**PREDICCIÓN DE POTENCIALES DEPÓSITOS DE LITIO (Li) MEDIANTE EL USO DE DATOS GEOQUÍMICOS Y MACHINE LEARNING EN COLOMBIA**

Luis Aguirre

Natalia Correa

Diego Ramírez

Bootcamp: Análisis de datos – Integrador

Septiembre de 2025

**Introducción**

La creciente dependencia global de los sistemas de almacenamiento de energía, impulsada por el transporte eléctrico, ha generado un incremento exponencial en la producción de baterías de iones de litio recargables, consolidando al litio (Li) como un elemento de importancia estratégica mundial. Para satisfacer la demanda actual y asegurar la cobertura de las necesidades futuras de Li y otros metales críticos como el cesio (Cs), rubidio (Rb), estaño (Sn), tantalio (Ta) y niobio (Nb), resulta vital el descubrimiento y la caracterización de nuevos depósitos en pegmatitas de litio-cesio-tantalio (LCT). Esta necesidad se alinea directamente con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas, específicamente el ODS 7 (energía limpia y asequible) y el ODS 11 (ciudades y comunidades sostenibles).

No obstante, la exploración de las pegmatitas LCT es inherentemente costosa, lenta, y se ve dificultada por su tamaño relativamente pequeño (típicamente menos de 50 metros de ancho) y el enmascaramiento por materiales de cobertura como el suelo y vegetación. Los métodos tradicionales de exploración, incluyendo la geoquímica de suelos, lateritas o sedimentos, aunque moderadamente exitosos, presentan limitaciones significativas. Las técnicas de análisis de muestra total, como la espectrometría de masas con plasma acoplado inductivamente (ICP-MS) o la espectroscopia de emisión óptica (ICP-OES), son costosas y lentas, y el uso de ácidos corrosivos para la digestión ácida puede estar restringido por razones de seguridad o medioambientales. Además, las tecnologías analíticas portátiles como la fluorescencia de rayos X portátil (pXRF), si bien es rápida y de bajo costo, no puede detectar directamente el litio. Esta limitación se debe a que la pXRF solo puede cuantificar elementos trazadores más pesados que el magnesio que se asocian al Li en las pegmatitas LCT (como K, Rb, Cs, Ba, Sr, Cu, Zn, Pb, Ga, Sn, Ta, ± W).

La urgencia de desarrollar métodos de exploración más económicos, rápidos y sostenibles para la identificación de depósitos de Li, ha impulsado la integración de estrategias de análisis de datos avanzadas. Estudios recientes han demostrado que es posible superar la limitación del pXRF mediante el uso de algoritmos de aprendizaje automático (Machine Learning, ML), para predecir el contenido de Li en rocas basándose en las concentraciones de elementos trazadores (Cs, U, Zn, As, Th, W, Sc, Y, Tb, Ga, Dy, In, Fe, Ho, Co, Er) y la relación K/Rb. El uso de ML ha demostrado ser robusto, logrando altas precisiones (un modelo de red neuronal para la predicción de Li alcanzó un R² de 0.90 y un RPD de 3.2, indicando una predicción precisa).

El presente proyecto se enfoca en aplicar y adaptar esta innovadora metodología de predicción al contexto geológico y geoquímico de Colombia. Colombia, en su estrategia de transición energética, requiere la identificación y caracterización de sus recursos de minerales estratégicos. La adaptación de este enfoque basado en la geoquímica y ML representa una alternativa más eficiente y menos invasiva en comparación con los métodos tradicionales, lo que puede reducir costos y acelerar la prospección de Li. Los resultados de este análisis beneficiarán directamente al Servicio Geológico Colombiano y a las empresas exploradoras, al proporcionar una herramienta de toma de decisiones informada y eficiente para dirigir los esfuerzos de prospección en campo.

**Pregunta de Investigación**

¿Cómo podemos optimizar y hacer más eficiente la identificación de potenciales depósitos de litio en Colombia, superando las limitaciones de la exploración geoquímica tradicional y la incapacidad de la pXRF para detectar litio directamente, mediante la adaptación y el reentrenamiento de algoritmos avanzados de aprendizaje automático utilizando datos geoquímicos disponibles?

# Objetivos

**Objetivo General**

Adaptar y validar la metodología de predicción geoquímica de litio (Li) mediante algoritmos de aprendizaje automático (Machine Learning) al contexto colombiano, utilizando elementos trazadores cuantificables por fluorescencia de rayos X portátil (pXRF), con el fin de desarrollar una herramienta de prospección más rápida, económica y eficiente para la identificación de zonas con potencial de pegmatitas de litio-cesio-tantalio (LCT) en el país.

**Objetivos Específicos**

1. Gestionar y estructurar la base de datos geoquímicos: Estructurar una base de datos geoquímica (en Access, MySQL y Pandas), que incluya las concentraciones de litio (Li) y los 16 elementos trazadores detectables por pXRF (Cs, U, Zn, As, Th, W, Sc, Y, Tb, Ga, Dy, In, Fe, Ho, Co, Er) y la relación K/Rb, asegurando la consistencia y normalización de los datos para su posterior análisis.

2. Modelar la predicción de litio: Diseñar y entrenar algoritmos de aprendizaje automático, para predecir el contenido de Li en suelos, adaptando la metodología de Pereira et al. (2025) al contexto geológico de Colombia.

3. Evaluar la eficacia y precisión de los modelos: Evaluar el desempeño de los modelos predictivos mediante métricas clave como el coeficiente de determinación, el error cuadrático medio y la desviación de predicción residual para la predicción de contenido de Li, y la precisión general, con el objetivo de seleccionar el algoritmo más robusto para la exploración en Colombia.

4. Generar herramientas de visualización interactiva: Crear visualizaciones de datos interactivas (utilizando herramientas como Pandas, Matplotlib, y Seaborn) que permitan explorar y comunicar claramente los patrones geoquímicos identificados y los resultados de las predicciones de los modelos de Machine Learning, facilitando así la toma de decisiones informada por parte del Servicio Geológico Colombiano y empresas exploradoras.

**Dataset**

| **Fuente Primaria** | Datos geoquímicos de rocas y sedimentos publicados en bases de datos libres del Servicio Geológico Colombiano (SGC) |
| --- | --- |
| **Enlace de Descarga** | https://www2.sgc.gov.co/sgc/mapas/Paginas/Bases-Datos.aspx |
| **Licencia** | Los datos provienen de bases de datos libres |

**Número de Registros, Variables y Periodo de Cobertura**

El conjunto de datos debe ser estructurado de modo que cumpla con los requisitos del modelado de ML, basándose en la metodología validada en el estudio de referencia (Pereira et al., 2025).

• **Número de variables:** 17 variables de entrada/predicción (concentraciones de elementos trazadores y la relación K/Rb) más la variable objetiva (Li).

• **Número de registros:** El número exacto de registros iniciales en la base de datos del Servicio Geológico Colombiano es, 168656 filas x 233 columnas. Sin embargo, el dataset final solo contiene los 97 datos que cumplen con los requisitos analíticos requeridos para el modelo ML (detección de Li y elementos trazadores).

• **Periodo de Cobertura:** Los datos disponibles en la base de datos fueron obtenidos hasta el año 2012.

**Diccionario de Variables**

El dataset combina mediciones de alta precisión del elemento objetivo (Li, obtenido típicamente por ICP-OES en el estudio de referencia) con las concentraciones de elementos trazadores medidos por pXRF. Todas las concentraciones de elementos se expresan en miligramos\*kilogramo (o ppm).

| **Nombre de la Variable** | **Tipo de Dato** | **Descripción** | **Método Analítico de Referencia** |
| --- | --- | --- | --- |
| Li | Numérico (Float) | Concentración de Litio, variable objetivo del modelo ML | ICP-OES (Espectroscopia de Emisión Óptica). |
| K | Numérico (Float) | Concentración de Potasio. | pXRF (Fluorescencia de Rayos X Portátil). |
| Rb | Numérico (Float) | Concentración de Rubidio. | pXRF. |
| K/Rb | Numérico (Float) | Razón K/Rb. Variable crucial que correlaciona negativamente con el grado de fraccionamiento de las pegmatitas LCT y su prospectividad. | Derivada de pXRF. |
| Cs | Numérico (Float) | Concentración de Cesio. Importante trazador en pegmatitas LCT | pXRF. |
| U | Numérico (Float) | Concentración de Uranio. | pXRF. |
| As | Numérico (Float) | Concentración de Arsénico. | pXRF. |
| Th | Numérico (Float) | Concentración de Torio. | pXRF. |
| W | Numérico (Float) | Concentración de Wolframio. | pXRF. |
| Sc | Numérico (Float) | Concentración de Escandio. | pXRF. |
| Fe | Numérico (Float) | Concentración de Hierro. | pXRF. |
| Y | Numérico (Float) | Concentración de Itrio. | pXRF. |
| Tb | Numérico (Float) | Concentración de Terbio. | pXRF. |
| Ga | Numérico (Float) | Concentración de Galio. | pXRF. |
| Ln | Numérico (Float) | Concentración de Lantano. | pXRF. |
| Ho | Numérico (Float) | Concentración de Holmio. | pXRF. |
| Co | Numérico (Float) | Concentración de Cobalto. | pXRF. |
| Er | Numérico (Float) | Concentración de Erbio. | pXRF. |
| Zn | Numérico (Float) | Concentración de Zinc. | pXRF. |

**Limitaciones y sesgos del dataset**

La adaptación de la metodología de Pereira et al. (2025) al contexto colombiano introduce limitaciones y sesgos que deben ser considerados:

**Limitaciones intrínsecas a la detección de Litio**

• **Detección Indirecta:** La principal limitación es que el pXRF no puede detectar directamente el litio (Li). La metodología se basa en predecir las concentraciones de Li a partir de la correlación con elementos trazadores.

• **Dependencia de datos de referencia:** El modelado (ML) requiere que el conjunto de datos posea tanto las concentraciones de los elementos trazadores (pXRF) como las concentraciones reales de Li medidas con técnicas de laboratorio más precisas, como ICP-OES, para el entrenamiento inicial.

**Sesgos geológicos (Contexto colombiano)**

• **Diferencias geológicas y climáticas:** El estudio de referencia se realizó en suelos (Holoceno) de terrenos glaciados con clima húmedo y templado (Wisconsin/Michigan). Colombia presenta una amplia variedad de climas, tipos de suelo y rocas. Es crucial comprender cómo los procesos de meteorización y formación de suelos colombianos, influyen en la dispersión y concentración de Li y sus trazadores.

• **Contaminación de la muestra:** En terrenos con topografía empinada, puede ocurrir contaminación de los suelos por elementos solubles o partículas detríticas transportadas ladera abajo (escorrentía), especialmente cerca del contacto con rocas encajables de composición contrastante.

**Limitaciones relacionadas con la prospección**

• **Alcance geográfico limitado:** La metodología es aplicable en suelos cuando se trata de perfiles de poco espesor formados directamente sobre el lecho rocoso. Posiblemente, no es apta para explorar depósitos de sedimentos distales o en perfiles de suelos gruesos (decenas o cientos de metros).

• **Heterogeneidad de las pegmatitas:** Las pegmatitas LCT son cuerpos altamente heterogéneos y con zonación interna discontinua. Esta variabilidad natural puede explicar la gran dispersión de los puntos de datos y afectar la precisión de los modelos.

# Metodología

Para predecir potenciales depósitos de Litio (Li) mediante datos geoquímicos de rocas y sedimentos y algoritmos de Machine Learning (ML) en el contexto geológico de Colombia, se tomó como base el enfoque exitoso demostrado en el estudio de Pereira et al. (2025). A continuación se describe la metodología implementada en el presente estudio.

**Herramientas y librerías usadas**

Para la gestión de datos, análisis exploratorio, modelado de Machine Learning y visualización, se emplearán las siguientes herramientas:

| **Categoría** | **Herramienta o Librería** | **Propósito** |
| --- | --- | --- |
| **Gestión de Datos** | MySQL, Access y Pandas(Python) | Estructuración y gestión de la base de datos geoquímica. |
| **Librerías de ML** | MLP | Modelar relaciones no lineales entre variables geoquímicas y el litio (Li\_ppm). |
| Ridge Regression | Predecir Li\_ppm penalizando coeficientes demasiado grandes. |
| Lasso Regression | Predecir Li\_ppm mientras selecciona las variables más relevantes, ya que puede forzar a cero algunos coeficientes. |
| Gradient Boosting Regressor | Crear un modelo robusto para estimar Li\_ppm con alta precisión. |
| randomForest (R) | Implementación del algoritmo Random Forest para predicción de Li. |
| neuralnet (R) | Implementación de algoritmos de Redes Neuronales, utilizados para predicción de Li. |
| **Análisis Exploratorio** | pandas y numpy | Para estadísticas descriptivas |
| matplotlib.pyplot y seaborn | Para heatmaps, histogramas, scatterplots y barplots de correlación. |
| **Visualización** | Matplotlib, Seaborn, Folium | Gráficos estáticos |
|  | plotly.express y Dash | Dashboards interactivos (scatter, histogramas y mapas de calor con filtros en tiempo real) |

**Pasos metodológicos**

El análisis se divide en cuatro fases, siguiendo el flujo estándar de un proyecto de ciencia de datos:

**1. Importación y estructuración de la base de datos**

• **Adquisición de datos:** Se descargaron los datos geoquímicos disponibles en las bases de datos libres del Servicio Geológico Colombiano (SGC).

• **Estructuración:** Los datos brutos se importaron a una base de datos relacional (Access y MySQL) para asegurar la consistencia y la normalización.

• **Definición de variables:** Se aseguró la presencia de la variable objetiva (Litio (Li) y los 16 elementos predictores cuantificables por pXRF (Cs, U, K/Rb, Zn, As, Th, W, Sc, Y, Tb, Ga, Dy, In, Fe, Ho, Co, Er)

**2. Exploración y limpieza de datos**

• **Preprocesamiento:** Se aplicaron correcciones a los datos brutos, similares a las realizadas en el estudio de referencia mediante el uso de regresiones lineales (ver siguiente capítulo).

• **Cálculo de variables derivadas:** Se calculó la relación K/Rb, una variable crucial que se correlaciona negativamente con el grado de fraccionamiento de las pegmatitas LCT y es un predictor potencial de Li.

• **Análisis exploratorio (EDA):** Se realizaron gráficos de dispersión (por ejemplo, Li vs. Cs, Li vs. K/Rb) para identificar las relaciones geoquímicas clave en el contexto colombiano.

• **Normalización/Estandarización:** Para el modelado con ML, los contenidos elementales fueron estandarizados mediante el método de centrado y escalado, como se sugiere en la metodología de referencia.

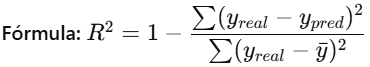
**3. Análisis y modelado predictivo (Machine Learning)**

Se entrenaron cinco algoritmos principales para la predicción de contenido de Litio (MLS, Ridge, Lasso, Gradient Boosting y Random Forest). El objetivo es identificar qué elementos trazadores (incluyendo la razón K/Rb) son los más relevantes para la predicción de Li en el dataset. El conjunto de datos será dividido aleatoriamente en conjuntos de entrenamiento (80%) y de validación (20%).

**4. Evaluación y Selección de Modelos (Criterios de Decisión)**

La selección del modelo más adecuado de ML se basa en la evaluación del rendimiento utilizando métricas específicas para regresión y clasificación. A continuación se hace un breve resumen de las métricas usadas en el presente estudio.

**Coeficiente de Determinación (**R²**):**  Proporción de la variabilidad de la variable objetiva (Li\_ppm) que el modelo logra explicar.



R² = 1: el modelo explica el 100 % de la variabilidad (predicción perfecta).

R² = 0: el modelo no mejora respecto a usar solo la media.

Valores negativos: el modelo es peor que la media.

**Error Cuadrático Medio (RMSE):** Magnitud promedio de los errores de predicción.



Cuanto más bajo, mejor.

Tiene las mismas unidades que la variable objetiva (ppm de litio).

**Desviación de Predicción Residual (RPD):** Relación entre la desviación estándar de los valores reales y el RMSE.



RPD < 1.5 → Modelo poco útil.

1.5–2.0 → Predicciones aproximadas.

> 2.0 → Buena capacidad predictiva.

> 3.0 → Excelente para aplicaciones prácticas.

**Precisión General (*Overall Accuracy*):**

Se busca el modelo que clasifique correctamente el mayor número de puntos de validación.

# Análisis Exploratorio de Datos (EDA)

**• Tratamiento inicial de la base:** Se inició con la selección de solo las muestras que contienen Litio. En la base de datos inicial, se evidenciaron algunos datos con gran desviación, con concentraciones de Li desde un mínimo de -1 ppm, hasta un máximo de 76,8 ppm. El Litio exhibe un amplio rango de concentraciones, lo cual es esencial para entrenar un modelo robusto de regresión. La base de datos geoquímica además contiene valores faltantes (representados como NA o null) que fueron descartados para poder realizar el modelamiento por ML.

El estudio cuenta con diferentes unidades de medición para todos los elementos: porcentaje, ppb, y ppm, por este motivo se unificaron las concentraciones para que todos los datos estén expresados en ppm.

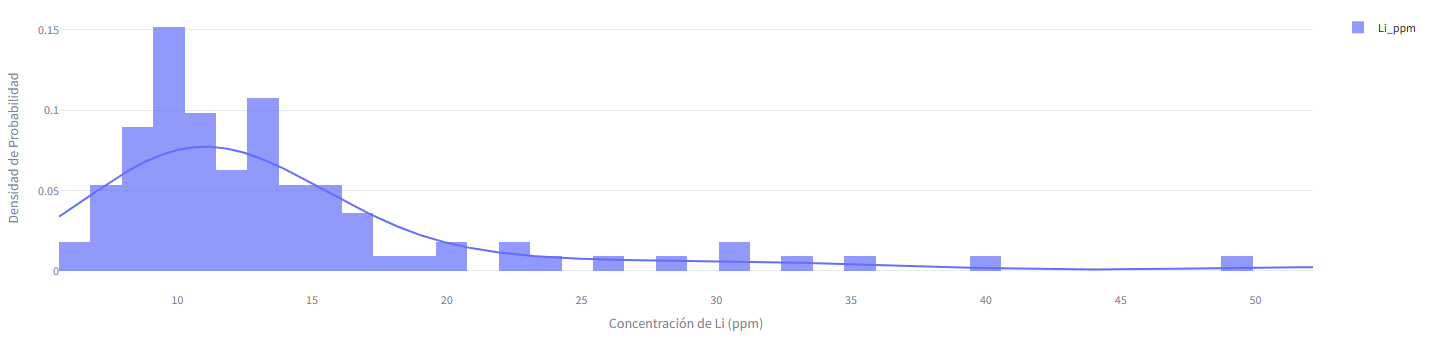
Para mayor precisión de la información y mejor entrenamiento del modelo de ML se seleccionaron solo las muestras de Roca y Sedimentos, con filtro en la técnica de medición Plasma (ICP) y ICP-OES, ya que las concentraciones obtenidas por estos instrumentos son los que presentan la mayor precisión analítica en el laboratorio.

Posterior a esta depuración, quedaron 97 registros que cumplen con todas las condiciones requeridas para correr un modelo de ML.

Adicionalmente, para la ubicación geográfica de las muestras fue necesario realizar una transformación del sistema de coordenadas de metros a grados decimales, ya que la mayoría de estas se encontraban en el primer sistema de referencia y la librería de python Folium no permite su uso.

• **Identificación y tratamiento de outliers (valores atípicos):** Los valores atípicos se analizaron cuidadosamente, ya que en el contexto geoquímico, un *outlier* puede representar una anomalía verdadera (es decir, un depósito de Li de alta concentración), o un error de medición o muestreo.Se utilizaron gráficos de caja (box plots) para visualizar la distribución de cada variable (Li, Rb, K/Rb, etc.) e identificar puntos que se encuentran fuera del rango intercuartílico (IQR) extendido.

La Figura 1 presenta el histograma y boxplot de la concentración de Litio en las 97 muestras después de los filtros mencionados. Los datos varían desde un mínimo de 5,6 ppm hasta un máximo de 52,2 ppm. También se observa que la mayor parte de los datos se agrupan entre aprox. 10 y 20 ppm y existen outliers de hasta 50 ppm.



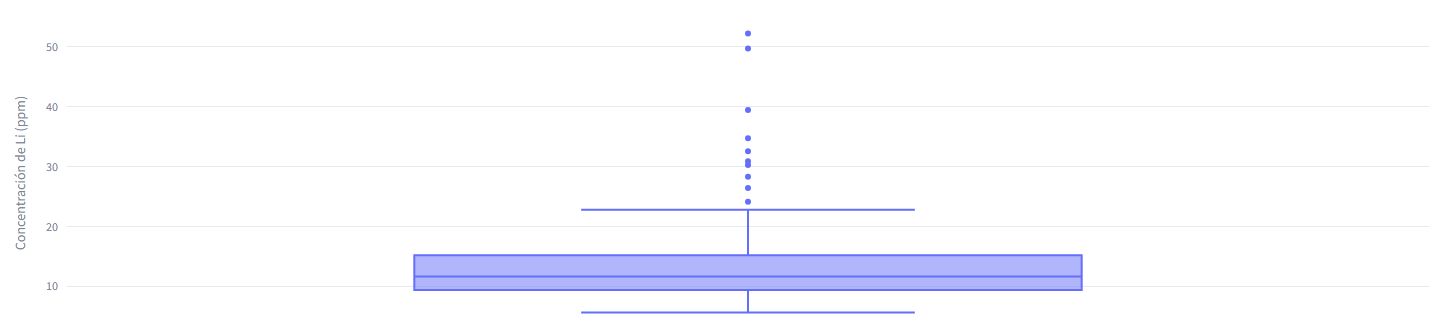


Figura 1. Histograma y boxplot de las concentraciones de Litio en el dataset después de los filtros.

• **Relación y dispersión del Litio vs. Cesio y K/Rb:** La razón K/Rb se considera una variable predictora crucial. En la Figura 2 se observa una fuerte correlación inversa con el contenido de Li, y es la principal herramienta visual para la discriminación: la mayoría de los puntos de datos ricos en Li se concentran en valores bajos de K/Rb.

Esta relación se correlaciona geológicamente con el grado de fraccionamiento de las pegmatitas LCT y es un predictor potencial de la prospectividad de Li

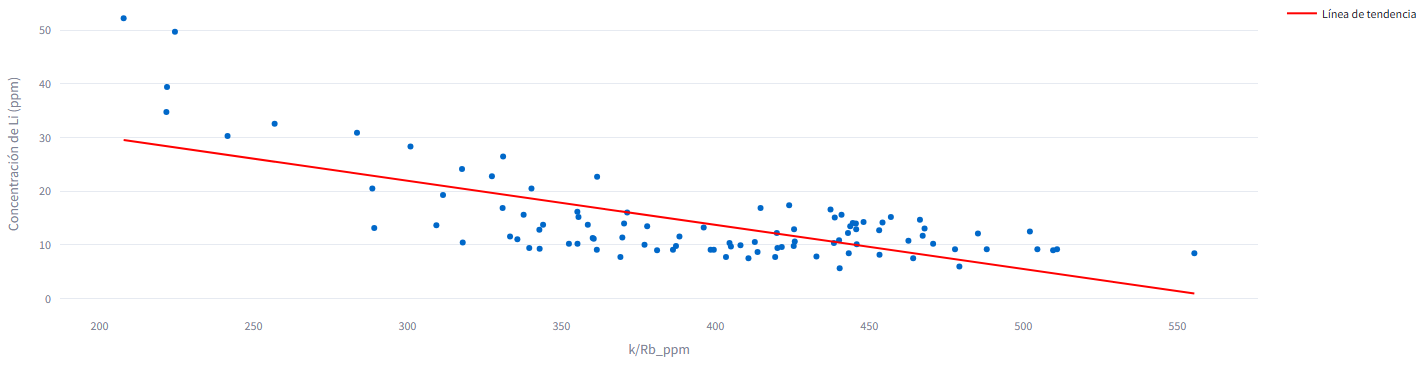


Figura 2. Scatterplot de K/Rb vs Litio en correlación inversa

Por otra parte el Cesio, muestra una relación estrecha con el Litio. Su parentesco químico muestra que durante la evolución de un magma, estos dos elementos se encuentran en las etapas finales que forman las pegmatitas graníticas (Figura 3).

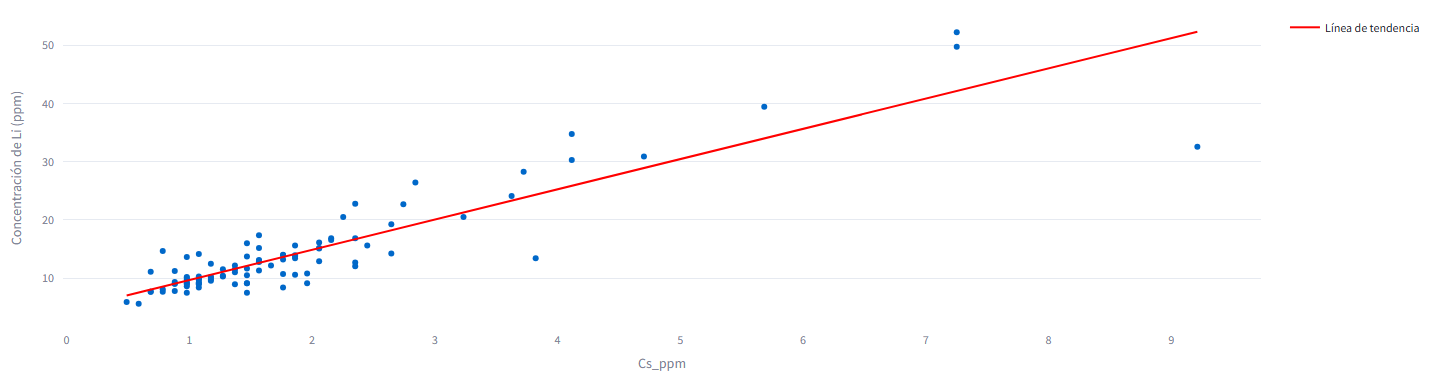


Figura 3. Scatterplot de Cesio vs. Litio en correlación directa

Elementos como el hierro (Figura 4), cobalto, zirconio y arsénico no muestran una relación lineal clara como los elementos anteriores, presentando mayor dispersión.

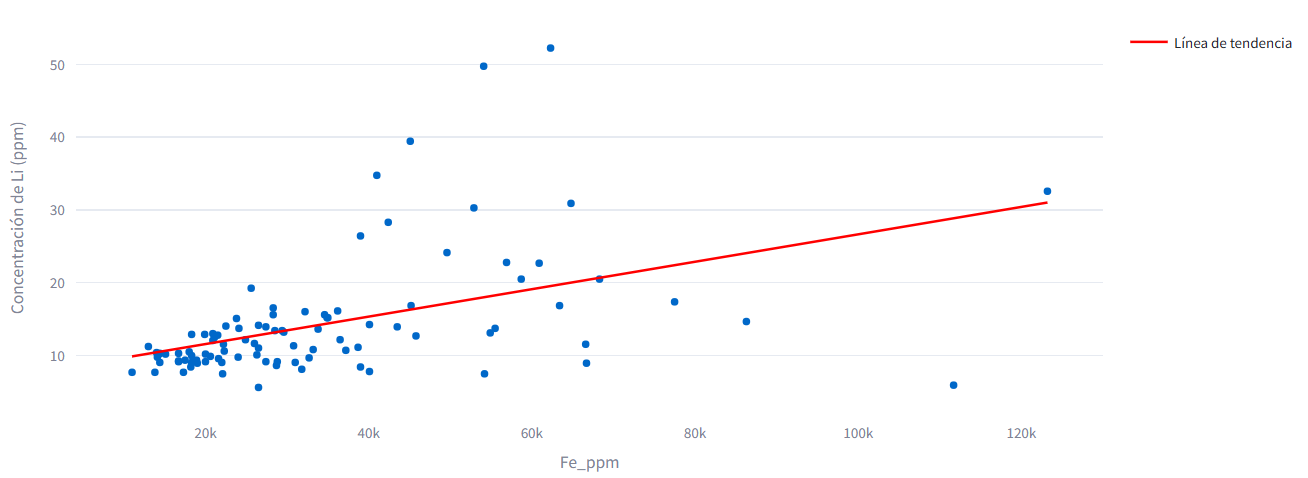


Figura 4. Scatterplot de Hierro vs. Litio, sin relación lineal clara

En la figura 5 se presentan todos los elementos trazadores organizados de acuerdo a su grado de correlación con el Litio, obteniendo el siguiente orden de mayor a menor correlación: Cs, U, Zn, As, Th, Sc, Y, Tb, Ga, Dy, Ln, Fe, Ho, Er, Co, W y K/Rb.

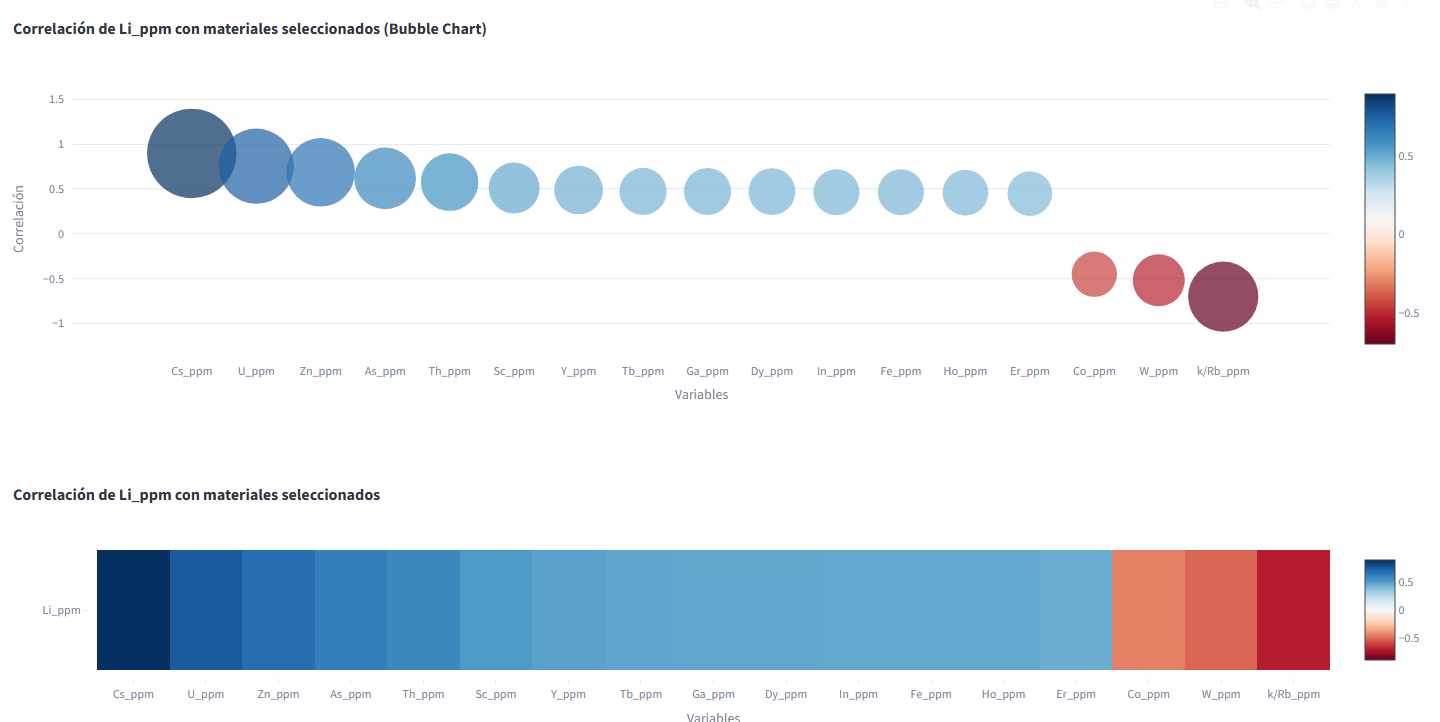


Figura 5. Correlación de los elementos trazadores con el Litio. En azul, elementos con correlación directa, en rojo con correlación inversa.

Estos patrones preliminares validan la hipótesis de que los algoritmos de ML son necesarios para modelar las relaciones multivariadas y no lineales que discriminan eficazmente entre las muestras. El modelo de Machine Learning deberá entrenarse para explotar las correlaciones inversas y positivas.

# Análisis principal

El dataset inicial obtenido del Servicio Geológico Colombiano cuenta con más de 160 mil datos geoquímicos recolectados en diferentes partes de Colombia. Sin embargo, posterior a los múltiples filtros realizados (descritos en la metodología), en la figura 6 se presenta la localización de las muestras que contienen concentraciones de Litio.

En la figura de la izquierda se presenta el Heatmap de los 4 clusters donde Litio ha sido medido en el país. Los resultados se distribuyen en los departamentos de 1) la Guajira, 2)Bolívar y Cesar, 3) Cundinamarca, Antioquia y Chocó, y 4) Huila, Cauca y Putumayo. A la derecha, se presentan las mismas muestras catalogadas por colores, donde amarillo indica bajas concentraciones, azul medias concentraciones y rojo altas concentraciones de Litio.

De acuerdo a nuestros resultados, las concentraciones de las muestras de rocas y sedimentos analizados presentan en su mayoría bajas concentraciones.

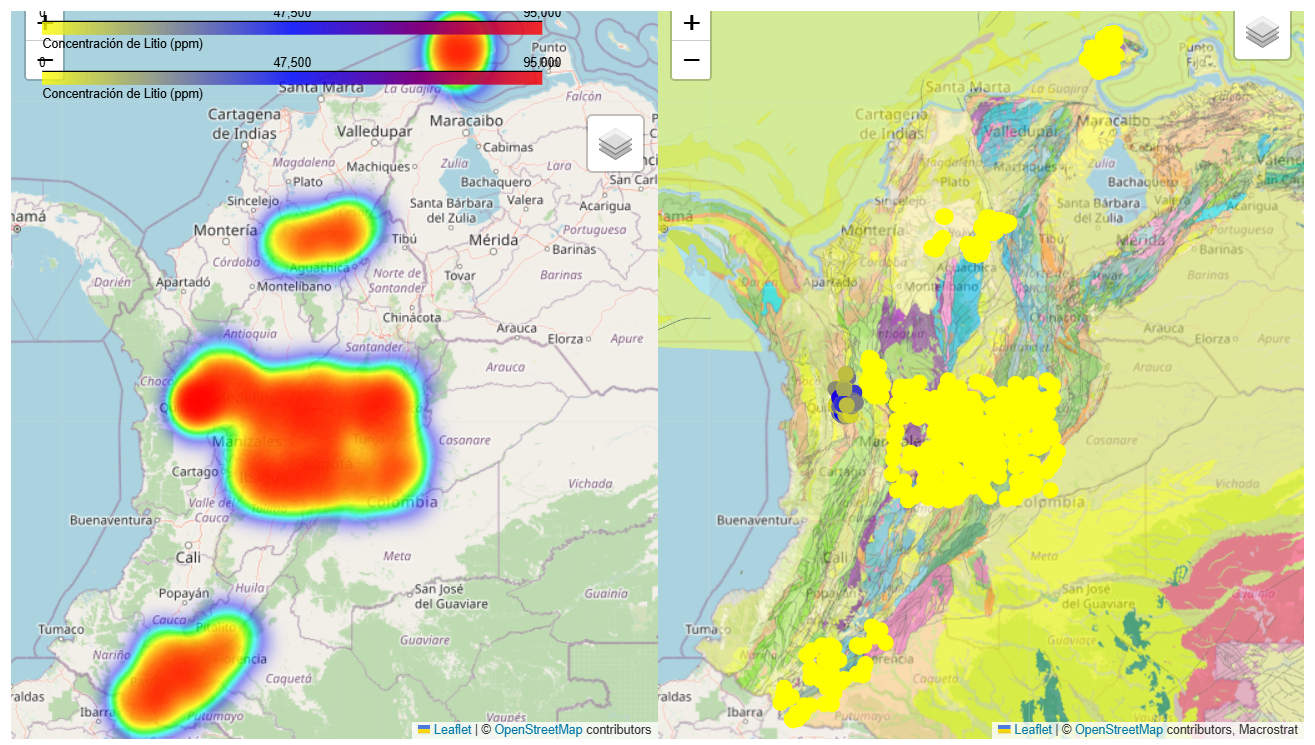
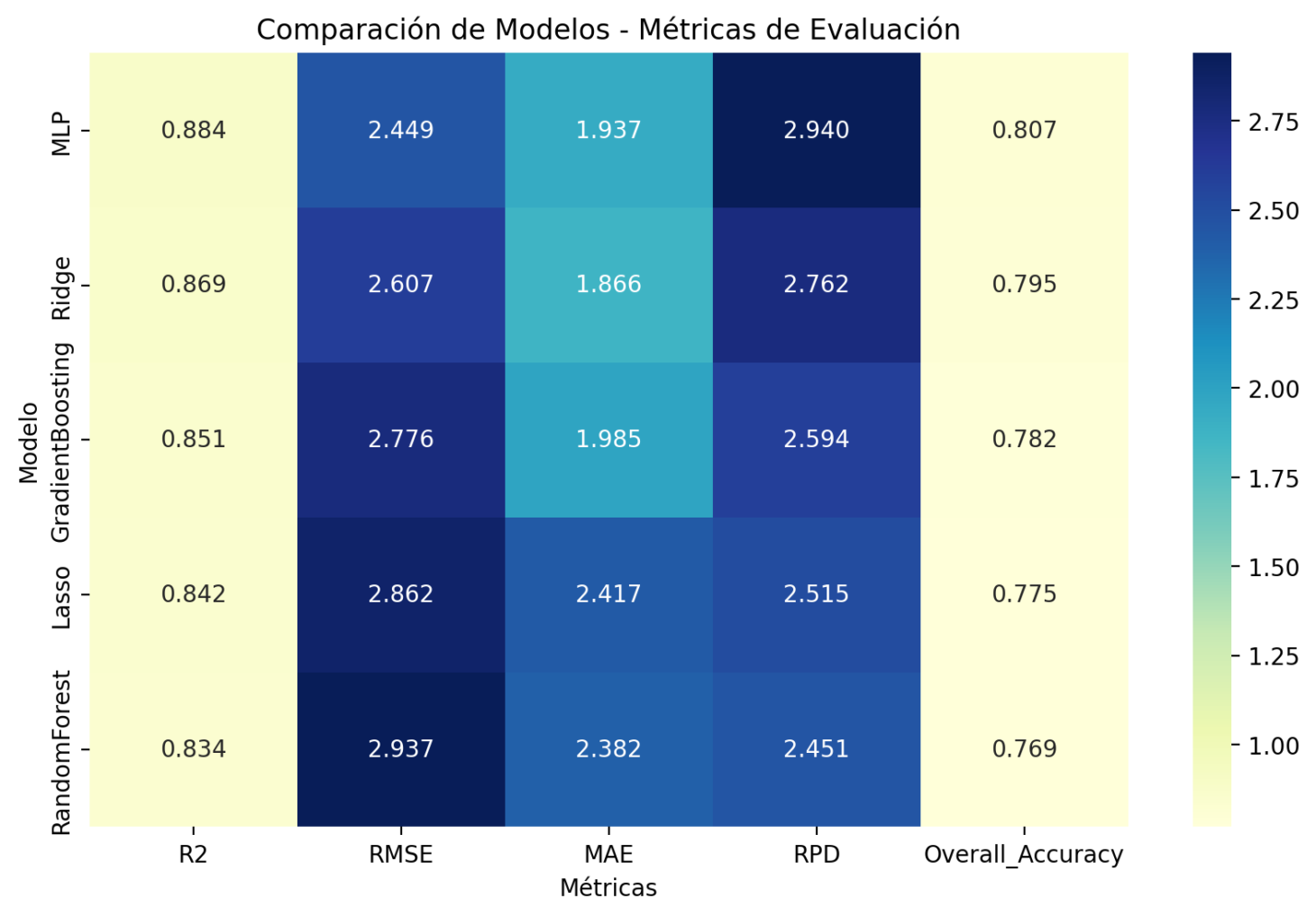


Figura 6. Ubicación en el mapa geográfico y geológico de las muestras que presentan concentraciones medidas de Litio en la base de datos del Servicio Geológico Colombiano.

Este dataset fue filtrado y utilizado para correr los modelos de ML, pues era necesario que cada muestra analizada tuviera mediciones tanto de Litio como de los elementos trazadores. En la figura 7 se presentan las métricas (R², RMSE, MAE, RPD, Overall Accuracy) obtenidas para todos los modelos ML utilizados en el presente estudio:



# Hallazgos clave

A partir de las métricas obtenidas se puede concluir:

**1) Mejor modelo: MLP (Red Neuronal Multicapa)**

R² ≈ **0.89** → explica casi el 90% de la variabilidad en Li.  
RMSE ≈ 2.4 ppm → error bajo en el rango de valores (muy bueno).  
RPD ≈ 2.97 (>2.5) → considerado excelente en predicción de propiedades geoquímicas.  
Accuracy ≈ 0.81 → alto nivel de confiabilidad.

**2) Ridge y Lasso** también muestran buen desempeño (R² entre 0.84–0.86). Indica que el Li tiene relación lineal y aditiva con algunos de los elementos trazadores, lo que explica por qué los modelos lineales funcionan bien.

**3) Gradient Boosting y Random Forest** (R² ≈ 0.83). Muestran resultados sólidos, pero no superan a los lineales ni a la red neuronal. Puede ser que las relaciones no sean altamente no-lineales en este dataset.

**Conclusión técnica**: Los datos son muy predecibles (R² > 0.8 en casi todos los casos). El Li tiene asociaciones geoquímicas consistentes que pueden ser capturadas por modelos tanto lineales como no lineales.

# Conclusiones y recomendaciones

## 1) Significado geológico de los elementos trazadores asociados a Li

Los elementos químicos que más se correlacionan con Li es muy reveladora:

**Metales de alta densidad y alta energía (Th, U, W, Sc, Ga, In, Cs)**: Th y U son indicadores de magmas altamente fraccionados, enriquecidos en elementos incompatibles, comunes en sistemas pegmatíticos y graníticos litíferos. Cs es el clásico “pathfinder” de litio en pegmatitas; asociado a polilitionita y lepidolita. W es indicador de fluidos hidrotermales asociados a magmatismo fraccionado. Sc, Ga, In son elementos de traza que se concentran en ambientes de cristalización tardía; y su asociación sugiere evolución magmática avanzada.

**Elementos de tierras raras (Y, Dy, Tb, Ho, Er):** Estos elementossuelen concentrarse en pegmatitas, granitos peraluminosos, y depósitos hidrotermales, y se relacionan con minerales accesorios (monacita, xenotima, allanita) que pueden coexistir con espodumena, lepidolita o turmalinas litíferas. Su presencia sugiere un entorno fraccionado y enriquecido en elementos incompatibles , típico de magmas evolucionados que hospedan litio.

**Metales de transición (Fe, Co, Zn, As):** Pueden marcar influencia hidrotermal o alteración. Zn y As son asociados a ambientes hidrotermales y mineralizaciones polimetálicas; si se correlacionan con Li, podría indicar redistribución por fluidos tardíos; mientras Co y Fe reflejan condiciones redox y presencia de minerales máficos/alteración hidrotermal.

**Interpretación global:** Los modelos captan la firma geoquímica de magmas altamente evolucionados y fluidos hidrotermales tardíos, los cuales son claves para la concentración de litio. La presencia de Cs + REE + Th + U es un “combo exploratorio” muy fuerte para pegmatitas litíferas. La asociación con Zn, As, Co, Fe puede señalar zonas donde hubo alteración hidrotermal que redistribuye o concentró litio.

## 2) Recomendaciones

* Incluir más muestras, especialmente en sectores con pocos datos, para mejorar la representatividad espacial y estadística del modelo.
* Combinar la geoquímica con mapas litológicos, estructurales y datos geofísicos, para mejorar la interpretación de las anomalías predichas.
* Prestar especial atención a Cs, REE (Y, Dy, Tb, Ho, Er), Th, U y W, que aparecen como los mejores predictores de Li, tanto en el modelo como en el marco geológico.
* Exportar los resultados predichos a un GIS y aplicar interpolación geoestadística (kriging, co-kriging con Cs u otros REE) para generar mapas continuos de anomalías.
* Validar las predicciones con muestreos adicionales o compararlas con depósitos conocidos para comprobar la capacidad del modelo de identificar zonas prospectivas.
* Usar las predicciones de Li no como valor absoluto, sino como zonas de anomalía relativa para priorizar campañas de muestreo, perforación o análisis mineralógico detallado.

**3) Limitaciones**

* El modelo sólo es confiable dentro del rango de composiciones químicas presentes en el dataset. No puede extrapolar a contextos geológicos diferentes donde no se entrenó.
* Aunque ciertos elementos correlacionan bien, el litio también depende de factores mineralógicos y texturales (espodumena, lepidolita, arcillas litíferas) que no se capturan solo con geoquímica total.
* El modelo se basa únicamente en geoquímica, sin integrar mineralogía, geología de campo o controles estructurales, lo cual puede limitar la interpretación geológica real.
* El mapa basado en muestras puntuales no refleja continuidad total; requiere interpolación y puede inducir falsas anomalías si la densidad de muestreo es baja.
* Aunque Random Forest y MLP son robustos, valores atípicos en concentraciones de elementos traza pueden influir en la importancia de variables y predicciones.
* En modelos como MLP (redes neuronales) la interpretación es menos intuitiva que en regresiones lineales, lo que dificulta justificar geológicamente algunas relaciones.

**4) Posibles investigaciones futuras**

* Hacer mapas de anomalías combinadas de Cs + REE + Th/U + W con el fin de priorizar sectores de exploración.
* Hacer un mapa de zonas con anomalías de Zn, As, Co asociadas al litio predicho por ML con el fin de identificar flujo de fluidos tardíos (lugares de removilización y enriquecimiento secundario).

**Referencias bibliográficas**

Luiza Maria Pereira Pierangeli, Mona-Liza C. Sirbescu, Sérgio Henrique Godinho Silva, David C. Weindorf, Thomas R. Benson, Nilton Curi; Soil Geochemistry Toward Lithium Pegmatite Exploration: Building a Machine-Learning Predictive Algorithm via Portable X-Ray Fluorescence. *Economic Geology* 2025; doi: <https://doi.org/10.5382/econgeo.5166>