

Predicción del flujo de calor mediante aprendizaje automático. Caso de estudio: Sierra Centro-Norte del Ecuador

Heat flow prediction by Machine Learning. Case of study: Central and north mountain region of the Ecuador

Maritza BarreraRamírez

Universidad Central del Ecuador, Quito, Ecuador

mebarrera@uce.edu.ec

<https://orcid.org/0000-0002-2401-6838>

Esteban Armas Cepeda

Universidad Central del Ecuador, Quito, Ecuador

esarmas@uce.edu.ec

<https://orcid.org/0000-0001-6657-0223>

Santiago ÁlvarezRuales

Universidad Central del Ecuador, Quito, Ecuador

sbalvarez@uce.edu.ec

<https://orcid.org/0000-0002-7768-2690>

Christian Mejía Escobar

Universidad Central del Ecuador, Quito, Ecuador

cimejia@uce.edu.ec

<https://orcid.org/0000-0001-6715-191X>

Cita del artículo: Barrera-Ramírez, M., Armas-Cepeda, E. y Álvarez-Ruales, S. (2020). Heat flow prediction by Machine Learning. Case of study: Central and north mountain region of the Ecuador.



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Resumen

Los procesos de enseñanza y aprendizaje de la educación superior deben generar cambios que impacten positivamente al proceso de aprendizaje de los estudiantes, para esto se proponen nuevas metodologías de aprendizaje y manejo de datos como el Machine Learning (ML). El conocer nuevas tecnologías para la predicción de datos de cualquier tipo de variables, sirve tanto en el ámbito geológico como en cualquiera otra área del conocimiento, siempre que se tome en cuenta una variable dependiente la cual se desee predecir y una serie de variables independientes con las que se entrena al algoritmo. En este artículo se realizó la predicción de datos de flujo de calor geotérmico, presentados mediante mapas construidos con datos de predicción, esto permite conocer la variación de estos datos con respecto a los datos reales. La información con la que se elaboraron los mapas fue obtenida de manera predictiva, mediante el empleo de algoritmos de aprendizaje automatizado como son las redes neuronales y los árboles de regresión. Estos algoritmos se emplearon en el software R Studio, con la finalidad de dar a conocer la utilidad de estos para la realización de predicciones de datos, sobre todo cuando no se cuenta con una base de datos extensa, como es el caso de los datos para el flujo de calor en el Ecuador, que por diversas razones son difíciles de conseguir para determinadas áreas del país o simplemente no existen. La zona de estudio se encuentra en la Sierra Centro-Norte, en las provincias de Cotopaxi y Pichincha.

Palabras clave

Aprendizaje automático, flujo de calor, redes neuronales, árboles de regresión, predicción.

Abstract

The teaching and learning processes of higher education must generate changes that positively impact the learning process of students, for this, new learning methodologies and data management such as Machine Learning (ML) are proposed. Knowing new technologies for the prediction of data of any type of variables, serves both in the geological field and in any other area of knowledge, as long as a dependent variable is taken into account which is to be predicted and a series of independent variables with those that the algorithm is trained. In this article, the prediction of geothermal heat flow data is performed, presented through maps built with prediction data, this allows to know the variation of these data with respect to the real data. The information with which the maps were prepared was obtained predictively, using automated learning algorithms such as neural networks and regression trees. These algorithms were used in the R Studio software, in order to publicize their usefulness for making data predictions, especially when there is not an extensive database, as is the case of data for the flow of heat in Ecuador, which for various reasons are difficult to achieve for certain areas of the country or simply do not exist. The study area is in the Sierra Centro-Norte, in the provinces of Cotopaxi and Pichincha.

Keywords

Machine learning, heat flow, neural networks, regression trees, prediction.

1. Introducción

En la actualidad vivimos en un mundo digital, y la Inteligencia Artificial (IA) juega un aspecto importante en nuestras vidas, debido a que muchas aplicaciones y máquinas nos ayudan a completar nuestras tareas diarias. Las nuevas tecnologías de aprendizaje como el Machine Learning (ML) o Aprendizaje Automático, parte de la IA, se ha introducido para el



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

manejo de grandes cantidades de datos. El propósito es que la computadora aprenda en base a un modelo que consta de procesos de prueba y error, con el fin de llegar a ajustarlo a un modelo óptimo. La implementación de estas nuevas tecnologías en el ámbito laboral hace que los procesos de enseñanza y aprendizaje de la educación superior sufran cambios que impactan positivamente al proceso de aprendizaje de los estudiantes, por lo que es importante conocer sobre las nuevas metodologías de aprendizaje mediante el empleo de las tecnologías de ML. La técnica de Machine Learning, se ha empleado con distintas finalidades, nuestro enfoque es para el área geológica, con el fin de dar a comprender a la comunidad, el uso de los algoritmos que en un principio son complejos pero empleados correctamente permiten solucionar una serie de problemas. El único requisito que debe considerar es la recolección de un conjunto de datos extensos y confiables, en este caso de estudio son los datos de las características geológicas de la región estudiada para la predicción del flujo de calor.

Tomando en cuenta el empleo del ML, el presente artículo tiene como propósito presentar y explicar la predicción de flujo de calor mediante la implementación de algoritmos de ML en lenguaje de programación R con soporte del software RStudio. De esta manera, dar a conocer a estudiantes y profesionales de varias carreras, como es el comportamiento del flujo de calor en la zona centro-norte de la sierra ecuatoriana, en un área comprendida entre las provincias de Pichincha y Cotopaxi (Figura 1.), a partir de valores de diferentes anomalías y parámetros, los cuales se pueden evidenciar de igual manera en el mapa de ubicación, para la predicción del flujo de calor, mediante la aplicación de los algoritmos en un software especializado. Como resultado se presentará mapas elaborados a partir de las mediciones reales de flujo de calor y los datos obtenidos de la predicción. La importancia que puede tener la predicción de datos con las técnicas de ML, mediante el empleo de algoritmos como en este caso son las redes neuronales y los árboles de regresión. No se pretende explicar la función que cumple el flujo de calor en determinados procesos geológicos, pero sí la importancia de éste en los glaciares, así como la relación que se puede tener entre datos reales y datos pronosticados por medio de algoritmos en un lenguaje computacional.

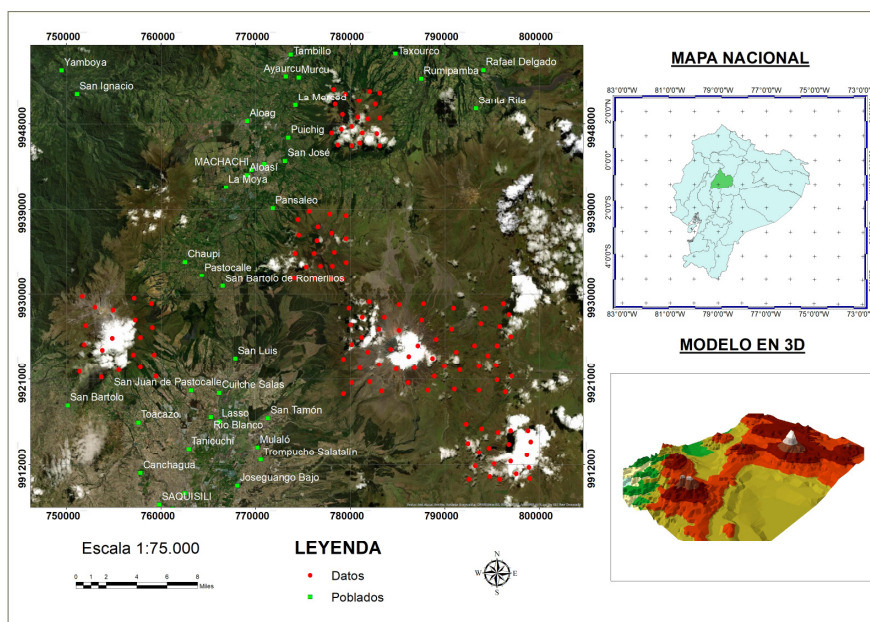


Figura 1. Mapa de ubicación y puntos donde se tomaron los datos



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

La principal función de aplicar herramientas computacionales y las técnicas de ML en la predicción del flujo geotérmico de una zona de estudio, en este caso el de la Sierra centro-norte entre las provincias de Pichincha y Cotopaxi, es la obtención de valores estimados, como ya se ha mencionado, los algoritmos seleccionados para dicha predicción son las Redes Neuronales y los Árboles de Regresión, con esto se pretenderealizar mapas que permitan comprender cuál es la relación de los datos estimados con el ML y los datos reales. La presentación de los mapas de flujos de calor para datos reales y estimados contiene los diversos parámetros como son la recolección de un conjunto de datos confiables acerca de las características geológicas de la zona de estudio con los datos de flujos de calor, medidos u obtenidos en este lugar. Así se implementará un modelo de aprendizaje automático para el tratamiento de los datos obtenidos anteriormente para la predicción del flujo de calor, con la ayuda del software especializado RStudio, y como último punto, la elaboración de los mapas de predicción del flujo de calor tanto para los datos reales y los datos estimados, con el fin de conocer la variación existente entre ambos tipos de datos.

Al obtener estos valores calculados de forma predictiva, es de interés no sólo desde el punto de vista de las matemáticas no convencionales, sino también se le agrega un ambiente de importancia ambiental y de ordenamiento territorial, debido a que los diferentes complejos volcánicos presentes en la Cordillera de los Andes son los responsables de que existan anomalías en el flujo de calor presente en la corteza terrestre. Con las herramientas computaciones es posible calcular el flujo de calor de una manera en que seestimen dichos valores, además de características geológicas que se presentan en la zona. Podrían ser de utilidad para conocer su relación con los deshielos de los glaciares de la cordillera o si están vinculados con el cambio climático, incluso cuál es la relación que tienen con estos factores y si en caso de elevarse pueden producir que todo el agua acumulada de forma sólida, se descongele y cambie o destruya la condiciones ambientales del lugar debido a que ya no sería el mismo entorno natural, producto de este deshielo ya no se tenga fuentes de agua como la del glaciar del volcán Cotopaxi, la cual es útil para el abastecimiento de agua a diferentes poblados como Quito, Machachi, Latacunga, Sangolquíetc., y para el ámbito económico, y ambiental ya que toda el agua que nutre a los páramos ya no existiría ycausaría impactos medio-ambientales.

Debido a que en el país los datos de flujo de calor no se encuentrandisponibles, o son restringidos, se puede realizar una predicción con la ayuda de herramientas computacionales, comoel software de programación RStudio, con que el que se pretendegenerar mapas de flujo de calor en base a predicción de una cantidad pequeña de datos.

2. Revisión de literatura

El calor que emite la Tierra se cuantifica mediante el flujo de calor superficial, que es la cantidad de calor que pierde el planeta por unidad de superficie y tiempo. El flujo de calor superficial se calcula como el producto del gradiente geotérmico y la conductividad térmica del medio. (Pollack et al., 1993).

$$Q = -k \left(\frac{\Delta T}{\Delta Z} \right) \quad (\text{Ecuación 1})$$

Donde, Q es el flujo de calor, k es la conductividad térmica y $\left(\frac{\Delta T}{\Delta Z} \right)$ es el gradiente geotérmico.



Pollack también propone que los flujos de calor medios de los continentes y océanos son 65 y 101 mW /m², respectivamente, que cuando se ponderan en términos reales producen una media global de 87 mW /m²; adicionalmente da a conocer que la velocidad a la que el espectro del flujo de calor pierde o gana fuerza está estrechamente relacionado con la topografía de la Tierra y que los valores de los campos gravitacional y magnético caen mucho más abruptamente, en consonancia con las fuentes de campo en el manto inferior y el núcleo, respectivamente.

Algunas investigaciones han desarrollado significativas contribuciones a la literatura utilizando minería de datos en la predicción de variables discretas. Por ejemplo, Rezvanbehbahani et al, (2017) usaron algoritmos de árboles de regresión potenciados por gradiente para calcular datos del flujo de calor en Groenlandia y predecir la pérdida de hielo de los glaciares. Entre otras aplicaciones a la Geología, Hulbert et al, (2018) propone una metodología donde plantea que un terremoto puede estar precedido por un conjunto de procesos organizados que podrían potencialmente ser predecibles con el uso de ML. Existen estudios en otras ramas de la ciencia relacionados a flujos como el estudio de Castillo en 2018 sobre redes neuronales para predecir coeficientes de pérdidas primarias en un régimen de flujo turbulento.

3. Metodología

Esta investigación inicialmente se realizó en 4 fases que se muestran esquemáticamente en la Figura 2.

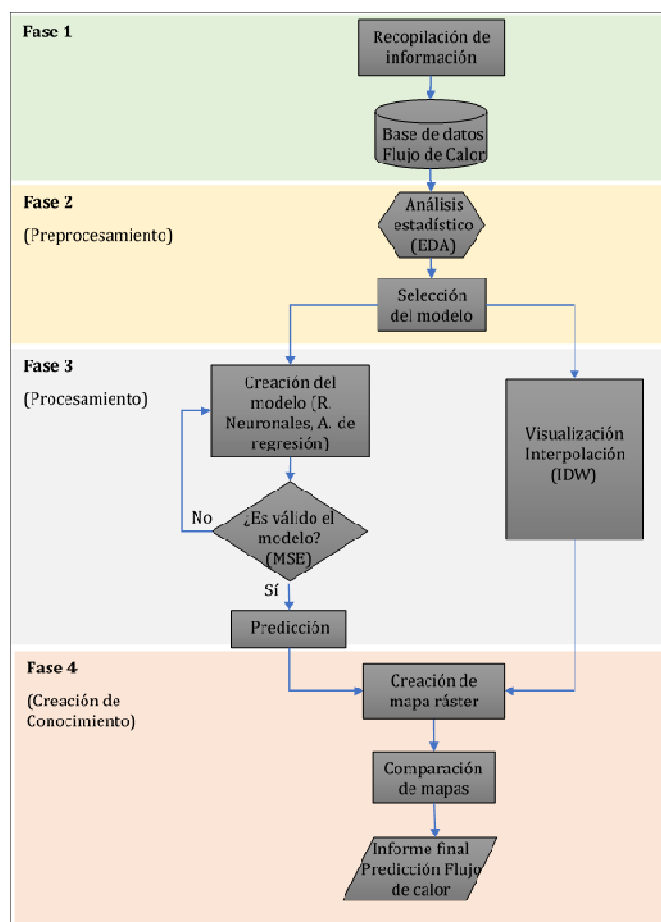


Figura 2: Diagrama de flujo para la metodología propuesta en este estudio. Simbología American National Standard Institute ANSI.

La primera fase fue la recopilación de datos cuantitativos y cualitativos, tanto de la variable dependiente (flujo de calor) como sus correspondientes covariables (Cuadro 1). Los datos fueron obtenidos de los diferentes mapas existentes en varios servidores a nivel mundial publicados por organizaciones como: el NOAA (por sus siglas en inglés de: National Oceanic and Atmospheric Administration), la CGMW (Commission for the Geological Map of the World) en actividades conjuntas con la NASA, etc. En base a esta información se obtuvieron 130 puntos aleatorios con la información de cada variable en la zona de estudio, obteniendo así una base de datos consolidada.

A continuación, se muestran los datos considerados para la predicción con sus respectivos parámetros principales

Nombre variable	Descripción	Unidades	Área de Conocimiento
Flujo de Calor	Cantidad de energía emitida por la tierra	Vatio sobre metro al cuadrado (W/m^2)	Geotérmica
Elevación	Topografía de la zona de estudio	Metros sobre el nivel del mar (m.s.n.m)	Geomorfología
Geología	Tipos de Rocas	-	Geológica
Distancia al Basamento	Con respecto al basamento del volcán	Metros	Geomorfología
Conductividad Térmica	Capacidad de conducir calor a través de un medio	Vatio sobre metro por k ($W/m \cdot k$)	Geotérmica
Gradiente Geotérmico	Variación de la temperatura en función de la profundidad	Metro por grado centígrado ($^{\circ}C/km$)	Geotérmica
Magnetometría	Susceptibilidad magnética del terreno	NanoTeslas (nT)	Geofísica
Gravimetría	Variación gravedad del terreno, medida en (mGal)	MiliGales (mGal)	Geofísica
Latitud	Posición en el espacio	Metros	Geográfica
Longitud	Posición en el espacio	Metros	Geográfica

Cuadro N.º 1: Parámetros variables

En el cuadro 2. se muestran el total de 130 datos recopilados bibliográficamente para las 10 variables ya mencionada anteriormente. (Ver [Cuadro 2](#))

Tanto el preprocesamiento y procesamiento de datos se realizaron en el programa R Studio, que es un software especializado en el tratamiento de datos estadísticos y gráficas, inicialmente el programa, fue desarrollado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland. Este software a través de su gran variedad de bibliotecas disponibles permite aplicar algoritmos tanto para: redes neuronales con la librería Neutalnet y árboles de regresión con la librería Rpart y Rpart.plot.



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

La segunda fase fue el preprocesamiento de datos, donde se realizó un análisis estadístico de datos exploratorios con el fin de conocer aquellas variables con mejor distribución, para considerarlas en la posterior modelización y en la selección de los algoritmos a utilizar. Se eligieron para la predicción los algoritmos de redes neuronales y árboles de regresión. También se realizó la normalización de los datos para trabajar con algoritmos de redes neuronales. El proceso de normalización es empleado debido a que este algoritmo trabaja con una base de datos de números entre uno y cero, lo que hace que calcule mejor el error con respecto a trabajar con todos los datos.

La fase tres se basó en la implementación de técnicas de ML, que se usaron para el procesamiento de datos y la creación, predicción y validación de los diferentes modelos. Durante la creación del modelo se implementó el uso de las diferentes librerías con las que trabaja el software, donde vienen ya estructurados los argumentos de los algoritmos. Se utilizó la librería “Neuralnet” para Redes Neuronales y “Rpart” para árboles de decisión. Para obtener más información sobre el código de programación de estos algoritmos descargue ([Código de programación](#)). Posteriormente se ejecutó la predicción de los datos obtenidos con el modelo y se realizó la validación de éste, mediante el uso del método estadístico del *error cuadrático medio* (MSE) para minimizar el error. Paralelo a este proceso se graficaron los datos de Flujo de Calor de la base de datos inicial con el método de interpolación del *inverso a la distancia* (IDW) para comparar la calidad de estos datos con aquellos obtenidos de los modelos predictivos.

Finalmente, la fase cuatro fue la generación del conocimiento, donde a través de los resultados obtenidos como mapas ráster, se realizó un análisis comparativo sobre la precisión y calidad de los datos obtenido tanto de métodos convencionales como aquellos obtenidos con aprendizaje automático. También se comparó aquellos datos obtenidos con formulaciones matemáticas y aquellos productos de la predicción.



4. Resultados

4.1 Redes Neuronales

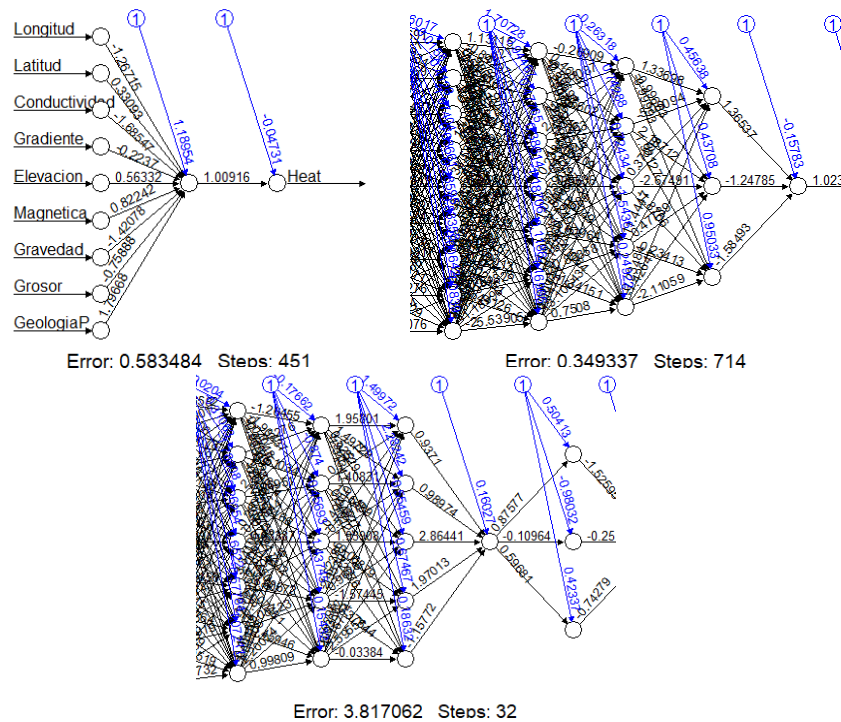


Figura N° 3: Resultados de error obtenidos para los gráficos de redes neuronales.

El software en el que trabajamos como se observa en la Figura N°3, nos arroja las gráficas para cada uno de los modelos, según el número de neuronas, con sus respectivos errores y pasos que ejecuto según los criterios de entrada. Se obtiene que para el de una neurona hay un error de 0.583484 y un número de 451 pasos, para el modelo de 5 neuronas, obtenemos un valor de error de 0.349337 con 714 pasos, mientras que para un último modelo de 7 neuronas obtenemos valores de 3.817062 para el error y de 32 para los pasos.

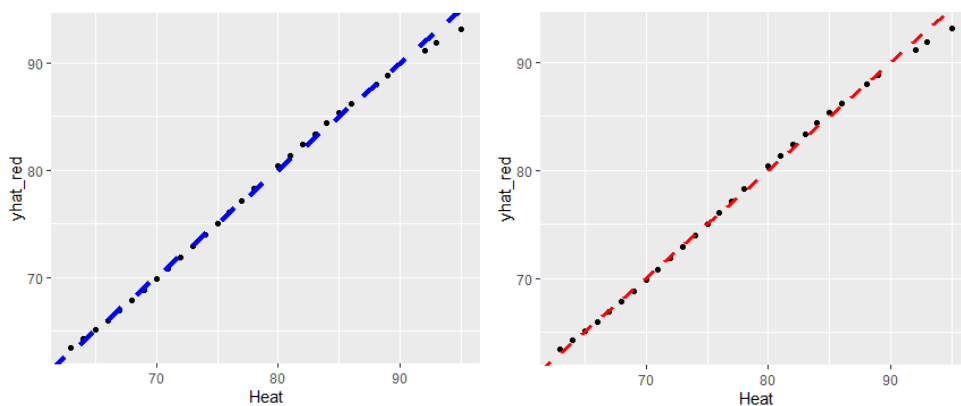


Figura 4. Comprobación lineal de los modelos de 1 y 5 neuronas para redes neuronales



El software RStudio, permite evidenciar el grado de error a través de una gráfica lineal, con la cual el objetivo o el menor grado de error es cuando todos los puntos o la mayoría de ellos caen sobre la recta, en el caso de estudio se realizó como muestra la Figura 4 para las redes neuronales de 1 y 5 capas, respectivamente, como se muestra en la gráfica tienen un grado de error muy bajo.

Los resultados para el error medio cuadrático de los modelos de redes neuronales son de 10% para el modelo de una sola neurona y de 9.25% para el modelo de 5 neuronas.

4.2 Árboles de Regresión

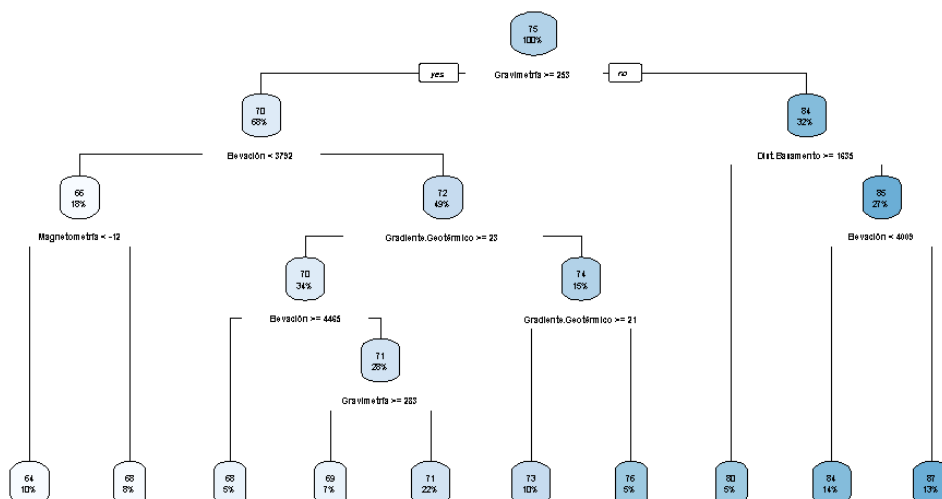


Figura 5. Modelo obtenido para árboles de regresión

La Figura5 nos arroja como resultados un árbol con un total de 5 ramificaciones, en las cuales en sus hojas podemos obtener el valor de error que fluctúa entre 5-22%, según se hayan tomado las consideraciones por el algoritmo para tomar cada una de las variables e ir interactuando entre sí. Se tiene que para el valor del 5% de error, esta ramificación ha interactuado de manera constante las variables del gradiente geotérmico y la elevación, mientras para el error del 22% se aumenta una nueva variable a las dos mencionadas la cual es la gravimetría.

El valor del error cuadrático medio para este modelo es del 6.6%.

4.3 Datos Estimados

Los datos calculados mediante el software RStudio a partir de los algoritmos se muestran en el Cuadro N°3 (ver [Cuadro 3](#)).

En base a los modelos predictivos del flujo de calor por medio de los algoritmos de las redes neuronales y de los árboles de regresión, se estimaron los siguientes datos Cuadro N°3: Datos de Flujo de Calor obtenidos con R Studio para redes neuronales y árboles de regresión, los cuales pueden ser comparados entre sí y ver la precisión que va a tener el uno u otro modelo.

Con los datos tanto los reales como los predictivos se procedió a procesarlos en el software QGIS para la realización de mapas, y poder realizar una comparación visual entre



4.3.2 Redes neuronales

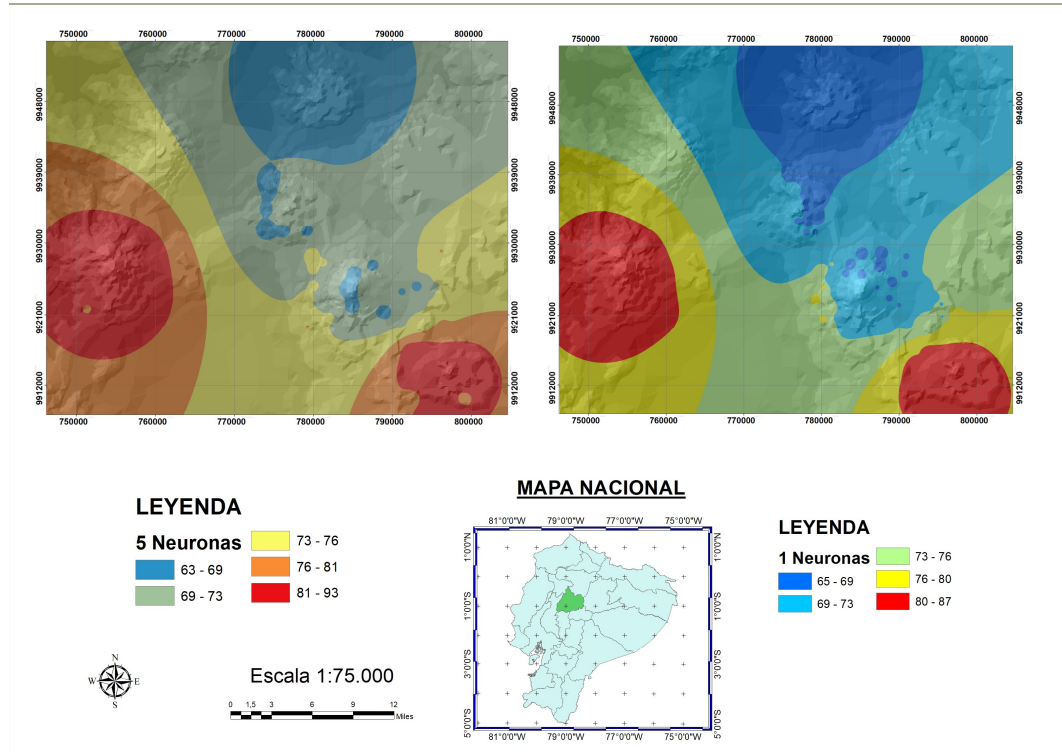


Figura 7. Mapa de flujo de calor con datos calculados mediante redes neuronales, para 5 neuronas y 1 neurona respectivamente

Redes Neuronales	Error en el Modelo	EMC	Pasos
5 neuronas	0.349337	9.25%	714
1 neurona	0.583484	10.45%	451

Cuadro N.º 4: Valores de error del modelo y error medio cuadrático para los modelos de 5 neuronas y 1 neurona respectivamente

Para los valores estimados mediante redes neuronales, se realizaron dos mapas ya que se tenía tanto como para una sola red neuronal como para 5 neuronas. El análisis de error medio cuadrático para cada modelo (Cuadro 4). El primero de estos mapas de la Figura 7, ubicado en la zona derecha de la figura, realizado para una sola neurona presenta un rango de 65-87 W/m² en los valores de flujo de calor, siendo representadas con una tonalidad azul los valores con menor intensidad, con valores cercanos a la base siendo la zona NE la de mayor densidad de estos datos, mientras los valores con mayor intensidad con un color rojo y valores cercanos al 87 W/m², la mayor cantidad de estos se encuentran en la zona SE y SW. Para el segundo mapa de los nombrados Figura 7 y que se encuentra al lado izquierda de ésta, se emplearon los datos obtenidos para las 5 neuronas, con un rango de valores que van de 63 -93 W/m² y teniendo valores de error y EMC de 0.349337 y 9.25% respectivamente, siendo clasificados igual en 5 clases y con los mismos criterios se obtuvo una distribución similar de los valores del flujo de calor del mapa de una sola neurona.



4.3.3 Árboles de regresión

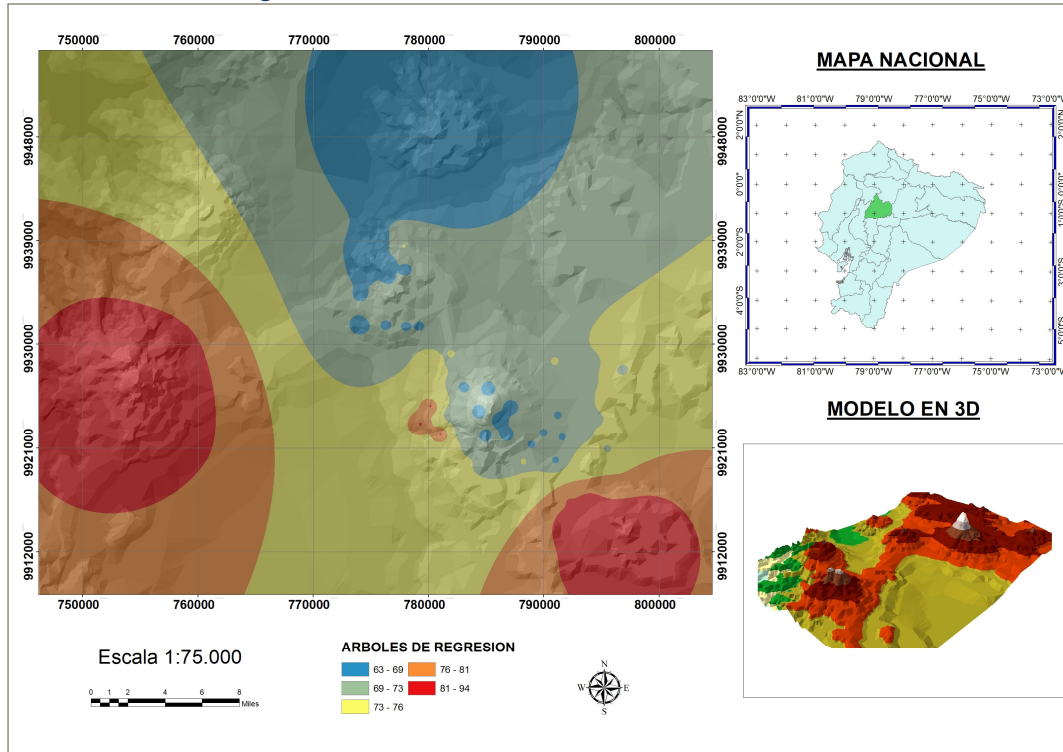


Figura8. Mapa de flujo de calor con datos calculados mediante árboles de regresión

Para los datos obtenidos por los árboles de regresión, se realizó el mapa de la Figura 8 donde los valores fluctuaban entre 63-94 W/m², de los cuales se utilizó el mismo criterio de clasificación que para los anteriores donde se obtiene que las áreas, donde se encuentran los valores más bajos es igual en la parte N y NE, y los valores con una mayor intensidad en la zona SE y SW. Para la comprobación del modelo se obtuvo un EMC de 6.61%

Se plantearon puntos al azar para ver la variación de los valores en cualquier punto del mapa donde se obtuvieron los siguientes resultados Cuadro 5: con la información de los datos comparativos (ver [Cuadro 5](#)).

5. Discusión

5.1 Correlación de los modelos de redes neuronales y árboles de regresión

Como se observa en la Figura 3 y la Figura 5 tenemos que los valores de error más pequeños son para el modelo de redes neuronales con 5 capas en donde el modelo nos arroja un error de apenas el 0.343397, mientras que este valor aumenta con respecto al de una sola capa y al de 7 neuronas, mientras que para los árboles de regresión se obtiene valores de error que fluctúan entre 5-22%. Con lo que se puede comprobar que para una mayor precisión se debe ajustar en el número de neuronas con el número de datos que ingresamos, mientras que los árboles por la cantidad de datos no resultan ser confiables y da un gran margen de error dependiendo las posibilidades de variables que actúen entre sí hasta llegar a las hojas.



Al comparar los valores del EMC obtenemos que el modelo que mejor se ajusta es el de los árboles de regresión, el cual tiene un valor de 6%. Mientras que el modelo que menor error genera en redes neuronales es el de 5 neuronas con un error de 9.25%.

5.2 Correlación de datos reales y estimados

En la Cuadro N°3 donde se observan los datos de los valores obtenidos por RStudio mediante los algoritmos de redes neuronales y árboles de regresión, se tiene planteado que para el primer algoritmo se han desarrollado 1 y 5 neuronas, siendo esta última la que presenta un error mucho menor y por ende una mayor aproximación (error aproximado al 10%). Al comparar los datos reales de flujo de calor con los datos calculados por medio del algoritmo de los árboles de regresión, se tiene que no es muy confiable, aunque el error se mantiene dentro de los límites ya que es de un 6.61%, no tienen mucha correlación y tiende a repetirse mucho los datos, debido a que para este último modelo se requiere de bases extensas de datos, y en el caso de estudio no fue así por eso el programa entra en un bucle y comienza a repetir valores.

Al trabajar con algoritmos de predicción se identificó una similitud con los datos tomados de servidores mundiales, los cuales calculan el flujo de calor a través de formulaciones matemáticas, y aquellos datos obtenidos con Machine Learning, ya que al ejecutar el algoritmo de árboles de predicción por defecto selecciona las variables con más peso para la predicción, en el caso de estudio de las 9 variables independientes seleccionó 4 con mayor peso: gradiente geotérmico, conductividad térmica, elevación y profundidades al basamento. Estas 4 variables son las mismas que intervienen en el cálculo con formulaciones matemáticas como se observa en la ecuación 1. Esto podría indicar que el algoritmo puede llegar a comprender la naturaleza del fenómeno solamente a partir de datos reales.

5.3 Comparación de Mapas

Como se puede observar en los resultados, todos los mapas presentan mínimas variaciones. La mayoría de estas variaciones son con respecto a los límites, ya que aquí se puede ver es donde discreparían los valores, teniendo en cuenta la interfaz gráfica o visual de los mapas, en todas las zonas con mayores valores de intensidad se encuentran ubicados en la zona S, tanto al SE donde están asociados al complejo volcánico Chalupas, y en el lado SW de los Ilinizas. Mientras que, para los colores de menor intensidad, están ubicados en la zona norte. Se puede observar que existe una ligera variación en la delimitación de las áreas para cada rango. En el área de estudio se puede decir que se tiene un valor mayor de flujo de calor en estos sitios debido a que en el caso del Ilinizas se encuentra más cercano a la zona de subducción, mientras que para el caso del complejo volcánico Chalupas juegan un papel importante la actividad volcánica, que aunque no está activa afecta, conjuntamente con el grosor de la corteza, la cual en este lugar hace que esté más cerca del manto elevando la temperatura en dicho lugar, y esto se puede ver evidenciado ya que en estos dos sitios para todos los mapas obtenemos zonas marcadas con un color rojo, el cual va a indicar un alto grado de flujo de calor.

5.4 Valores de puntos aleatorios

Como un último punto se planteó, el ploteo aleatorio de 50 puntos anexados en el Cuadro 5, cuyo objetivo es tratar de hacer una comparación en cualquier parte del área de estudio, entre los datos de flujo de calores reales y los obtenidos por medio de la aplicación y elaboración de los algoritmos ML, mencionados anteriormente, en los que igual se puede decir que los valores que más se parecen a los obtenidos con los datos reales, son con la predicción por medio de las redes neuronales, especialmente para el de 5 neuronas, ya que



el de 1 neurona tenía inclusive una variación mayor con respecto a los datos obtenidos con el de árboles de regresión.

6. Conclusiones

- Al realizar la comparación de los datos obtenidos mediante cada uno de estos diferentes algoritmos tanto, en el Cuadro 3 y el Cuadro 5, el mejor algoritmo para la predicción de nuestros datos de flujo de calor son las redes neuronales, debido a que entre más neuronas se agreguen será mejor la precisión de los datos calculados. Se debe considerar que, pasado un número de neuronas, el algoritmo también comienza a fallar como se observó en el modelo de 7 neuronas, por lo que hay que tener una secuencia de prueba y falla hasta obtener un resultado óptimo. Teniendo en cuenta este inciso, las redes neuronales con 5 neuronas, cumplen los criterios y nos dan un valor más próximo a los reales. Por ende, es el algoritmo que mejor se va a ajustar para los datos de flujo de calor trabajados en el presente artículo.
- El flujo de calor en la zona de estudio, varía y decrece en dirección oeste-este, se debe a que entre más cercano nos encontremos a la zona de subducción mayor va a ser el flujo de calor, el caso exclusivo en el área de estudio son los valores anómalos en la zona sureste, corresponde a los complejos volcánicos del Chalupas y Cotopaxi, los cuales por el hecho de tener una actividad volcánica y su cercanía con el manto, hace que aumente el valor del flujo de calor.
- Los valores obtenidos mediante el algoritmo de los árboles de regresión caen en un bucle y se genera un error, en el cual se tiende a repetir los datos, como se observa en el cuadro 3, esto sucede debido a la cantidad con la que se está trabajando que apenas son 130 mediciones y hay una gran cantidad de variables en este caso, haciendo que para poder cumplir con el algoritmo se vayan repitiendo los datos que ya se trabajaron anteriormente, por lo que no es adecuado para un número muy pequeño de datos, los cuales se los puede predecir con otro algoritmo como es el de las redes neuronales.
- Las redes neuronales para la base de datos propuesta van a tener un error mucho menor de error al calcular el modelo que los árboles de regresión, para una pequeña base de datos es recomendable este primer algoritmo y en función de ésta se van a ir ajustando las capas que como se puede observar en la Figura 3 el modelo de 5 capas cuenta con un error menor que el de 1 y 7 capas.
- Al analizar la Figura 3 mayor sea el número de capas se va a ir disminuyendo el error, pero que estas capas también deben ser controladas ya que si se abusa su empleo van a aumentar el error en vez de disminuir como es el ejemplo del modelo con 7 que aumenta el valor en vez de disminuir.
- Las variaciones entre datos reales y estimados como se apreció a lo largo del artículo van a tener una ligera discrepancia, y este índice de error va a ir disminuyendo o aumentando, según el algoritmo y las condiciones empleadas en éste.
- El algoritmo de árboles de regresión aprende sobre la naturaleza intrínseca del fenómeno del flujo de calor, ya que ha logrado discriminar variables independientes que se ejecutan en modelos matemáticos para conocer valores de flujo de calor.



7. Recomendaciones

- Entre más extensa y confiable sea la base de datos con la que se va a trabajar, mejores van a ser los resultados que se obtengan al emplear estos algoritmos de predicción.
- Para aplicar la predicción de datos de flujos de calor, por medio de redes neuronales se requiere trabajar con la cantidad total de datos, no se debe dividir la muestra en test y entrenamiento, debido a que esto va a generar que no se aproxime a un valor más certero y va a aumentar el margen de error.
- Cuando se ejecute el algoritmo de las redes neuronales hay que tener en cuenta que el número de capas con las que se desee trabajar va a estar en función de la cantidad total de datos que se posean en la base de datos.
- El método de árboles de regresión entre mayor sea la cantidad de datos ingresados va a ser mucho mejor la predicción, ya que, si se emplean muchos datos, el algoritmo va a tender a ocupar un mismo resultado para diferentes bases de datos.
- Para intentar predecir cualquier tipo de variable en un determinado lugar es mucho más factible ocupar algún tipo de tecnología que emplee el ML, debido a que, si lo comparamos con una simple interpolación, ésta no va a ser la más adecuada debido a que lo va a realizar de una manera muy geométrica.
- Para realizar la interpolación, se debe seleccionar un tamaño de píxel pequeño para una mejor relación de los datos, dándonos como resultado una relación inversa de entre más pequeño el tamaño de píxel la calidad de los datos va a ser mucho mejor.

Vínculo: [Informe Completo](#)

8. Agradecimiento

Agradecemos al ingeniero Christian Iván Mejía Escobar por las discusiones fructíferas sobre la el aprendizaje automatizado y la aplicación de los diferentes algoritmos, para el procesamiento de nuestros datos obtenidos mediante el empleo del software R Studio durante las tutorías de la asignatura de software especializado a la ingeniería y por los conocimientos impartidos en las clases que nos han permitido entender las características y el comportamiento tanto del programa como de los diferentes algoritmos siendo estos en el caso de estudio los de las redes neuronales y árboles de decisión.

9. Bibliografía

Castillo-Calderon, J., Solorzano-Sastillo, B. y Moreno-Moreno, José. Diseño de una red neuronal para la predicción del coeficiente de pérdidas primarias en régimen de flujo turbulento. Ingenius [online]. 2018, n.20, pp.21-27. ISSN 1390-860X. <https://doi.org/10.17163/ings.n20.2018.02>.

CGMW. (2007). [Magnetic Anomaly Map of the World]. Recuperado el 28 de julio, 2020, de: <https://ccgm.org/en/maps/113-carte-des-anomalies-magnetiques-du-monde.html>

Contreras, Juan. (2015). Modelado de Procesos Terrestres. Flujo de Calor en la Litosfera Terrestre-Capítulo de libro. Research Gate.



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](#)

- Davies, J. H. (2013). Global map of solid Earth surface heat flow. *Geochemistry, Geophysics, Geosystems*, 14(10), 4608–4622. doi:10.1002/ggge.20271
- Grant, J. (2020) Aprendizaje Automático Profundo: Complete consejos y trucos para el aprendizaje automático profundo (Libro En español/Deep Machine Learning Spanish Book Version). Independently Published, de <https://books.google.com.ec/books?id=bIOBzQEACAAJ>
- Hall, M., Ramón, P., Samaniego, P., Le Pennec, Johnson, J.B., 2008. Ecuadorian Andes volcanism: A review of Late Pliocene to present activity, Ecuador. *Rev. Journal of Volcanology and Geothermal Research* 176 (2008) 1-6
- Hantschel, T., & Kauerauf, A. I. (2009). Heat Flow Analysis. In *Fundamentals of Basin and Petroleum Systems Modeling* (pp. 103–150). <http://doi.org/10.1007/978-3-540-72318-9>
- Hamza, V. & Vieira, Fábio. (2012). Global distribution of the lithosphere-asthenosphere boundary: A new look. *Solid Earth*. 3. 199-212. 10.5194/se-3-199-2012.
- Hulbert, C., Rouet-Leduc, B., Johnson, P.A. et al. Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience* 12, 69–74 (2019). <https://doi.org/10.1038/s41561-018-0272-8>
- NOAA. (2010). [World Digital Magnetic Anomaly Map]. Recuperado el 3 de septiembre, 2020, de: <https://www.ngdc.noaa.gov/geomag/NGDC720/index.html>
- Maus, S., An ellipsoidal harmonic representation of Earth's lithospheric magnetic field to degree and order 720, *Geochem. Geophys. Geosyst.*, 11, Q06015, doi: 10.1029/2010GC003026, 2010
- Pasquale, V., Verdoya, M., & Chiozzi, P. (2014). Heat Conduction and Thermal Parameters. In *Geothermics*. <http://doi.org/10.1007/978-3-319-02511-7>
- Pollack, H. N., Hurter, S. J., & Johnson, J. R. (1993). Heat flow from the Earth's interior: Analysis of the global data set. *Reviews of Geophysics*, 31(3), 26
- Rezvanbehbahani, S., Stearns, L. A., Kadivar, A., Walker, J. D., and van der Veen, C. J. (2017). Predicting the geothermal heat flux in Greenland: A machine learning approach. *Geophysical Research Letters*, 44, 12,271–12,279. <https://doi.org/10.1002/2017GL075661>
- Russell, S. J., Norvig, P. (2009) *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall
- Sterrett, S. (2000) Turing's two tests for intelligence, *Minds and Machines* 10(4): 541-559.

Autores

MARITZA BARRERA-RAMIREZ cursando el noveno semestre de la carrera de Ingeniería en Geología en la Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental de la Universidad Central del Ecuador (Ecuador) en 2020

ESTEBAN ARMAS-CEPEDA cursando el noveno semestre de la carrera de Ingeniería en Geología, en la Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental de la Universidad Central del Ecuador (Ecuador) en 2020.



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

SANTIAGO ALVAREZ RUALES cursando el noveno semestre de la carrera de Ingeniería en Geología, en la Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental de la Universidad Central del Ecuador (Ecuador) en 2020.

CHRISTIAN MEJÍA-ESCOBAR docente en la Facultad de Ingeniería en Geología, Minas, Petróleos y Ambiental de la Universidad Central del Ecuador (Ecuador) en 2020.



[Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)