

03

Fecha de presentación: marzo, 2022
Fecha de aceptación: junio, 2022
Fecha de publicación: septiembre, 2022

APLICACIÓN DE ALGORITMOS

DE APRENDIZAJE AUTOMÁTICO PARA CLASIFICAR LA FERTILIDAD DE UN SUELO BANANERO

APPLICATION OF AUTOMATIC LEARNING ALGORITHMS TO CLASSIFY THE FERTILITY OF A BANANA SOIL

Harry Vite-Cevallos¹

E-mail: hvite@utmachala.edu.ec

ORCID: <https://orcid.org/0000-0003-2056-7111>

Héctor Carvajal-Romero¹

E-mail: hcarvajal@utmachala.edu.ec

ORCID: <https://orcid.org/0000-0001-6303-6295>

Salomón Barrezueta-Unda¹

E-mail: sabarrezueta@utmachala.edu.ec

ORCID: <https://orcid.org/0000-0003-4147-9284>

¹Universidad Técnica de Machala

Cita sugerida (APA, séptima edición)

Vite-Cevallos, H., Carvajal-Romero, H., & Barrezueta-Unda, S., (2022). Aplicación de algoritmos de aprendizaje automático para clasificar la fertilidad de un suelo bananero. *Revista Universidad y Sociedad*, 14(5), 29-33.

RESUMEN

La investigación propone un algoritmo para integrar técnicas de aprendizaje automático a la gestión del ciclo nutricional del banano. El tipo de investigación es correlacionar y descriptiva, detallando las actividades que se debe realizar para lograr articular el uso de aprendizaje automático en la toma de decisiones del productor bananero, utilizando a través de métodos supervisados, técnicas que permitieron clasificar los datos de estudio, seleccionando al algoritmo de Árboles de Decisión, el cuál clasificó correctamente la información, facilitando la predicción del comportamiento de los nutrientes del suelo, focalizando la zona que presentó variaciones en los nutrientes, facilitando la toma de decisiones al productor bananero y la optimización de recursos.

Palabras clave: Aprendizaje Automático, nutrientes del suelo, toma de decisiones, agricultura de precisión.

ABSTRACT

The research focuses on describing how to integrate machine learning techniques into the management of the banana nutritional cycle. The type of research is correlating and descriptive, detailing the activities that must be carried out in order to articulate the use of automatic learning in the decision making of the banana producer, using supervised methods, techniques that allowed classifying the study data, selecting to the algorithm of Decision Trees, which correctly classified the information, facilitating the prediction of the behavior of soil nutrients, focusing the area that presented variations in nutrients, facilitating decision making to the banana producer and optimization of resources.

Keywords: Automatic learning, soil nutrients, decision making, precision agriculture.

INTRODUCCIÓN

El sector agrícola, tiene la necesidad de mejorar los procesos de producción para ser más eficientes con el uso de los recursos que nos provee el ecosistema como son los nutrientes del suelo (Riba et al., 2018); y con esto disminuir costos. Pero lograr esta eficiencia, se requiere de un sistema que permitan integrar nuevas tecnologías, para generar nuevas variables que permitan la construcción de modelos predictores para la toma de decisiones (Moreno-Carriles, 2018).

Por otra parte, la gestión de un sistema agrícola involucra actividades relacionados a planificar, organizar, controlar y dirigir los recursos como los nutrientes del suelo (Fernández, Fernández, Rivera & Calero, 2016; Fajardo, Aguilar, Flores, Parra & Acurio, 2017). En este aspecto, el requerimiento nutricional es diferente en cada especie vegetal, un desbalance de los nutrientes por causas naturales o por un mal manejo agrícola, tiene efectos negativos en la producción final y por ende afecta en el rendimiento financiero de la empresa (Espinoza-Freire y Tinoco-Cuenca, 2015; Rodríguez & Fusco, 2017).

El banano (*Musa acuminata*) es el principal producto agropecuario de exportación para el Ecuador, el cual se ha ido perfeccionándose en el ámbito de los controles fitosanitarios y de la fertilización, con el objetivo de aumentar los rendimientos de la fruta (Marín-Usuga, Casamitjana, Loaiza-Usuga, & Loaiza-Usuga, 2016) el cual estuvo expuesto a un cambio extremo en la tendencia de crecimiento de los factores agronómicos de productividad total (PTF). En este marco, el desarrollar un modelo para la toma de decisiones concerniente al ciclo nutricional es de vital importancia (Valverde, García, Moreno, & Socorro, 2019).

El aprendizaje automático, nace de la inteligencia artificial como apoyo al entrenamiento de modelos que respondan a la predicción de los datos en diferentes ámbitos de la ciencia (Baviera, 2017). Su estructura requiere el análisis de diferentes algoritmos que permitan evaluar a un conjunto de datos a fin de establecer si el problema es de clasificación o de regresión (Navarrete-López, Calderón, Díaz Arévalo, Herrera & Izquierdo, 2017). Para la resolución de problemas de predicción se pretende a través de diversos algoritmos, entrenarlos con la finalidad de lograr que el modelo sea capaz de predecir el nuevo conjunto (Ramírez, Laguna & Anda, 2015).

La aplicación de aprendizaje automático emplea métodos supervisados y no supervisados, como apoyo a los procedimientos de análisis de los datos. La fase de predicción de los datos es el resultado del entrenamiento del modelo y el análisis de los hiperparámetros que permiten dar mayor rigurosidad a los datos (Goya, Barquero &

Figuera, 2017). Su aporte genera elementos suficientes para dar solución estadística a los problemas que requieran la utilización de estas técnicas.

En tal virtud se plantea como problema de investigación lo siguiente, ¿Qué algoritmo de aprendizaje automático permite predecir el comportamiento de los nutrientes del suelo en la producción de banano? En este marco el objetivo de la investigación fue de predecir el comportamiento de los nutrientes que afectan la fertilidad del suelo en la producción de banano mediante el uso de algoritmos de Aprendizaje Automático.

MATERIALES Y METODOS

Área de estudio

El estudio se realizó en una finca bananera de aproximadamente 20 hectáreas, ubicada en el municipio de Machala (coordenadas en UTM: 3215037; 79904698), a una altura de 6 msnm. La precipitación promedio fue de 505 mm y la temperatura media 26.5 °C. El suelo en la zona es de tipo franco a franco arenoso y pertenecen al orden Alfisol. La variedad de banano es Cavendish y los ciclos fitosanitarios y de manejo fertilidad son normales.

Dentro de la finca al azar fueron delimitadas cuatro parcelas de 1 ha, donde se recolectaron las muestras de suelo, las cuales fueron clasificadas en zona 1, zona 2, zona 3 y zona 4. En las cuales se logró extraer varias submuestras de 0 a 30 cm de profundidad de todo el terreno, mezclar y homogenizar para obtener una muestra compuesta, que fue llevada al laboratorio de suelos del INIAP, en Ecuador, donde se realizaron las siguientes determinaciones (Tabla 1).

Tabla 1. Listado de nutrientes y métodos utilizados para su determinación

Propiedades	Métodos
N (%)	Kjeldahl
P (mg/kg)	Olsen modificado
K (cmol(+)/g)	Espectrometría de absorción atómica
Ca (cmol(+)/g)	Espectrometría de absorción atómica
Mg (cmol(+)/g)	Espectrometría de absorción atómica
Fe (mg/kg)	Espectrometría de absorción atómica
Mn (mg/kg)	Espectrometría de absorción atómica
Cu (mg/kg)	Espectrometría de absorción atómica
Zn (mg/kg)	Espectrometría de absorción atómica

Fuente: Elaboración propia

Los resultados del análisis de suelos, las variables fueron ingresadas al software estadístico Weka versión 3.8 (Universidad de Waikato, 2019), el cual, a través de sus algoritmos, Redes Bayesianas, Árboles de decisión, Regresión logística, Vecino más cercano, Árboles aleatorios y Máquina de vectores de soporte, permitieron entrenar el modelo y establecer el algoritmo que logre clasificar a las parcelas en función de la disponibilidad de nutrientes.

Proceso estadístico

Los datos recolectados fueron tabulados en Excel y analizados su medias y desviación estándar. En la investigación se buscó que el algoritmo clasifique los elementos nutricionales, de acuerdo a los niveles descritos por Stoorvogel & Segura (2018). En tal virtud, se aplicó métodos supervisados para clasificar los datos, utilizando las siguientes técnicas, Redes Bayesianas, Árboles de decisión, Regresión logística, Vecino más cercano, Árboles aleatorios y Máquina de vectores de soporte.

El objetivo de la utilización del aprendizaje automático es de que el modelo logre predecir el comportamiento de los componentes nutricionales, para lo cual se lo clasificó en A suelo óptimo, B suelo aceptable y C suelo no aceptable. Una vez clasificado el tipo de suelo, se busca entrenar al modelo, para que sea capaz de reconocer en base a los nutrientes el tipo de suelo clasificado, procesando la data en el software seleccionado.

RESULTADOS Y DISCUSION

La Tabla 2, muestra el análisis descriptivo de los nutrientes. En general todos nutrientes se ubicaron en el rango óptimos sugeridos, indicando la mayor variabilidad el calcio (4.4 cmol(+)/g \pm 1.3) y el magnesio (2.3 cmol(+)/g \pm 0.8);

bases intercambiable sensible, cuando los niveles de potasio (0.35 cmol(+)/g \pm 0.03) son altos. Chávez-Velazco & Araya-Vargas (2009) y Barrezueta-Unda, Paz-Gonzalez & Chabla-Carrillo (2017), determinaron en suelos de bananeras y cacaoeras respectivamente de la provincia de El Oro rangos similares a los descriptos en la investigación.

Tabla 2. Análisis Descriptivo de las variables de estudio

Nutriente	Estadístico obtenido	Rango sugerido
Nitrógeno	0.4 (\pm 0.1)	> 0.31
Fósforo	15.6 (\pm 2.3)	11.0-20.0
Potasio	0.35 (\pm 0.03)	0.20-0.50
Calcio	4.4 (\pm 1.3)	3.00-6.00
Magnesio	2.3 (\pm 0.8)	1.00-3.00
Hierro	55 (\pm 4)	21-80
Manganeso	60 (\pm 4.8)	10-100
Cobre	15 (\pm 4)	3-20
Zinc	28 (\pm 5)	6-36

Valor entre paréntesis significa desviación estándar

Los algoritmos utilizados en el estudio se enlistan en la Tabla 3, los cuales se analizaron en función del estadístico Kappa y el error absoluto de la media. Los algoritmos Redes Bayesianas y Algoritmo de aprendizaje de perceptrón presentaron un estadístico Kappa de 0.84 y 0.96, respectivamente con error de la media entre 0.605 y 0.014. Los algoritmos Árboles de decisión, Regresión Logística y Bosques aleatorios, presentaron un Kappa 1 con error de la media 0; siendo árboles de decisiones y bosques aleatorios los que obtuvieron mayor significancia 0.00 y 0.01 respectivamente, motivo por el cual fue seleccionado para el entrenamiento.

Tabla 3. Resultados del análisis de los datos por Algoritmo

N°	Algoritmos	Estadístico Kappa	Error de la media	p-value (0.05)
1	Redes Bayesianas	0.84	0.605	0.04
2	Regresión Logística	1	0	0.01
3	Aprendizaje de perceptrón	0.96	0.014	0.03
4	Árbol de decisión	1	0	0.00
5	Bosques aleatorios	1	0	0.01
6	Vecino más cercano	1	0.007	0.02
7	Máquina de vectores de soporte	1	0.056	0.02

Luego de haber seleccionado y entrenado el modelo a utilizar, los datos del día siete fueron clasificados correctamente presentando las siguientes observaciones.

Tabla 4. Resultado de la clasificación con el algoritmo entrenado

Zona	% de Clasificación	OBSERVACIONES
Zona 1	100%	Clasificó en zonas A y B
Zona 2	100%	Clasificó en zonas A y B
Zona 3	100%	Clasificó en zonas A y B
Zona 4	100%	Clasificó en zonas A,B y C

En la Tabla 5, con los datos obtenidos de la zona 4, la cual presentó una clasificación que genera una disposición diferentes de los elementos nutricionales, se presenta la clasificación realizada con el algoritmo Árbol de decisión y Bosques Aleatorios, estableciendo el 8% de los datos se clasificados en suelo A, el 67% en suelo B y el 25% en suelo C, mientras que con Bosques aleatorios se obtuvo una clasificación del 10% en suelo A, el 69% en suelo B y el 21% en suelo C, lo que estableció a través de los dos algoritmos que se debe focalizar sobre esta zona actividades que permitan analizar los motivos que ocasionan esta clasificación considerando que los niveles identificados en zona de estudio están aceptables.

Tabla 5. Resultado de la clasificación de la zona 4

Algoritmo	% de Clasificación	A	B	C
Arboles de decisión	100%	8%	67%	25%
Bosques aleatorios	100%	8.33%	66.6%	25.07%

Para la presente investigación se determinó trabajar con el modelo de clasificación obtenido a través de los Arboles de decisión, sin embargo, para contrastar el estudio se entrenó a los datos con el algoritmo de los Bosques Aleatorios el cual presentó una clasificación similar, lo que establece que en la zona 4 existen alteraciones nutricionales que afecta al lote de estudio.

Los Bosques aleatorios como técnica de clasificación en métodos supervisados ha permitido realizar caracterizar las coberturas del suelo en la demarcación hidrográfica mediante minería de datos (Martín, Cánovas, Alonso, Gomariz, & Moreno, 2012). La técnica ha facilitado la clasificación en diferentes estudios relacionados a suelo. De la misma manera la técnica de árboles de decisión ha permitido clasificar adecuadamente la cartografía de suelo, siendo muy utilizado en este tipo de estudios (Blanco, López, Colditz, Hardtke, Mari & Fischer, 2012).

En este contexto se ha podido realizar estudios en la cual para el cultivo de soja se evaluaros tres algoritmos

de clasificación, identificando al de árboles de decisión y máxima verosimilitud como los más idóneos para realizar el mapeo de los datos durante la producción (Nolasco, Bocco, y Scavuzzo, 2015). Además se han usado algoritmos de clasificación, a través de árboles de decisión para identificar enfermedades como la Sigatoka negra, para identificar el patógeno en su diversos estadios (Freitez, Ablan, & Gómez, 2009). De la misma forma a través de series temporales se ha logrado generar modelos de predicción en el sector agropecuario resolviendo la problemática de encontrar mejores predicciones para las series de lluvias que se realizan en el sector (Rodríguez, 2016).

CONCLUSIONES

El algoritmo que clasificó más adecuado, según el estudio, fue el de Arboles de Decisión, el cual fue seleccionado a través de parámetros estadísticos, permitiendo establecer como a través de aprendizaje automático se generan nuevos procedimientos para la toma de decisiones en la producción de banano. El cual suministra información que facilita la predicción del comportamiento de los nutrientes del suelo, focalizando la zona que presentó variaciones en los nutrientes, lo cual suministra información importante para elevar la eficiencia y eficacia de la producción de banano.

REFERENCIAS BIBLIOGRÁFICAS

- Barrezueta-Unda, S., Paz-Gonzalez, A., & Chabla-Carrillo, J. (2017). Determinación de indicadores para calidad de suelos cultivados con cacao en provincia de El Oro. *Cumbres*.
- Baviera, T. (2017). Técnicas para el Análisis de Sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Revista DÍgitos*, 1(3), 33-50. <https://revistadigitos.com/index.php/digitos/article/view/74/39>
- Blanco, P., López, G., Colditz, R., Hardtke, L., Mari, N., & Fischer, A. (2012). Propuesta metodológica para la cartografía periódica de la cobertura del suelo en Latinoamérica y el Caribe: estado de situación y avances. *Revista de Teledetección*, 38, 65-70. <https://www.researchgate.net/publication/293256835>
- Chávez-Velazco, C., & Araya-Vargas, M. (2009). Correlación entre las características del suelo y los nematodos de las raíces del banano (Musa AAA) en Ecuador. *Agromía mesoamericana*, 20(2), 361-369.

- Espinoza-Freire, E., & Tinoco-Cuenca, N. (2015). La problemática ambiental resultante de la fumigación aérea con plaguicidas a bananeras de la provincia de El Oro Ecuador. *Ciencia en su PC*, (4). <https://www.redalyc.org/html/1813/181349354008/>
- Fajardo, J., Aguilar, L., Flores, M., Parra, D., & Acurio, C. (2017). Aspectos socio-económicos influyentes en el sector agrícola y su relación con la creación de empresas populares y solidarias de Quevedo. *Ciencia Sociales y Económicas*, 1(1), 107-116. <https://doi.org/10.18779/csye.v1i1.141>
- Fernández, A., Fernández, R., Rivera, C., & Calero, S. (2016). Desafíos de la gestión de producción agropecuaria en Cuba. *Agroalimentaria*, 22, 119-132. <http://www.redalyc.org/articulo.oa?id=199245407007>
- Freitez, J., Ablan, M., & Gómez, C. (2009). Propuesta de modelos predictivos del brote de la Sigatoka Negra para las plantaciones de plátano al sur del Lago de Maracaibo, Venezuela. *UDO Agrícola*, 9(1). <http://revistas.udo.edu.ve/index.php/udoagricola/article/view/3327>
- Goya, R., Barquero, O., & Figuera, C. (2017). Enseñanza del aprendizaje automático utilizando las competencias de Kaggle. *Nuevos enfoques en la Innovación Docente Universitaria, 2017, ISBN 978-84-697-6817-4, págs. 66-70*, 66-70. <https://dialnet.unirioja.es/servlet/articulo?codigo=6275344>
- Marín-Usuga, M., Casamitjana, M., Loaiza-Usuga, J., & Loaiza-Usuga, J. C. (2016). Políticas de modernización agropecuaria y desarrollo rural en Colombia (1996-2008). *Revista EIA*, 13(25), 99-117. <https://doi.org/10.24050/REIA.V13I25.1020>
- Martín, J., Cánovas, F., Alonso, F., Gomariz, F. J., & Moreno, J. (2012). *Clasificación de coberturas del suelo en la Demarcación Hidrográfica del Segura mediante técnicas de minería de datos*. https://www.researchgate.net/profile/Fulgencio_Canovas-Garcia/publication/273144096_Clasificacion_de_coberturas_del_suelo_en_la_Demarcacion_Hidrografica_del_Segura_mediante_tecnicas_de_mineria_de_datos/links/54fa3c170cf20b0d2cb63793.pdf
- Moreno-Carriles, R. M. (2018). Big data, ¿pero qué es? *Angiología*, 70(5), 191-194. <https://doi.org/10.1016/J.ANGIO.2018.05.001>
- Navarrete-López, C., Calderón, D., Díaz Arévalo, J., Herrera, M., & Izquierdo, J. (2017). Modelos geostatísticos para la predicción de fallos de una zona de la red de abastecimiento de agua de Bogott, integrando algoritmos de Machine Learning. *Ibero-American Seminar on Water and Drainage Networks*. <https://doi.org/10.2139/ssrn.3113048>
- Nolasco, M., Bocco, M., & Scavuzzo, C. (2015). Mapeo de cultivos agrícolas en Córdoba: una comparación del desempeño de tres algoritmos de clasificación. *Congreso Argentino de AgroInformática*, 1, 58. http://sedici.unlp.edu.ar/bitstream/handle/10915/62830/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Ramírez, N., Laguna, M., & Anda, A. (2015). Aplicación de Métodos de Inteligencia Artificial en el Área Médica. *Pistas educativas*, 35, 438. Recuperado a partir de <http://itcelaya.edu.mx/ojs/index.php/pistas/article/view/361>
- Riba, A., et al. (2018). Data-driven Agriculture, las tecnologías aplicadas a problemas de agricultura regional. *SEDICI*. <http://sedici.unlp.edu.ar/handle/10915/68088>
- Rodríguez, C. (2016). *Modelos no lineales de pronóstico de series temporales basados en inteligencia computacional para soporte en la toma de decisiones agrícolas*. Universidad Nacional de Córdoba. Facultad de Ciencias Exactas, Físicas y Naturales. <https://rdu.unc.edu.ar/handle/11086/4604>
- Rodríguez, D., & Fusco, M. (2017). Gestión de riesgos agropecuarios en el sector del cacao en Ecuador. *Revista de Investigación en Modelos Financieros*, 6, 57-74. <http://www.economicas.uba.ar/wp-content/uploads/2016/02/Rodriguez-D.-Fusco-M.-Gestion-de-riesgos-agropecuarios-en-el-sector-del-cacao-en-Ecuador.pdf>
- Stoorvogel, J., & Segura, R. (2018). Nutrition and soil management in banana cultivation. En G. Kema & A. Drentj (Eds.), *Achieving sustainable cultivation of bananas* (Vol. 1, pp. 1-15). Cambridge, UK: Burleigh Dodds Science Publishing.
- Valverde, E., García, R., Moreno, A., & Socorro, A. (2019). Alternativas nutricionales eficientes en banano orgánico en la provincia El Oro, Ecuador. *Revista Metropolitana de Ciencias Aplicadas*, 2(1), 151-159. <http://remca.umet.edu.ec/index.php/REMCA/article/view/104>