



REVISTA FIGEMPA

Aplicación de algoritmos de clasificación para la estimación de carbono orgánico del suelo en la provincia de Chimborazo, Ecuador.

Application of classification algorithms for the estimation of soil organic carbon in the province of Chimborazo, Ecuador

Óscar Padilla | Escuela Superior Politécnica de Chimborazo, Ecuador. oscar.padilla@esPOCH.edu.ec
Silvia Haro | Escuela Superior Politécnica de Chimborazo, Ecuador

HISTORIAL DEL ARTÍCULO

Recepción: 03/08/2021
Aceptación: 08/11/2021

PALABRAS CLAVE

Árboles de clasificación, algoritmos de clasificación supervisada, carbono edáfico.

ARTICLE HISTORY

Received: 03/08/2021
Accepted: 08/11/2021

KEY WORDS

Classification trees, supervised classification algorithms, edaphic carbon.

RESUMEN

El estudio tuvo como objetivo evaluar la técnica de árboles de decisión mediante el mejor algoritmo de clasificación supervisada, que permita predecir el contenido de carbono edáfico en la provincia de Chimborazo en zonas nativas o endémicas, considerando la base de datos del Ministerio de Agricultura y Ganadería (MAG). En el estudio se realizó la limpieza del conjunto de datos y se determinaron 10 variables útiles para la categorización de carbono orgánico del suelo, obteniendo 4 clases: muy alto, alto, medio y bajo. Se determinó la eficiencia de tres algoritmos: C5.0, SMV y CART, seleccionándose el CART mediante el método de validación cruzada para la construcción del árbol. Los resultados con el conjunto de datos de prueba generó una precisión del 63,41 puntos porcentuales y un error de predicción de 36,59 por ciento; estos alcances se presentan como una nueva alternativa de cuantificación de COS, el modelo calibrado puede ser extendido sin necesidad de muestrear in situ, muy útil en zonas complejas como el ecosistema de bosque alto andino. El mapeo digital permitió revelar los niveles de COS existentes en suelos de la provincia de Chimborazo

ABSTRACT

The objective of the study was to evaluate the decision tree technique using the best supervised classification algorithm, which allows predicting the edaphic carbon content in the province of Chimborazo in native or endemic areas, considering the database of the Ministry of Agriculture and Livestock (MAG). In the study, the data set was cleaned and 10 useful variables were determined for the categorization of soil organic carbon, obtaining 4 classes: Very High, High, Medium and Low. The algorithm that provided the best percentage of efficiency and relevant results was Classification and Regression Trees (CART) using the cross-validation method. The efficiency of three algorithms was determined: C5.0, SMV and CART, selecting the CART by means of the cross-validation method for the construction of the tree. The results with the test data set generated a precision of 63,41 percentage points and a prediction error of 36,59 percent, these scopes are presented as a new alternative for SOC quantification, the calibrated model can be extended without the need to sample in situ, very useful in complex areas such as the forest ecosystem. The digital mapping allowed to reveal the existing SOC levels in soils of the Chimborazo province.

INTRODUCCIÓN

Los análisis de datos principalmente se basan en técnicas estadísticas; sin embargo, a medida que la información almacenada crece considerablemente, los métodos estadísticos tradicionales han empezado a enfrentar problemas de eficiencia, valoración y de clasificación. La minería de datos (Graham, 2011), es una solución para el

análisis de fenómenos no explícitos en bases de datos y la búsqueda de patrones ocultos entre éstos, información que posteriormente puede ser utilizada en la predicción de comportamientos (Acosta, *et al.*, 2018; Romero, 2018).

Una técnica en minería de datos son los árboles de decisión, mismos que se basan en algoritmos de clasificación (Pérez y González, 2018), los cuales son efectivos y

de fácil interpretación. Se remontan a 1944, por una idea de Von Neuman (1903-1958) utilizada en la teoría de juegos (Gastaldi, *et al.*, 1998). Esta técnica permite identificar la mejor opción de diversas alternativas y se muestra en forma gráfica, facilitando la decisión a tomar (Roche, 2009). Son una técnica de aprendizaje supervisado, la cual consiste en construir modelos para clasificar nuevas observaciones (Ruiz, *et al.*, 2019), o para interpretar la información y transformarla en conocimiento (Rodríguez y Camacho, 2018). Los árboles de decisión, son muy utilizados para la predicción, pues al trabajar con datos que almacenan información pasada, permite realizar pronósticos (Quintero y Amézquita, 2003).

Los suelos son los principales recursos naturales del mundo, siendo el sistema edáfico uno de los ejes fundamentales para la producción de alimentos (Aguilera, 2020). El carbono del suelo (COS) es un componente que conforme se incrementa mejora la calidad y fertilidad del mismo; es por ello que el presente trabajo se enfocó en la clasificación del COS a nivel de la provincia de Chimborazo en zonas nativas o endémica, con el fin de determinar un modelo predictivo basado en datos históricos que permita conocer el comportamiento del COS en los diferentes ecosistemas, es así que para su descripción se emplearon variables que inciden directamente en sus niveles, considerándose: el *ecosistema* (Villegas, 2013); sistema biológico constituido por una comunidad de organismos vivos y el medio físico, la *taxonomía* (Sigtierras, 2017); variable que indica la clasificación de los suelos en función de varios parámetros y propiedades, la *textura* (Blanquer, *et al.*, 2010); muestra el contenido relativo de partículas de diferente tamaño, la *pendiente* (Ministerio del Ambiente, 2013); el *declive del terreno* y la *inclinación* (DEM) (López, 2002), la *altura* sobre el nivel del mar a la que se realizó el punto de muestreo, y 4 índices espectrales: el Índice de resistencia atmosférica visible (García, *et al.*, 2020), mismo que resalta la vegetación en la parte visible del espectro, el Índice diferencial de agua normalizado (Delpino, *et al.*, 2018), empleado para evaluar cuerpos de agua cubierta, el Índice de área calcinada (Mena, 2017); mismo que identifica superficies de suelo desnudo en ambientes de baja intervención antrópica y el Índice de vegetación mejorado de dos bandas (Gilabert, *et al.*, 1997), el cual optimiza la respuesta espectral de la vegetación con alta densidad.

Materiales y métodos

La zona de estudio fue la provincia de Chimborazo-Ecuador, localizada en el centro sur del país; en la zona geográfica 17 (INEC, 2010). Se realizó una

investigación exploratoria, cuasi experimental del tipo de inferencia inductiva, en un periodo transversal.

Los datos empleados para la construcción de los modelos de clasificación, se obtuvieron a partir de la base proporcionada por el Ministerio de Agricultura y Ganadería (MAG), del Proyecto Regional de Cooperación de Capacitación de Mapeo de Suelos de la FAO. Para clasificar los niveles de COS se generó una variable categórica denominada Niveles, para lo cual se empleó la categorización realizada por Vela, López y Rodríguez (Vela, *et al.*, 2012); quienes en su investigación plantearon límites en los intervalos; en función de los valores máximos y mínimos de concentración de COS, tal como muestra la tabla 1 (ver Tabla 1).

Se trabajó con una data de 591 observaciones que describen 10 variables, tabla 2. Se consideró al COS como la variable predictora, las muestras fueron recolectadas mediante el método Walkley Black (Eyherabide, *et al.*, 2014); a través del cual se midió la densidad aparente (g.cm⁻³) y materia orgánica (%), con un factor de transformación de 1,725 en CO (%), para lo cual se empleó el método de barrenación. Las variables explicativas: ráster de ecosistema, taxonomía, textura, pendiente, DEM y los índices espectrales; estas variables y los cuatro índices espectrales se obtuvieron a partir de las bandas de imagen satelital Landsat 8 (Zhiminaicela, *et al.*, 2020) y su medición se realizó al tope de la atmósfera, para lo cual se utilizó el sistema de información geográfica y el software libre QSIG V.3.14 (ver Tabla 2).

Con el fin de seleccionar el algoritmo adecuado para el conjunto de datos, se aplicaron los modelos:

C5.0: desarrollado por Ross Quinlan en 1993 (Quinlan, 1993), como una mejora del C4.5 el cual es una extensión del ID3, desarrollado en 1986 por el mismo autor (Quinlan, 1986). Se basan en el concepto de entropía a la hora de construir el árbol, misma que representa una medida de impureza o incertidumbre (Díaz, *et al.*, 2004), su ecuación es:

$$E(S) = \sum_{i=1}^c -p_i \log_2(p_i) \quad \text{Ec. (1)}$$

Donde: S es el conjunto de muestras, c: es el número de diferentes clasificaciones y p_i : es la proporción de ejemplos que hay de la clasificación en la muestra.

SMV: la técnica de máquinas de vectores de soporte se utiliza en clasificación y regresión (Cristianini y

Shawe-Taylor, 2000), está basada en la idea de minimizar un límite superior del error de generalización en vez del principio de minimización de riesgo empírico (Remp). Emplea la ecuación:

$$f(\mathbf{x}) = \mathbf{w}^t * \mathcal{O}(\mathbf{x}) + b \quad \text{Ec. (2)}$$

donde: $\mathcal{O}(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^H$ ($d < H$) es una transformación del espacio dimensión. El vector define el hiperplano de separación en dicho espacio y representa el sesgo respecto al origen de coordenadas (Ureña, *et al.*, 2006).

CART: es un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos (Bartolini, *et al.*, 2013). Con la idea de buscar particiones que discriminen o consigan nodos más puros se considera el Índice Gini (Haro, 2019; Quinlan, 1993).

$$I(s) = \sum_{j=1}^c P_j^*(P_j^1, P_j^2, \dots, P_j^c)$$

donde el índice de Gini, en el nodo t se puede formular de la siguiente manera:

$$1 - \sum (P_j)^2$$

El estadístico utilizado para la selección del modelo adecuado fue el coeficiente de Kappa; mismo que se emplea para cuantificar el grado de acuerdo entre las observaciones, se caracteriza por corregir el factor azar. La escala fundamental para interpretar este coeficiente, se muestra en la tabla 3 (López de Ulibarri y Pita, 1999) (ver Tabla 3).

En el estudio se utilizó el software estadístico libre RStudio V.1.3.1 bajo la plataforma R V.3.6.3. Para la construcción del árbol de decisión se procedió a dividir la base de datos en conjuntos de entrenamiento y prueba (Ferreira, *et al.*, 2018), se utilizó el criterio de Houlduot mediante el cual se consideró el 70% de la data para el conjunto de entrenamiento y el restante 30% para el testing. Con el objetivo de generar un modelo óptimo se aplicó la técnica de Validación Cruzada (cv), misma que modifica los parámetros y garantiza que los resultados sean independientes de la partición entre los datos de prueba y de entrenamiento (Pérez, 2020).

La validación de los modelos se realizó a través de la matriz de confusión (Fawcett, 2006), la cual compara el conjunto de variables de la tabla testing con la predicción obtenida, versus el rango de cos a la que los datos realmente pertenecen (Haro, *et al.*,

2018). Las columnas determinan el número de predicciones de cada clase (rango de cos), y las filas representan las instancias clasificadas en la clase real. La tabla 4, muestra la matriz de confusión para dos clases.

Las medidas de rendimiento (Haro, *et al.*, 2018) empleadas en el estudio fueron:

Precisión global P (exactitud): esta medida permite calcular la proporción de datos predichos correctamente VN y VP en relación a la suma de observaciones evaluadas, se calcula mediante la ecuación:

$$P = \frac{VN + VP}{VN + FP + FN + VP}$$

Error de predicción (error): esta tasa es la proporción de observaciones mal clasificadas y se calcula mediante la expresión:

$$E = \frac{FN + FP}{VN + FP + FN + VP}$$

Después de la generación y selección del modelo se generó un mapa georreferencial del contenido de cos en la provincia de Chimborazo, empleándose el software geográfico QGIS.

Resultados

La tabla 5, muestra las medidas de rendimiento de los modelos empleados en el estudio; se puede observar que el algoritmo CART es óptimo para el conjunto de datos, presenta una mayor precisión cuando es y no aplicada la técnica cv, mientras que el índice Kappa mostró un mayor rendimiento con cv, obteniendo un valor de 0,492; lo cual establece una fuerza de concordancia moderada (ver Tabla 5).

En la tabla 6, se observan las variables independientes posicionadas en una serie ordenada de acuerdo a su importancia, siendo *textura* la variable predominante y el Índice de área calcinada (BI) como la variable sin importancia.

En la figura 1, se observa el árbol de decisión construido tras el entrenamiento del algoritmo CART con los datos del MAG en la provincia de Chimborazo. El modelo generó *textura* como variable predominante con 15 reglas de decisión, de las cuales 5 representan a la clasificación del carbono como muy alto, 4 a la clasificación alto, 4 a la clasificación medio y 2 a la clasificación del carbono bajo. Una de las

Niveles	Rango de COS
Muy Alto	> 215 Mg/ha
Alto	152 - 215 Mg/ha
Medio	80 - 152 Mg/ha
Bajo	< 80 Mg/ha

Tabla 1: Rangos de COS.

Variables	Valores
COS	Carbono orgánico del suelo con valores Bajo, Medio, Alto, Muy Alto
Ecosistema	Intervención, BSV_Piemontano, BSV_Montano, Arbustal y Herbazal
Taxonomía	Alfisol, Entisol, Inceptisol, Mollisol
Textura	Fina, Media Moderada y Gruesa
Pendiente	Colinado, Escarpado, Montañoso, Ondulado y Plana
DEM	Altura sobre el nivel del mar en metros
VARI	Índice de resistencia atmosféricamente visible con valores [-1, 1]
NDWI	Índice diferencial de agua normalizada con valores [-1, 1]
BI	Índice de área calcinada con valores [-1, 1]
NBR2	Índice normalizado de Áreas Quemadas 2 con valores [-1 y 1]

Tabla 2: Valores que puede tomar las variables.

Valor de k	Fuerza de concordancia
< 0.20	Pobre
0.21 - 0.40	Débil
0.41 - 0.60	Moderada
0.61 - 0.80	Buena
0.81 - 1.00	Muy Buena

Tabla 3: Valoración del índice Kappa.

Clase original	Clase Predicha	
	Positivo (P)	Negativo (N)
Positivo (P)	VP (Verdadero Positivo)	FN (Falso Negativo)
Negativo (N)	FP (Falso Positivo)	VN (Verdadero Negativo)

Tabla 4: Matriz de confusión.

Algoritmos	Validación Cruzada (CV)			IC: 95%
	Precisión	Error	Kappa	
C5.0	50.62	49.48	0.367	(0.45, 0.60)
SVM	57.63	42.37	0.386	(0.47, 0.63)
CART	63.41	36.59	0.492	(0.54, 0.65)

Tabla 5: Comparación de los algoritmos de clasificación.

Variables Independientes	Importancia	Importancia Normalizada
Textura	38.00	100 %
NBR2	32.49	85.50 %
Taxonomía	30.83	81.13 %
VARI	28.02	73.74 %
NDWI	24.69	64.97 %
DEM	11.42	30.05 %
Pendiente	11.12	29.26 %
Ecosistema	7.77	20.47 %
BI	0	0 %

Tabla 6: Importancia de las variables.

Carbono Taxonomía	Alto	Bajo	Medio	Muy Alto	
Alfisol	6	0	0	63	69
Entisol	35	10	59	27	131
Inceptisol	5	1	6	121	133
Mollisol	49	29	67	113	258
	95	40	132	324	591

Tabla 7: Matriz de confusión.

reglas de decisión para la clasificación de carbono edáfico que arrojó el modelo fue: si la muestra obtenida se encuentra en una textura media y moderada (textura = media, moderada) considerando el índice de resistencia atmosféricamente visible mayor o igual a 0,025 ($\text{VAR} \geq 0,025$), con ecosistemas; bosque siempre verde montano y herbazal (ecosistema = BSV montano, herbazal), entonces el carbono es muy alto (ver Figura 1).

La tabla 7 muestra la matriz de confusión del modelo árbol de decisión. Se puede observar que, de las 591 muestras 121 se clasificaron en el tipo de suelo inceptisol con un nivel muy alto de COS, seguido por mollisol con 113 datos clasificados en el mismo nivel (ver Tabla 7).

Mediante el mapeo digital del COS, figura 2(a), se puede observar la predicción de carbono edáfico generado a través de la técnica de árboles de decisión, se muestra un mapa con los niveles de COS representado de color de tierra. El tono más oscuro indica niveles muy altos de clasificación para el COS, en tanto que el color claro muestra la clasificación de COS como bajo. Mediante la predicción del COS en toneladas por hectárea se obtuvo un mapa con los valores del carbono edáfico estableciendo un máximo de 358,92 Mg/ha y un mínimo de 19,23 Mg/ha (ver Figura 2).

En la figura 2(b) se visualizan las áreas de acuerdo a los niveles de carbono orgánico del suelo, en la provincia de Chimborazo. Se determinaron niveles de carbono edáfico muy altos en zonas nativas de bosque siempre verde montano andino; ubicados al oeste de los cantones de Penipe, Chambo y Riobamba, en todo el territorio de los cantones de Pallatanga y Cumandá; se establecieron niveles altos de COS en zonas de bosque siempre verde piemontano andino, en la parte central de los cantones de Colta, Guamoto, Alausí y Chunchi; se observó un nivel medio en zonas con arbustal y herbazal; así como en la parte central del cantón Guano; además, se determinó un nivel bajo en zonas intervenidas, en la parte central del cantón Riobamba.

Discusión

Los resultados obtenidos en la investigación corroboraron que el tipo de suelo inceptisol presentó un nivel muy alto de COS, esto se debe a que éste es un suelo joven que conserva su materia orgánica, y que de acuerdo a Martínez (2008) es el tercero en

acumular mayor cantidad; con un valor de 0,016 COS/área, acorde a la clasificación de carbono orgánico en todo el mundo realizado por Edwars (Edwars y Bohlen, 1996), después le sigue el suelo mollisol; mismo que tiene un gran impacto en los niveles de COS tal como lo determinan Novelli *et al.* (2013), en su estudio relacionado con la intensificación en el uso del suelo.

Las zonas de la provincia de Chimborazo con niveles de COS muy altos (> 215 Mg/ha) fueron las nativas de bosque siempre verde montano andino, estos lugares se caracterizan por tener una vegetación de transición entre el bosque de niebla y el páramo, este resultado es confirmado por Loayza (2020); donde se establece que los niveles de COS son favorecidos en tierras con mayor altitud, climas de bajas temperaturas y zonas húmedas de la Sierra ecuatoriana. Se evidenció niveles medios de COS en la zona arbustal y herbazal, es decir, en los páramos de la provincia; lugares que presentan vegetación fragmentada, donde se mezclan suelos desnudos con herbazales y arbustos. Niveles bajos de COS (< 80 Mg/ha) se localizaron en zonas intervenidas, esto se debe a que existe una reducción del contenido de materia orgánica; y que de acuerdo a Cárdenas (Cárdenas, 2015) se manifiesta por el uso que se da al suelo, como la agricultura y ganadería; siendo éstas las principales fuentes económicas de la provincia de Chimborazo.

La implementación del árbol de clasificación para identificar niveles de COS en la provincia de Chimborazo, fue una herramienta que caracterizó adecuadamente los suelos y zonas de estudio, pero un limitante fue el tamaño de la muestra; por lo que para futuras investigaciones es necesario incrementar la base de datos.

Conclusiones

La técnica de árboles de decisión mediante el algoritmo de clasificación CART, permitió identificar variables que inciden en los niveles de carbono edáfico en la provincia de Chimborazo, siendo éste un componente que determina la calidad de los suelos, ya que es fuente y almacenamiento de nutrientes, lo cual tiene incidencia directa en la flora y fauna de los ecosistemas. La *textura* del suelo fue la variable predominante en el estudio, por lo que el contenido de partículas y sus tamaños son significativas en los niveles de carbono edáfico en la provincia de Chimborazo. Las zonas nativas o endémicas con nivel muy

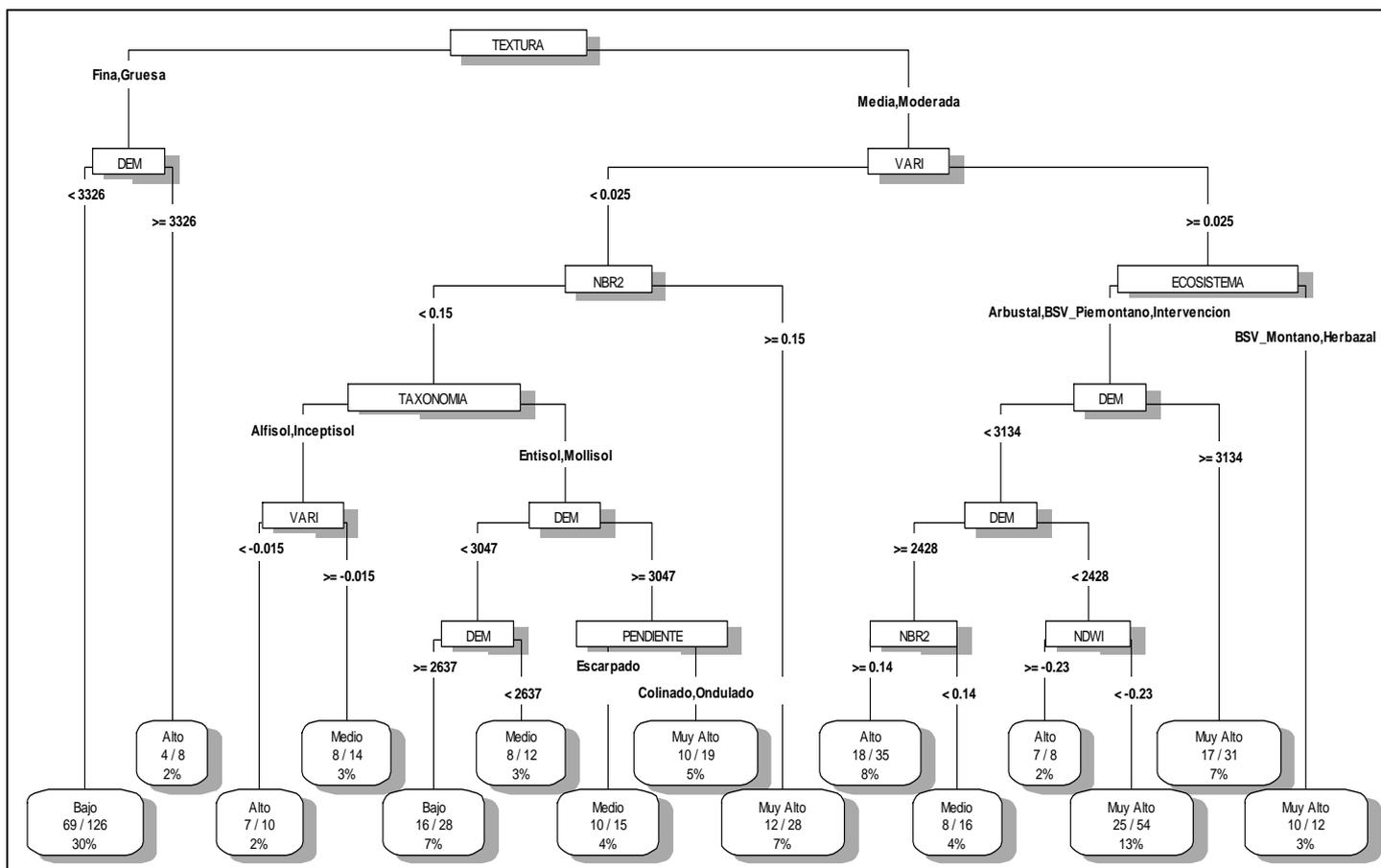


Figura 1: Árbol de Decisión.

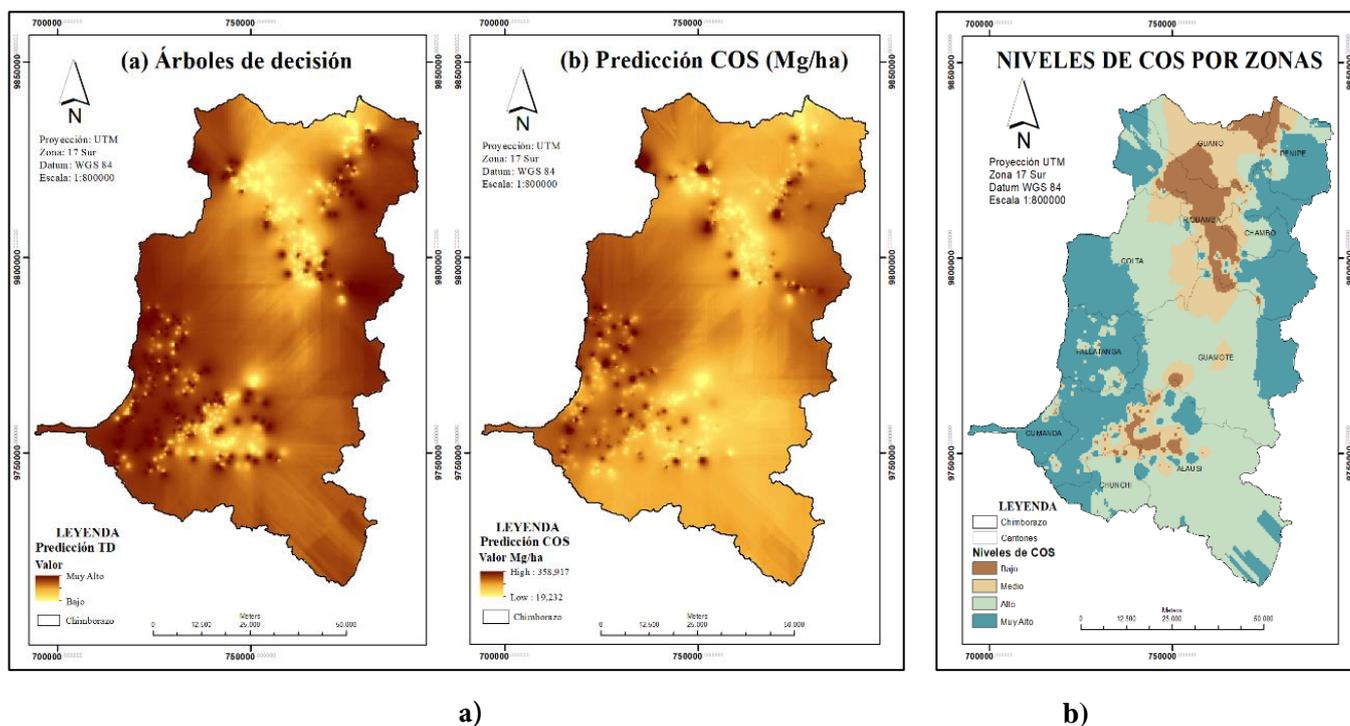


Figura 2: Niveles de COS: a) predicción b) por Zonas

alto de COS, fueron las situadas en los cantones de Penipe, Chambo, Riobamba, Guamote, Pallatanga y Cumandá; poder identificar éstas áreas permitirá su protección y creación de políticas que conlleven a su conservación, pues el COS influye directamente en el cambio climático del planeta, así como en los ciclos del agua. Además, estudios de este tipo son de gran relevancia ya que siendo la agricultura una de las principales actividades económicas en la provincia de Chimborazo, el poder identificar las zonas con bajos niveles de COS ayudará a mejorar la calidad de los suelos; pues su degradación incide directamente en la pérdida de producción agrícola y por ende en la seguridad alimentaria.

Agradecimientos

Al Ministerio de Agricultura y Ganadería (MAG) como entidad auspiciante de la base de datos del carbono orgánico del suelo (COS) en la provincia de Chimborazo.

Referencias

- Acosta, J., Martínez, D. y Primorac, C. (2018). *Determinación de perfiles de rendimiento académico en la UNNE con minería de datos educacionales*. Argentina. xx Wokshop de Investigación en Ciencias de la Computación, 1078-1082, abril.
- Aguilera, S. (2020). *Importancia de la protección de la materia orgánica en suelos*. Repositorio Académico, 77-85.
- Bartolini, J. L., Alonso, P. y Alvarez, F. (2013). Árboles de clasificación de Potimirim mexicana (*Decapoda: Caridea*), organismo hermafrodita protándrico secuencial. *Latin american journal of aquatic research*, 41(4), 739-745.
- Blanquer, J., Ibáñez, S. y Moreno-Ramón, H. (2010). La textura del suelo. *RiuNet*. Universidad Politécnica de Valencia.
- Cárdenas, M. (2015). *Evaluación de la calidad de los suelos de páramo intervenidos y no intervenidos en la comunidad Monjas Bajo, parroquia Juan Montalvo, cantón Cayambe*. Universidad Politécnica Salesiana, sede Quito, Ingeniería en biotecnología de los recursos naturales.
- Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Press CU.
- Delpino, M., Portillo, V. y Mora, C. (2018). Evaluación de índices espectrales derivados de sensores remotos para la caracterización de ambientes humedales. *Agropecuaria, Anais 7.º Simposio de Geotecnologías no Pantanal. Embrapa Informática Agropecuária/INPE*, p. 112-121.
- Díaz, Z., Fernández, J. y Segovia, M. (2004). Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 12(1), enero.
- Edwards, A. and Bohlen, P. (1996). *Biology and ecology of earthworms*. Springer.
- Eyherabide, M., Saínz, H., Barbieri, P. y Echevería, H. (2014). Comparación de métodos para determinar carbono orgánico en suelo. *Ciencia del Suelo*, 32(1), 13-19.
- Fawcett, T. (2006). An introduction to ROC analysis. *ScienceDirect*, 27(8), 861-874, junio.
- Ferreira, M., Hosoda, B., Franco de Los Santos, E. y Fernandes, A. (2018). Uso de datos abiertos y técnicas de minería de datos para la clasificación de estudiantes de instituciones privadas de educación superior de Belem-PA. *Research in Computing Science*, 147(5), 27-36, mayo.
- García, H., Flores, H., Khalil-Gardezi, A., Roberto, A., Tijerina-Chávez, L., Peña, M. (2020). Estimación de la fracción de cobertura de la vegetación en maíz (*Zea mays*) mediante imágenes digitales tomadas por un vehículo aéreo no tripulado (UAV). *Fitotecnia Mexicana*, 43(12).
- Gastaldi, C., Urrea, M. y Fernández, P. (1998). Teoría de decisión: contribuciones de von Neumann. *Divulgaciones Matemáticas*, 6(1), 37-42.
- Gilabert, M., González-Piqueras, J. García, J. (1997). Acerca de los índices de vegetación. *Teledetección*, 8(10).
- Graham, W. (2011). *Data mining with Rattle and R*. Springer.
- Haro, S. (2019). Árbol de decisión, aplicación con datos meteorológicos. Riobamba, VI Congreso Internacional de la Ciencia, Tecnología, Emprendimiento e Innovación, pp. 63-71.
- Haro, S., Pazmiño, R., Conde, M. Peñalvo, F. (2018). *Minería de datos para descubrir tendencias en la clasificación de los trabajos de titulación*. Quito, III Congreso de Ciencia & Tecnología, ESPE, 125-128.
- Haro, S. Zúñiga, L., Meneses, A., Vera, L. y Escudero, A. (2018). Métodos de clasificación en minería de datos meteorológicos. *Perfiles*, 2(20), 107-113, octubre.

- INEC, Instituto Nacional de Estadísticas y Censos, 2010. [En línea] <https://www.ecuadorencifras.gob.ec/ambiente-y-agropecuario-2/> [Último acceso: 2020].
- Loayza, V., Sevilla, V., Olivera, C., Guevara, M., Olmedo, G., Vargas, R. *et al.* (2020). Mapeo digital de carbono orgánico en suelos de Ecuador. *Revista Científica de Ecología y Medio Ambiente*, 29(2), 1-13, mayo-agosto.
- López de Ulibarri, G. Pita, S. (1999). Medidas de concordancia. El índice Kappa. *Cuadernos de atención primaria*, 6(4), 223-226.
- López, R. (2002). *Degradación del Suelo; causas, procesos, evaluación e investigación*. CIDIAT.
- Martínez, E., Fuentes, J. P. y Acevedo, E. (2008). Carbono orgánico y propiedades del suelo. *Revista de la ciencia del suelo y nutrición vegetal*, 8(1), 68-96.
- Mena, S. (2017). Identificación de áreas afectadas por incendios forestales utilizando imágenes satelitales LANDSAT 8. *GEO.SUR*, 8(12), 15-25.
- Ministerio del Ambiente. (2013). Modelo de unidades geomorfológicas para la representación cartográfica de ecosistemas del Ecuador continental. [En línea]. <http://app.sni.gob.ec/sni-link/sni/Gobierno%20Provincial%20Azuay/limites%20GPA/EstudioLimitesAzuayLoja/Geoformas453103936428118386.pdf> [Último acceso: 2020].
- Novelli, L. Caviglia, O. P., Melchiori, R., Wilson, M., Quintero, C. E., Sasal, M. C. *et al.* (2013). Impacto de la intensificación en el uso del suelo en molisoles y vertisoles de la zona agrícola de Entre Ríos. *Ciencia, Docencia y Tecnología Suplemento*, 3(3), 1-17.
- Pérez, B. (2020). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. *UIS Ingenierías*, 19(1).
- Pérez, J. y González, F. (2018). Árboles de clasificación vs. regresión logística en el desarrollo de competencias genéricas en ingeniería. *Computación y Sistemas*, 1519-1541, diciembre.
- Quinlan, J. R. (1986). Introduction of decision trees. *Machine Learning*, 1(1), 81-106, marzo.
- Quinlan, J. R. (1993). *C4.5. Programs for machine learning*. Kluwer Academic Publishers.
- Quintero, M. Amézquita, E. (2003). CGIAR [En línea]. <https://cgspace.cgiar.org/handle/10568/69621> [Último acceso: 2020].
- Roche, A. (2009). Árboles de decisión y series de tiempo (tesis de maestría en Ingeniería Matemática). Montevideo. Universidad de la República, Facultad de Ingeniería.
- Rodríguez, S. y Camacho, J. (2018). Los métodos de aprendizaje automático supervisado en la clasificación textual según el grado de especialización. *Tonos Digital*, (35), 131-149, julio.
- Romero, C. (2018). Estudio comparativo de algoritmos de inteligencia artificial y minería de datos enfocados a la toma de decisiones empresariales de elecciones de personal, (Trabajo de grado, Ingeniería en sistemas). Bogotá. Universidad de Cundinamarca.
- Ruiz, S., Sued, M., Vidal, L., Rodríguez, D., Nesbitt, S. y García, Y. (2019). Técnicas de clasificación supervisada para la discriminación entre ecos meteorológicos y no meteorológicos usando información de un radar de banda C. *Meteorológica*, 44(1).
- SIGTIERRAS. (2017). Ministerio de Agricultura y Ganadería, Ecuador. [En línea]. http://metadatos.sigtierras.gob.ec/pdf/memoria_mapa_de_ordenes_de_suelos_mag_sigtierras.pdf [Último acceso: 2020].
- Ureña, R., Pérez, F. y Díaz de María, F. (2006). *Estimación de probabilidades a posteriori en SMV multiclase para reconocimiento de habla continua*. Zaragoza, IV Jornadas en Tecnología del Habla, 213-218.
- Vela, G., López, J. y Rodríguez, M. d. L. (2012). Niveles de carbono orgánico total en el suelo de conservación del Distrito Federal, centro de México. *Investigaciones Geográficas, Boletín del Instituto de Geografía, UNAM*, (77), 18-30, julio.
- Villegas, T. T. W. (2013). *Sistemas de clasificación de ecosistemas del Ecuador continental*. Ministerio del ambiente. [En línea]. <http://app.sni.gob.ec/sni-link/sni/pdot/nivel%20nacional/mae/ecosistemas/documentos/sistema.pdf> [último acceso: 2020].
- Zhiminaicela, J., Quevedo, J. y Morocho, A. (2020). Deforestación y cambios en la cobertura vegetal del archipiélago de Jambelí, mediante el uso de imágenes satelitales Landsat-8. *Manglar*, 17(2), 153-157.