las variables utilizando la función Softmax

4. Experimentación y resultados

Uno de los enfoques clave de este estudio radica en evaluar la importancia de las variables para predecir litologías, lo que hace que la ingeniería de características sea fundamental para su éxito.

Los atributos sísmicos seleccionados permiten, en mayor o menor medida, la detección de cambios en la litología. Gracias a la extracción de datos en las ubicaciones de los pozos, fue posible obtener los valores cuantitativos de cada atributo. Entre estos, la impedancia acústica (en $\text{g/cm}^3\cdot\text{m/s}$) se destaca por su capacidad para identificar contrastes entre diferentes litologías y delimitar la presencia de hidrocarburos. De manera complementaria, la impedancia S (en $\text{g/cm}^3\cdot\text{m/s}$) juega un papel crucial en la interpretación de fluidos, ya que las ondas S no se propagan en líquidos, lo que permite distinguir zonas con gas o agua en los yacimientos. La densidad (en g/cm^3), por su parte, proporciona información clave sobre la composición del subsuelo, ya que está relacionada con la masa por unidad de volumen de la roca.

La relación de Poisson (adimensional) también resulta ser un indicador relevante, pues las rocas con gas tienden a tener valores inferiores a 0.2, lo que las diferencia de otras formaciones más compactas o porosas. De forma similar, la relación VP/VS (adimensional), que mide la diferencia entre las velocidades de ondas P y S, presenta valores superiores a 2.5 en presencia de gas.

Otros atributos importantes incluyen la descomposición espectral (en Hz), que permite analizar la distribución de frecuencias de la señal sísmica, y la similaridad (adimensional), que facilita la distinción entre capas con propiedades semejantes, sugiriendo posibles similitudes litológicas. La envolvente (en unidades de amplitud) resulta útil para resaltar cambios de energía sin depender de la polaridad de la onda, mientras que la RMS (en amplitud o m/s) mide la amplitud promedio de la señal sísmica, lo que ayuda a identificar zonas con mayor porosidad, fracturación o presencia de fluidos. Finalmente, la curvatura (en $1/\text{m}$) se incluyó para analizar variaciones en la geometría de las capas rocosas, proporcionando información clave sobre la deformación estructural del subsuelo. Atributos como la curvatura fueron seleccionados con el objetivo de comparar la jerarquía de variables y evaluar su importancia teórica en la caracterización del subsuelo.

En la Tabla 6 se presenta la estructura del dataset utilizado, el cual contiene atributos sísmicos cuantitativos.

COORD_X	COORD_Y	COORD_Z	P	IS	DENSIDAD	POISSON	VPVS	DECOMP	SIMILARIDAD	ENVOLVENTE	CURVATURA	RMS	LITO
234.21305	122951	1248	5351.425575	2295.025459	2.391796112	0.387223961	2.330099421	15412.80371	0.985106409	1060.604452	0.007624985	1143.079022	3
234.21305	122951	1250	5353.695313	2313.385742	2.372696396	0.382266402	2.20090545	73078.96675	0.98235996	673.3432007	0.000266504	1144.04136	3
234.6199.25	122199.9766	1002	4931.963281	2263.584961	2.149603439	0.37212401	2.215996089	210944.9219	0.988076036	5005.411621	-0.020183742	4013.972968	1
234.30415	122732.2891	834	4536.515625	1940.0625	2.156702042	0.402370572	2.474547797	234515.4375	0.999891766	1536542.725	0.067099623	6215.399454	2
234.30415	122732.2891	874	4422.195313	2036.529541	2.196399953	0.393351316	2.38501358	9843.78273	0.971113622	4637341.757	0.08528975	6997.03305	2
234.6199.25	122199.9766	1004	4977.023438	2193.942393	2.185999948	0.37590255	2.241829057	483708.2813	0.986733674	3750322.998	-0.021317413	4431.399629	1

Tabla 6. Estructura del Dataset

Posteriormente, se llevó a cabo el entrenamiento de un modelo supervisado. Este enfoque integra datos de pozos con información sísmica, proporcionando una solución para la clasificación de litologías. El objetivo principal de este estudio es la Predicción de litología basada en jerarquía de atributos sísmicos con Machine Learning. Para ello, se implementaron técnicas avanzadas de clasificación supervisada, realizando una jerarquización de las variables y experimentación con varios hiperparámetros.

La computadora empleada cuenta con 64 GB de memoria RAM, lo que permite manejar grandes volúmenes de datos de manera eficiente. Además, se utilizó almacenamiento en la nube para gestionar la información sísmica. El lenguaje de programación utilizado fue Python, junto con bibliotecas especializadas en Machine Learning y Deep Learning. Para optimizar los resultados, se probaron múltiples configuraciones del modelo, variando la combinación de variables, el número de capas y la cantidad de neuronas. Como resultado, se obtuvo un modelo con altos valores de predicción y una matriz de confusión aceptable, validando así su desempeño en la clasificación de litologías.

Curvas de Aprendizaje

La curva de pérdida del entrenamiento muestra una disminución progresiva a medida que aumentan las épocas, lo que resulta favorable para el modelo. Figura 4 De manera similar, la curva de precisión se incrementa con el número de épocas, y tanto las curvas de entrenamiento como de validación mantienen una trayectoria paralela. Esto indica un entrenamiento óptimo, sin evidencias de sobreentrenamiento.

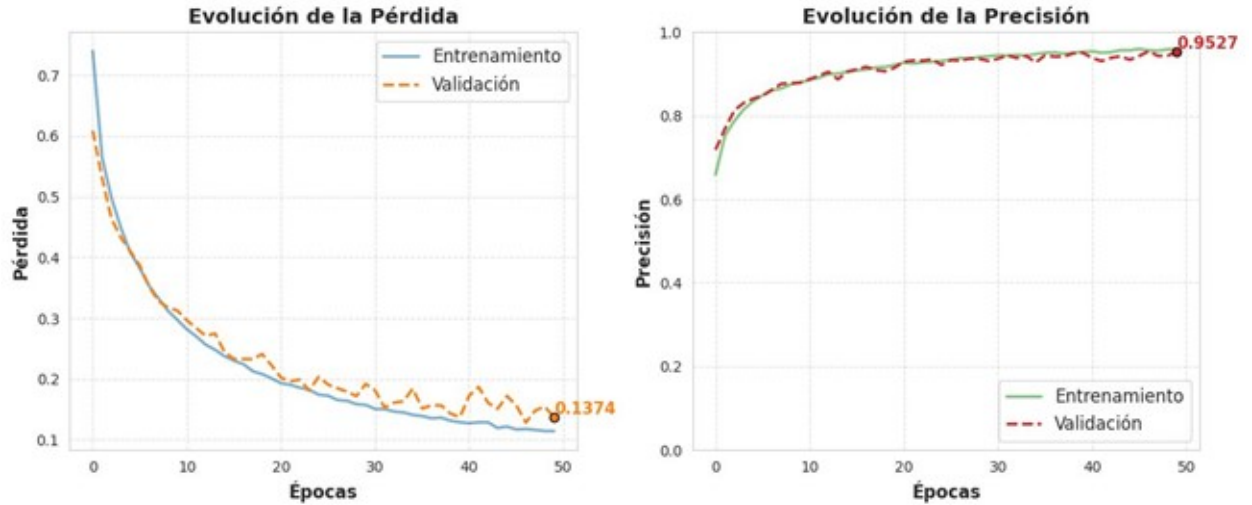


Figura 6. La curva de pérdida muestra una disminución constante a lo largo de las épocas. La curva de precisión aumenta con el número de épocas, manteniendo una trayectoria paralela en entrenamiento y validación.

Matriz de Confusión

La matriz de confusión en la etapa de evaluación refleja altos niveles de precisión en todas las clases. La clase 1, que representa las lutitas, alcanzó una precisión del 93.33 %. La clase 2, correspondiente a las areniscas sin hidrocarburo, obtuvo un 94.76 %. Finalmente, la clase 3, de mayor importancia por representar las areniscas con hidrocarburo, logró una precisión del 98.27 %. (Figura 5)

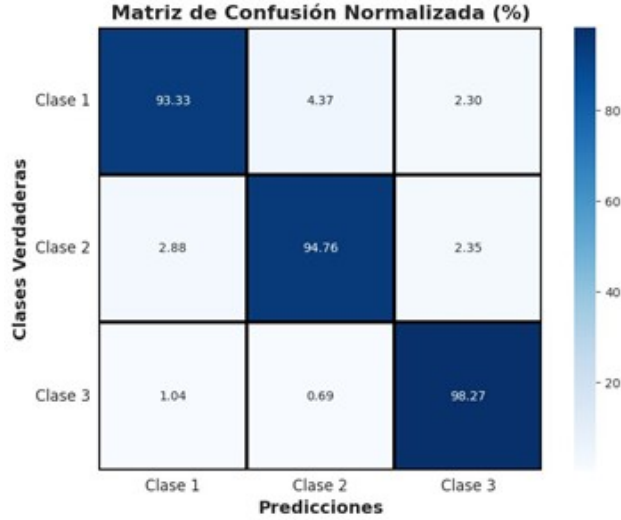


Figura 7. La matriz de confusión presenta una alta precisión en todas las clases 1 (lutitas), 2 (areniscas sin hidrocarburo) y 3 (areniscas con hidrocarburo).

Visualización 2D y 3D

A partir del volumen sísmico y el archivo .HSEP generado en el código, ambos fueron importados al programa especializado en sísmica InterWell. Se ajustaron los colores de las clases, asignando verde a las lutitas, amarillo a las areniscas y marrón a las areniscas con hidrocarburos. En la Figura X se presenta la vista 3D del volumen sísmico, mientras que en la Figura se muestra la visualización en sección Inline, donde se añadió uno de los pozos con su respectiva caracterización litológica. En la zona delimitada por un círculo rojo se observa que la predicción no es precisa. Sin embargo, a mayor profundidad, en los reservorios más grandes, las predicciones se ajustan con mayor exactitud, especialmente en las areniscas que contienen gas, la litología de mayor interés para este estudio.

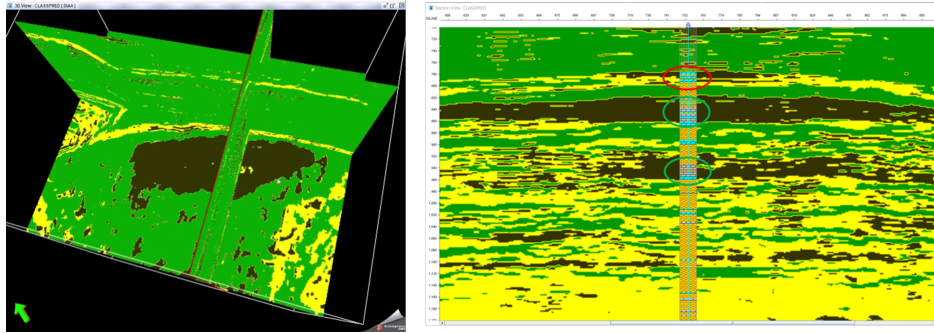


Figura 8. Visualización del cubo sísmico utilizado en el análisis de litologías. La imagen representa un volumen sísmico 3D y representación 2D dentro del cual se han aplicado los diferentes atributos sísmicos.

5. Discusión

Los resultados obtenidos en este estudio confirman la efectividad del uso de redes neuronales supervisadas y atributos sísmicos en la predicción de litologías. La metodología implementada permitió jerarquizar los atributos sísmicos, identificando aquellos con mayor relevancia en la caracterización del subsuelo. Esto representa un avance significativo respecto a los enfoques tradicionales, que dependen en gran medida del análisis manual y subjetivo de registros de pozos.

Uno de los hallazgos más relevantes es que la combinación de atributos sísmicos con redes neuronales mejora la precisión de la predicción de litologías, lo que se traduce en una interpretación más confiable de la distribución de los yacimientos. Además, la metodología propuesta permite reducir el tiempo de análisis y optimizar los recursos utilizados en la exploración y explotación de hidrocarburos.

En los primeros intentos de equilibrar las clases del dataset, se probó el uso de class weights con el fin de abordar el desbalance entre las clases mayoritarias y minoritarias. Sin embargo, esta técnica no resultó ser efectiva, ya que no logró mejorar significativamente la representación de las clases minoritarias (2 y 3) en el modelo. Como consecuencia, se optó por un enfoque de balanceo manual utilizando software especializado, lo que permitió añadir más detalle y resolución a las clases minoritarias.

Este ajuste no solo aumentó la representatividad de las clases menos frecuentes, sino que también mejoró la calidad del dataset, generando un conjunto de datos equilibrado con una mayor cantidad de información útil. Durante este proceso, se detectó un problema en la herramienta de extracción, que al procesar volúmenes sísmicos más pequeños que otros, asignaba valores de "0.^{en} lugar de "NaN". Este error, que podría haber afectado la calidad de los datos, fue reportado a los desarrolladores para su corrección. Tras estos ajustes, el dataset equilibrado mostró mejoras en el desempeño de los modelos de entrenamiento, permitiendo una mayor precisión en la clasificación y una mejor generalización de los resultados.

Se realizó un análisis exploratorio de datos con el objetivo de caracterizar la distribución y las características generales de cada variable. A partir de este análisis, se identificaron tres aspectos clave: la presencia de sesgos en varias variables, el desbalanceo en la distribución de clases de litología y la existencia de valores incorrectos en las variables Densidad, Poisson y Vp/Vs, los cuales fueron generados por la herramienta utilizada y no representaban datos válidos. Para evaluar el impacto de estos factores en el modelo predictivo, se realizaron pruebas sin aplicar correcciones. Se observó que los valores atípicos y la asimetría tenían un efecto limitado en la precisión global del modelo basado en redes neuronales supervisadas. Sin embargo, la matriz de confusión reflejó una mejora en la predicción de las clases 2 y 3, con un incremento aproximado del 0.5 % tras la aplicación de transformaciones a las variables sesgadas.

Para mitigar estos sesgos, se empleó el Power Transformer con el método Yeo-Johnson, ajustando la distribución de las variables hacia una forma más normal y reduciendo la influencia de valores atípicos. Sin embargo, su eficacia varió según la variable analizada. Mientras que en Similaridad el método logró corregir el sesgo de manera efectiva, en Envolvente la distribución no se normalizó completamente y algunos valores atípicos persistieron. Ante estas diferencias en el desempeño de la transformación, se optó por conservar las variables en su estado original y abordar la normalización en una etapa posterior. Dado que la corrección de sesgos individuales no siempre fue efectiva, se decidió abordar la normalización mediante un escalador más robusto. Para este proceso, se empleó RobustScaler, una técnica basada en la mediana y el rango intercuartil, lo que minimiza la influencia de valores atípicos y mejora la homogeneidad en la representación de las variables. A diferencia de otros enfoques de normalización, RobustScaler permite mantener la integridad de la información crítica del dataset sin distorsionar la magnitud relativa de las variables.

El archivo tipo .HSEP se estructura de manera que complementa y organiza la información extraída del archivo SEG Y, permitiendo un análisis detallado de los atributos sísmicos. Mediante un fragmento de código específico, se logró leer y procesar el archivo SEG Y, que contiene las trazas sísmicas en su cubo de datos. Este cubo SEG Y incluye información crucial sobre las ubicaciones en Inline, Crossline y tiempo, lo que permite realizar un análisis completo de las características sísmicas en el volumen.

Específicamente, el modelo predictivo lee cada uno de los valores dentro de los cubos de los atributos sísmicos, y a partir de estos datos realiza las predicciones correspondientes para asignar valores a cada una de las posiciones dentro del cubo. A través de un proceso cíclico, el modelo asigna valores basados en las correlaciones observadas entre los atributos sísmicos de las trazas. Aunque el archivo .HSEP en sí no es particularmente pesado, la riqueza de la información reside en las trazas dentro del cubo SEG Y, que contiene una gran cantidad de datos sísmicos. Esto permite que el modelo realice predicciones más precisas y asignaciones de valores, contribuyendo a una mejor interpretación del volumen sísmico y a la clasificación de las litologías de interés en el estudio.

6. Conclusiones

Este estudio ha logrado desarrollar un modelo de predicción de litologías altamente efectivo, utilizando redes neuronales supervisadas y atributos sísmicos, lo que representa un avance significativo en la caracterización del subsuelo para la exploración de hidrocarburos. A través de un enfoque robusto que incluyó la creación de un dataset de calidad, la aplicación de técnicas de ingeniería de características y un análisis estadístico detallado, se optimizó el proceso de entrenamiento del modelo, alcanzando un notable 95 % de precisión en las predicciones.

Uno de los avances más importantes de este estudio ha sido la identificación y selección de las variables más relevantes para el modelo. Mediante un análisis exhaustivo, se discriminaron aquellos atributos sísmicos que realmente aportaron valor a la predicción, fundamentando esta selección tanto teóricamente como prácticamente. Variables como la impedancia acústica, la densidad, la impedancia S y la relación VP/VS demostraron ser determinantes en la clasificación precisa de las litologías, especialmente en la identificación de reservorios con gas. Esta validación tanto empírica como

teórica asegura que los atributos seleccionados no solo son adecuados desde el punto de vista técnico, sino que también proporcionan una base sólida para su implementación en futuros estudios.

Los resultados obtenidos validan la eficacia de la metodología propuesta, demostrando que la integración de atributos sísmicos con redes neuronales permite superar las limitaciones de los enfoques tradicionales, brindando una capacidad superior para predecir y clasificar litologías. Además, la visualización en 2D y 3D permitió identificar con claridad las geometrías de los reservorios de hidrocarburos, ofreciendo una visión más precisa y objetiva del subsuelo, lo que tiene un potencial considerable para mejorar la toma de decisiones en la industria petrolera.

Este estudio también resalta la importancia de contar con herramientas y metodologías abiertas, ya que el código desarrollado y disponible en fuente abierta proporciona una base sólida para futuras investigaciones y aplicaciones en el ámbito de la exploración geofísica. Con ello, se facilita la replicabilidad y el ajuste del modelo a diferentes contextos geológicos y de datos, lo que abre nuevas posibilidades para la optimización de los recursos energéticos.

Para futuros trabajos, se recomienda profundizar en el análisis de la importancia relativa de los atributos sísmicos, validar el modelo con nuevos datos de perforación y explorar arquitecturas de deep learning más avanzadas. Estos avances fortalecerán el uso de inteligencia artificial en la exploración geofísica, mejorando la eficiencia y reduciendo la incertidumbre en la toma de decisiones.

Agradecimientos

Queremos expresar nuestro más sincero agradecimiento al Dr. Christian Mejía, docente de la cátedra de Software Aplicado a la Geología, por transmitirnos su conocimiento sobre Machine Learning y por su valiosa orientación a lo largo de este trabajo. Su enseñanza ha sido fundamental para nuestra comprensión y aplicación de estas técnicas en el ámbito geológico.

Asimismo, agradecemos a la empresa Beicip-Franlab por facilitarnos la licencia del software InterWell, así como por proporcionarnos los datos necesarios para la realización de este estudio.

Finalmente, extendemos nuestra gratitud a Romain Baillet y Sarra Ben-Brahin por su apoyo y sus valiosas ideas, que fueron esenciales para el desarrollo y mejora del código implementado en este trabajo.

Conflicto de interés

Este proyecto fue desarrollado con fines investigativos como parte del proyecto semestral de la cátedra de Software Aplicado a la Geología. La línea de investigación sobre jerarquía de variables fue impulsada por la empresa Beicip-Franlab, en la cual uno de los autores se encuentra vinculado.

Queda estrictamente prohibido el uso del código desarrollado en este trabajo con fines lucrativos sin la autorización expresa de los autores. Su utilización debe limitarse a propósitos académicos e investigativos, respetando siempre los principios de acceso y uso ético del conocimiento.

Disponibilidad de datos y código

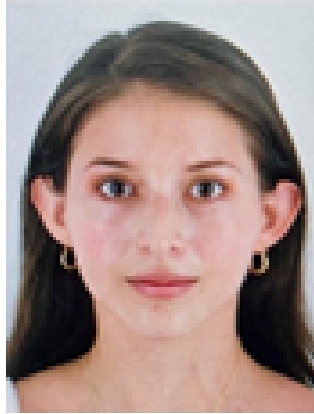
Los datos utilizados en este estudio son confidenciales y no pueden ser compartidos públicamente, en caso de ser necesario comunicarse al correo meveliz@uce.edu.ec. Sin embargo, el código que soporta los hallazgos del presente estudio está libremente disponible en: [<https://colab.research.google.com/drive/1bvWgVrqGZbZ9WnTJ1WDoE56SpFtGN1M?usp=sharing>].

Contribuciones de los autores

Todos los autores de este artículo participaron activamente en todas las fases y etapas de este proyecto, incluyendo la conceptualización, desarrollo metodológico, implementación del código, análisis de resultados y redacción del manuscrito.

Foto y biografía de los autores

María Elena Véliz Zambrano



Estudiante de décimo semestre y tesista de la carrera de Ingeniería en Geología en la Universidad Central del Ecuador. Actualmente, se desempeña como soporte local para Latinoamérica de la herramienta geofísica de Beicip-Franlab, donde aplica sus conocimientos técnicos en el ámbito de la inversión y caracterización sísmica. Sus principales áreas de interés incluyen la geofísica, geología del petróleo y estratigrafía.

Roberto Sebastián Andrade Albán



Estudiante de décimo semestre de la carrera de Ingeniería en Geología en la Universidad Central del Ecuador. Actualmente, se desempeña como pasante en la empresa petrolera PCR-ECUADOR, donde aplica sus conocimientos técnicos. Sus principales áreas de interés incluyen la geofísica, geología del petróleo y geología del subsuelo.

Joan Sebastian Torres Lasso



Estudiante de décimo semestre de Ingeniería en Geología en la Universidad Central del Ecuador. Tesista en Petroecuador, trabajando en la evaluación del potencial hidrocarburífero en los campos Tetete Tapi y Frontera. Con interés en petrofísica y geología del petróleo, enfocado en la caracterización de yacimientos y la optimización de la exploración de hidrocarburos.

Soledad Estefanía Berrones Garcés



Estudiante de décimo semestre de la carrera de Ingeniería en Geología en la Universidad Central del Ecuador. Pasante PCR-ECUADOR cumpliendo funciones como la descripción de rípios de perforación de los pozos Sami y Suyana, prácticas en software Petrel y Techlog. (julio 2024-enero 2025).

Referencias

- Alférez, S., Chen, Y., Li, J. (2024). Advanced machine learning artificial neural network applications in lithology identification. *Frontiers in Earth Science*. Recuperado de <https://www.frontiersin.org/articles/10.3389/feart.2024.1473325/full>
- Alférez, S., Chen, Y., 'I&' Li, J. (2024). Real-Time Lithology Prediction at the Bit Using Machine Learning. *Geosciences*, 14(10), 250. Recuperado de <https://www.mdpi.com/2076-3263/14/10/250>
- Alnuaimi, N. (2019). Formation Lithology Classification: Insights into Machine Learning Algorithms and Feature Selection Techniques. *SPE Annual Technical Conference and Exhibition*. Recuperado de <https://onepetro.org/SPEATCE/proceedings/19ATCE/19ATCE/D021S033R005/217664>
- Berzal, F. (s.f.). Redes Neuronales Deep Learning. Recuperado de <https://scholar.google.com/citations?hl=enuser=RbaQAtoAAAAJ>
- Caballero Tovar, L. Y. (s.f.). Diseño de un sistema automático de recepción, almacenamiento y procesamiento de datos experimentales. Recuperado de <https://scholar.google.com/citations?user=jfGikp0AAAAJ>
- Boiger, R., Churakov, S. V., Ballester Llagaria, I., Kosakowski, G., Wüst, R., Prasianakis, N. I. (2024). Direct mineral content prediction from drill core images via transfer learning. Recuperado de <https://arxiv.org/abs/2403.18495>
- Brown, N., Roubickova, A., Lampaki, I., MacGregor, L., Ellis, M., Vera de Newton, P. (2020). Machine learning on Crays to optimise petrophysical workflows in oil and gas exploration. Recuperado de <https://arxiv.org/abs/2010.02087>
- Chaki, S. (2015). Reservoir Characterization: A Machine Learning Approach. Recuperado de <https://arxiv.org/abs/1506.05070>
- Chen, Y., Alférez, S., Li, J. (2024). Advanced machine learning artificial neural network applications in lithology identification. *Frontiers in Earth Science*. Recuperado de <https://www.frontiersin.org/articles/10.3389/feart.2024.1473325/full>
- Ehsan, M., Manzoor, U., Chen, R., Hussain, M., Abdelrahman, K., Radwan, A. E., Ullah, J., Iftikhar, M. K., Arshad, F. (2024). Pore pressure prediction based on conventional well logs and seismic data using an advanced machine learning approach. *Journal of Rock Mechanics and Geotechnical Engineering*. <https://doi.org/10.1016/j.jrmge.2024.09.049>
- Mishra, A., Sharma, A., Patidar, A. K. (2022). Evaluation and development of a predictive model for geophysical well log data analysis and reservoir characterization: Machine learning applications to lithology prediction. *Natural Resources Research*, 31, 3195–3222. Recuperado de <https://link.springer.com/article/10.1007/s11053-022-10121-z>

- Nguyen, M., Nguyen, H., Thi Bui, N. (2023). Aplicación de redes neuronales artificiales y atributos sísmicos para predecir la distribución de areniscas del Oligoceno tardío en la cuenca de Cuu Long. Hanoi University of Mining and Geology, Hanoi, Vietnam.
- Nordloh, V. A., Roubícková, A., Brown, N. (2020). Machine Learning for Gas and Oil Exploration. Recuperado de <https://arxiv.org/abs/2010.04186>
- Ntibahanana, M., Jianguo, S., Luemba, M., Tondozi, K., Imani, G., Mohamed, B. (2024). Ensemble of neural networks utilizing seismic attributes for rock-property inversion with uncertainty estimation. *Earth and Space Science*, 11, e2023EA003101. <https://doi.org/10.1029/2023EA003101>
- Oumarou, S., Mabrouk, D., Tabod, C. T., Marcel, J., Ngos, S. III, Abate Essi, J. M., Kamguia, J. (2021). Seismic attributes in reservoir characterization: An overview. *Arabian Journal of Geosciences*, 14(402). <https://doi.org/10.1007/s12517-021-06626-1>
- Ren, X., Hou, J., Song, S., Liu, Y., Chen, D., Wang, X. (2019). Lithology identification using well logs: A method by integrating artificial neural networks and sedimentary patterns. *Journal of Petroleum Science and Engineering*, 182, 106336. Recuperado de <https://doi.org/10.1016/j.petrol.2019.106336>
- Ružić, T., Cvetković, M. (2021). Geological characterization of the 3D seismic record within the gas bearing upper Miocene sediments in the northern part of the Bjelovar subdepression—Application of amplitude versus offset analysis and artificial neural network. *Energies*, 14(14), 4161. <https://doi.org/10.3390/en14144161>
- Saikia, P., Baruah, R. D., Singh, S. K., Chaudhuri, P. K. (2019). Artificial Neural Networks in the domain of reservoir characterization: A review from shallow to deep models. *Computers and Geosciences*. <https://doi.org/10.1016/j.cageo.2019.104357>
- Sharma, A., Sharma, R., Upadhaya, P., Ram, J. (s.f.). Seismic attribute analysis and neural network application for reservoir characterization in Lanwa Field, India: A case study. 10th Biennial International Conference Exposition, IRS, ONGC, Chandkheda, Ahmedabad.
- Shan, L., Liu, Y., Tang, M., Yang, M., Bai, X. (2021). CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. *Journal of Petroleum Science and Engineering*, 205, 108838. Recuperado de <https://doi.org/10.1016/j.petrol.2021.108838>
- Tian, M., Omre, H., Xu, H. (2021). Inversion of well logs into lithology classes accounting for spatial dependencies by using hidden Markov models and recurrent neural networks. *Journal of Petroleum Science and Engineering*, 196, 107598. Recuperado de <https://doi.org/10.1016/j.petrol.2020.107598>
- Wang, K., Zhang, L. (2021). Predicting formation lithology from log data by using a neural network. *Petroleum Science*, 5(3), 251–257. <https://doi.org/10.1007/s12182-008-0038-9>
- Xie, Y., Zhu, C., Zhou, W. (2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering*, 160, 182-193. Recuperado de <https://doi.org/10.1016/j.petrol.2017.10.026>
- Zeng, L., Ren, W., Shan, L., Huo, F. (2022). Well logging prediction and uncertainty analysis based on recurrent neural network with attention mechanism and Bayesian theory. *Journal of Petroleum Science and Engineering*, 208, 109458. Recuperado de <https://doi.org/10.1016/j.petrol.2021.109458>