

Modelo predictivo para la medición del estado de madurez de *Musáceas* sp., cultivadas al suroeste del Ecuador, usando algoritmos de Machine Learning

Predictive model to measure the degree of ripeness of Musacea sp. cultivated in Southwest Ecuador through Machine learning algorithms

DOI: <https://doi.org/10.33262/rmc.v7i2.2809>

Omar Martínez-Mora

Universidad Nacional de Tumbes-Perú

Oscar Capuñay-Uceda

Universidad Nacional Pedro Ruiz Gallo – Perú

Luis Caucha-Morales

Universidad Nacional de Tumbes-Perú

María Bernarda Ruilova-Cueva

Universidad Técnica de Babahoyo-Ecuador

DIRECCIÓN PARA CORRESPONDENCIA: emartinez@utmachala.edu.ec

Fecha de recepción: 08/01/2022

Fecha de aceptación: 25/03/2022

RESUMEN

El objetivo de la presente investigación fue, definir un modelo predictivo para la medición del estado de madurez de *Musáceas* sp., cultivadas al suroeste del Ecuador, usando algoritmos de Machine Learning. La metodología utilizada en la presente investigación, se dividió en dos etapas: la primera etapa, consistió en obtener los datos de las características fisicoquímicas, a través de los grados Brix y acidez titulable (AOAC,2009); la segunda etapa, mediante el uso de algoritmos de Machine Learning, se procesaron dos sets de datos, tanto para el entrenamiento, como para el Test. El presente trabajo llegó a la conclusión de que, se puede predecir el estado inmaduro, maduro o sobre maduro del banano, a través de un modelo predictivo, con una exactitud del 91 %.

PALABRAS CLAVE: *Musáceas*, modelo predictivo, medición del estado.

ABSTRACT

The objective of the present investigation was to define a predictive model for the measurement of the state of maturity of *Musáceas* sp., cultivated in the southwest of Ecuador, using Machine Learning algorithms. The methodology used in the present investigation was divided into two stages: the first stage consisted of obtaining data on the physicochemical characteristics, through Brix degrees and titratable acidity (AOAC, 2009); The second stage, through the use of Machine Learning algorithms, two sets of data were processed, both for training and for the Test. The present work concluded that the immature, mature or overripe state of bananas can be predicted, through a predictive model, with an accuracy of 91 %.

KEYWORDS: *Musáceas*, predictive model, state measurement.

INTRODUCCIÓN

Desde hace algunos años, la Agroindustria ha tomado en cuenta el uso de nuevas tecnologías, destinadas al análisis de los alimentos y, dentro de ellas, se encuentra la Inteligencia Artificial (IA) y las Redes Neuronales Artificiales (RNA), las que permiten desarrollar controles exhaustivos de las condiciones nutricionales y funcionales de los alimentos, contribuyendo a la oferta de productos estables y aceptables por los consumidores. Además, estas nuevas tecnologías ayudan a simplificar los procesos de análisis que, con instrumentación de laboratorio convencionales, serían algo tediosos. Se han desarrollado investigaciones con resultados muy alentadores, dentro de las cuales están, el uso de RNA para modelar el proceso de secado de cereales (Farkas *et al.*, 2000); lesiones de melocotones en líneas de proceso (Bielza *et al.*, 2003), control de recepción de uvas (Roussel *et al.*, 2003); clasificación de tamaños de manzanas (Shahin *et al.*, 2001), mediante el uso de biosensores, basados en amperometría. Periasamy *et al.* (2011) y Romero (2016), lograron determinar glucosa y capacidad antioxidante en vino, respectivamente.

A nivel global, Ecuador es el primer productor y exportador de banano. En el año 2019 se produjeron 6,583,477 TM (CFN, 2020). Esta *Musácea* es muy consumida por la población ecuatoriana por su bajo costo, nivel nutricional y atributos sensoriales cuando alcanza su madurez, cambio bioquímico detectable al formarse la fructosa, además, se

producen compuestos aromáticos como, por ejemplo, los esterres (Beaudry, *et al.*, 1989). Así mismo, alcanzada la madurez en las *musáceas*, la corteza pierde su color verde, para dar paso al color amarillo, debido a pérdida de la clorofila, resultado de la migración de humedad hacia la zona comestible de la fruta. La aparición de puntos oscuros en las cortezas amarillas de los bananos, es un indicador del inicio de la sobre maduración de los mismos. Además, el banano para ser considerado como maduro, su índice de madurez debe promediar los 64 (Torres *et al.*, 2013). Este índice relaciona el contenido en sólidos solubles, con rango de entre los 20 y 22 grados Brix (Yanes *et al.*, 2020); y la acidez titulable, con valores promedio de 0,28 g/mL (ácido málico). En base a 100 gramos de pulpa de banano, en estado fisiológico de madurez, su composición química proximal (miligramos) en potasio es de 500; magnesio 35; fósforo 30 y calcio 3. Entre los macro componentes (gramos), el agua se presenta con 65, hidratos de carbono 32, fibra dietaria 2, proteína 1,3; y en lípidos totales, 0,37 (Casallas, L. 2010). Valores que demuestran el valor nutricional de esta fruta.

Una de las disciplinas de las ciencias informáticas, relacionada con el desarrollo de la Inteligencia Artificial (IA), es la denominada “Aprendizaje de Máquina” (*Machine Learning - ML*). Bengio (2009) define al ML como, “El conjunto de algoritmos ordenados de tal forma que, simplifiquen tareas complejas”. Además, Mucherino *et al.*, (2009) sostienen que, al escoger un algoritmo, éste deberá realizar estimaciones, sin diferencias significativas, con los resultados de una nueva base de datos de la misma naturaleza, además, estos algoritmos pueden ser supervisados y no supervisados. El primero, es cuando se conoce la información de salida y el entrenamiento requerido; y el segundo, es cuando se desconoce la información de salida, para ello se aconseja echar mano de las técnicas de agrupamiento no supervisadas. Cuando se habla de entrenamiento, los datos pueden provenir de expertos o por instrumentos de alta precisión. Sea cual fuere la fuente, los datos pueden ser divididos en dos grupos, unos para entrenamiento y otros para el *Test*. Es de gran ayuda, para el modelo que se pretenda obtener, utilizar un tercer conjunto de datos, lo que optimaría la certeza y confiabilidad (Refaeilzadeh *et al.*, 2009). Por tal motivo, el objetivo de la presente investigación fue, definir un modelo predictivo para la medición del estado de madurez de *Musáceas sp.*, cultivadas al suroeste del Ecuador, usando algoritmos de *Machine Learning*.

METODOLOGIA

La metodología utilizada, se dividió en dos etapas: Análisis de laboratorio y uso de algoritmos de *Machine Learning*

1. *Análisis de laboratorio*. Las muestras de banano de la variedad Cavendish, fueron seleccionadas en estado inmaduro, según lo establecido por Moreno (2009).

-*Sólidos solubles*, se midió directamente de la pulpa de la fruta, previamente macerada. Se utilizó un refractómetro digital de la marca Mettler Toledo-30 PX, siguiendo el método AOAC 932.12 (AOAC, 1990). Los valores se reportaron en grados Brix

-*Acidez titulable*, mediante el método AOAC 942.05 (AOAC, 1990). La ecuación utilizada fue:

$$\text{Acidez \%} = \frac{A * B * C}{D} * 100$$

donde:

A es la cantidad en mililitros de hidróxido de sodio (NaOH), gastados.

B es la normalidad de la base.

C es el peso equivalente, expresado en gramos de ácido málico y,

D es el peso de la pulpa de banano, expresado en gramos. La de acidez se determinó como porcentaje de ácido málico (ácido dominante).

2. *Algoritmos de Machine Learning*, la figura 1 resume las etapas desde la adquisición de los datos físico – químicos, provenientes de los análisis de laboratorio de las muestras de banano, hasta la evaluación del modelo predictivo.

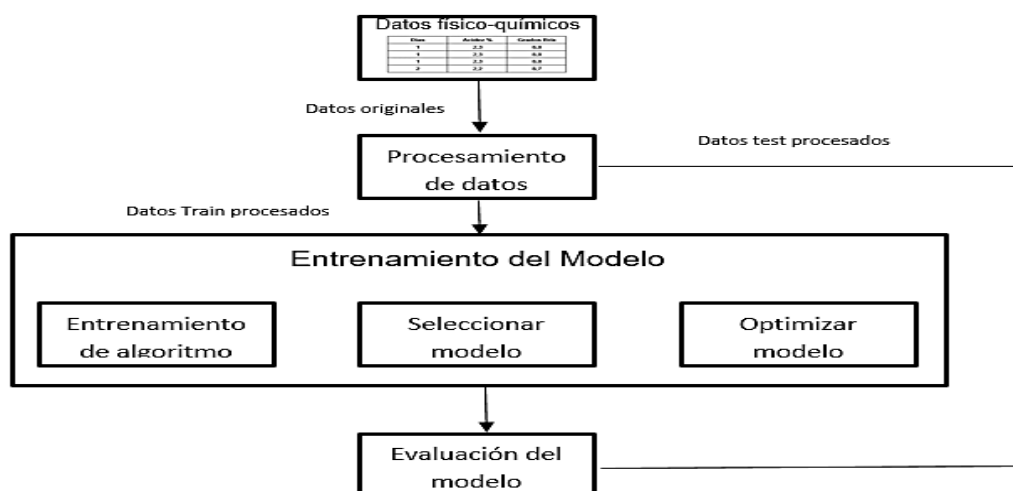


Figura 1: Etapas del uso de los algoritmos *ML*.

-*Adquisición de datos físico-químicos.* Se trabajó en los laboratorios de Análisis de Alimentos, de la Facultad de Ciencias Químicas y de la Salud, de la Universidad Técnica de Machala - Ecuador. Los sólidos solubles y la acidez titulable, se midieron a las muestras de banano, entre los días en que su cáscara presentó un color verde, pasando por el amarillo, hasta la presencia de manchas oscuras; estadios en los que se consideran al banano como inmaduro, maduro y, en estado de sobre maduración, respectivamente.

-*Procesado de los datos.* Se ordenó la información en tres variables: la primera, que tiene que ver con los sólidos solubles; la segunda, acidez titulable y, la tercera, con la decisión del estadio madurativo. Se repartieron dos conjuntos de datos, tanto para el entrenamiento y para el *Test*.

-*Entrenamiento del modelo.* Con el conjunto de datos físico- químicos, obtenidos en el laboratorio, se entrenaron los algoritmos de Clasificación de *Machine Learning*: *k - Nearest Neighbor (k-NN)*, *Naïve Bayes*, *Logistic Regression*, *Support Vector Machines (SVMs)*, *Decision Trees*, *Random Forests* y *Neural Networks*. Se buscó los mejores parámetros de los algoritmos y se evaluó la exactitud y error de cada modelo. Los algoritmos utilizados provienen de la librería Scikit-Learn de *Python* para *Machine Learning*, la cual se complementa con; *Pandas (pd)*, *Numpy (np)* y *Matplotlib.pyplot (plt)*.

-*Evaluación del Modelo.* Los datos físico – químicos de las muestras de banano, tomados a lo largo de su maduración, se relacionaron con el Índice de Madurez, indicador que está en función de los sólidos solubles y acidez de una fruta. Se utilizaron métricas de Exactitud, *Test Accuracy*, y de error, utilizando las siguientes ecuaciones:

$$Test Accuracy = \frac{IM + M + SM}{Total}$$

Donde:

IM, es el número de las observaciones de los bananos inmaduros

M, es el número de las observaciones de los bananos maduros

SM, es el número de las observaciones de los bananos sobre maduros

Total, es el número total de observaciones

Tasa de error:

$$Tasa\ de\ error = 1 - Test\ Accuracy$$

RESULTADOS Y DISCUSION

Las condiciones ambientales, a las cuales fueron tomados los datos, tuvieron una temperatura promedio de 26 °C, humedad relativa de 87 % y una altura de 6 msnm.

La tabla 1 y 2 muestran las características físico – químicas de la pulpa del banano de la variedad *Cavendish*, tanto de entrenamiento como de *Test*, respectivamente.

Tabla 1. Datos fisicoquímicos de entrenamiento.

Parámetro fisicoquímico	Días después de la cosecha										
	2	4	6	8	10	12	14	16	18	20	22
Acidez %	0,10 (0,03)	0,20 (0,04)	0,25 (0,01)	0,30 (0,03)	0,33 (0,01)	0,39 (0,04)	0,40 (0,02)	0,43 (0,02)	0,45 (0,03)	0,48 (0,02)	0,50 (0,01)
Solidos solubles °Brix	2 (0,01)	5 (0,01)	7 (0,02)	9 (0,01)	11 (0,03)	14 (0,02)	15 (0,01)	17 (0,01)	20 (0,03)	25 (0,04)	27 (0,01)
Índice madurez	20	25	28	30	33	36	38	40	47	52	54
Estado	IM	IM	IM	IM	IM	MA	MA	MA	MA	SM	SM

Los valores reportados son la media de 3 determinaciones y su desviación estándar en paréntesis. IM, Inmaduro; MA, Maduro; SM, Sobre maduro.

Tabla 2. Datos fisicoquímicos de *Test*.

Parámetro fisicoquímico	Días después de la cosecha										
	2	4	6	8	10	12	14	16	18	20	22
Acidez %	0,12 (0,02)	0,24 (0,01)	0,25 (0,03)	0,28 (0,03)	0,37 (0,01)	0,38 (0,05)	0,42 (0,02)	0,44 (0,07)	0,47 (0,02)	0,48 (0,02)	0,52 (0,05)
Solidos solubles °Brix	1 (0,03)	4 (0,02)	6 (0,02)	10 (0,01)	14 (0,05)	16 (0,02)	18 (0,07)	20 (0,01)	23 (0,07)	26 (0,09)	29 (0,04)
Índice madurez	8	17	24	35	37	42	43	45	49	54	56
Estado	IM	IM	IM	IM	MA	MA	MA	MA	MA	SM	SM

Los valores reportados son la media de 3 determinaciones y su desviación estándar en paréntesis. IM, Inmaduro; MA, Maduro; SM, Sobre maduro.

En la tabla 3 se definen las variables de entrada y salida, así como también, la variable de respuesta, información necesaria para la ejecución de los algoritmos.

Tabla 3. Variables de entrada, salida y de respuesta

Variable	Definición
xtrain	Parámetros fisicoquímicos de entrenamiento (Acidez, Brix y IM)
xtest	Parámetros fisicoquímicos en el <i>Test</i> (Acidez, Brix y IM)
ytrain	Estado de madurez del banano (IM, MA y SM) en entrenamiento
ytest	Estado de madurez del banano (IM, MA y SM) en el <i>Test</i>

A continuación, se presentan las Matrices de Confusión de todos los algoritmos seleccionados, así como su *Test Accuracy*, y tasa de error.

Algoritmo *k - Nearest Neighbor* (k-NN)

Test Accuracy = 0,91

Tasa de error = 0,99

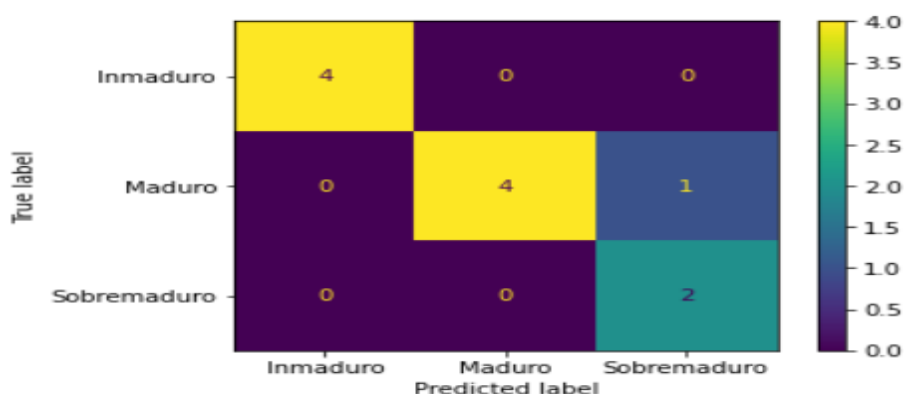


Figura 1. Matriz de confusión del algoritmo “k-NN”

Algoritmo Naive Vayes,

Test Accuracy = 1,0

Tasa de error = 0

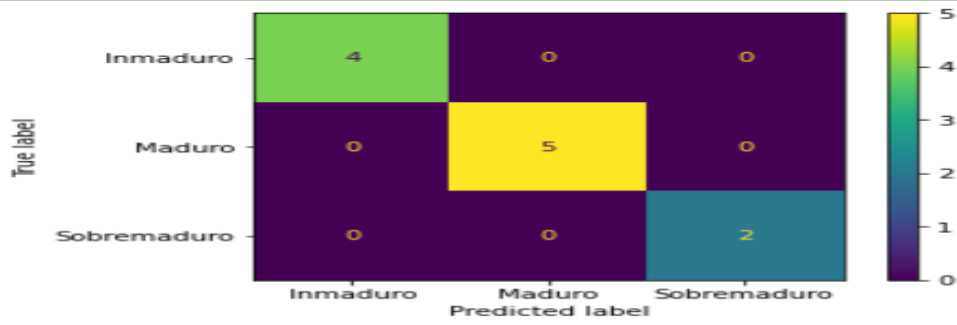


Figura 2. Matriz de confusión del algoritmo “Naive Vayes”

Algoritmo Logit

Test Accuracy = 0,91

Tasa de error = 0,99

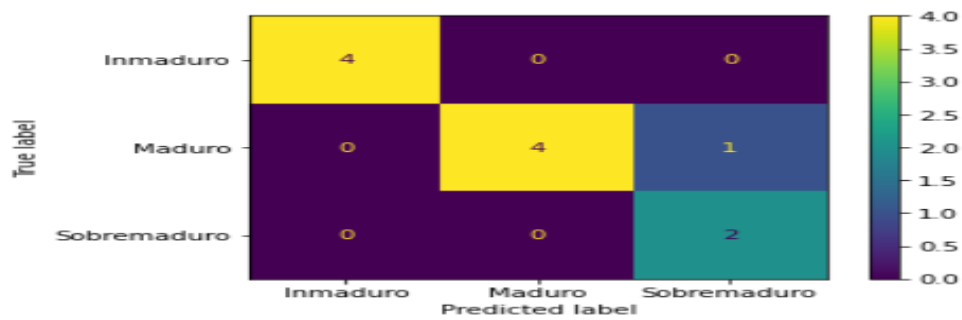


Figura 3. Matriz de confusión del algoritmo “Logit”

A continuación, en la tabla 3 se resume los *Test Accuracy*, así como la tasa de error de los algoritmos restantes

Tabla 3. *Test Accuracy*

Algoritmo	<i>Test Accuracy</i> %	Tasa de error %
<i>Support Vector Machines (SVMs)</i>	91,0	9,0
<i>Random Forests</i>	91,0	9,0
<i>Neural Networks</i>	91,0	9,0
<i>Decision Tree</i>	100,0	0,0

A continuación, en la figura 4 se muestra el árbol de decisión

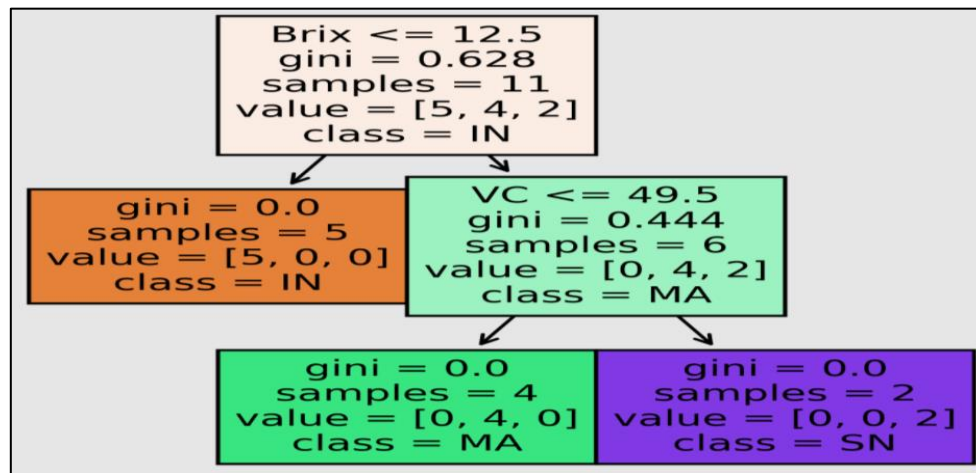


Figura 4. Árbol de decisión

DISCUSIÓN

Las condiciones ambientales como: temperatura, humedad relativa y altitud, entre las principales, afectan el tiempo de maduración de las *Musáceas*. Humedades relativas y temperaturas bajas, retardan la maduración de esta fruta. Investigaciones sobre la predicción de maduración de bananos, utilizando la Inteligencias Artificial, han logrado establecer porcentajes de exactitud mayores al presente trabajo, es decir, un 96 % (Villalba A., *et al.* 2019), así mismo, se han reportado valores muy inferiores, 65 % (Aguilar y Campoverde, 2020). Es importante tener en cuenta la temporada de recolección de muestras, es decir, si es época seca o lluviosa, debido a que los procesos bioquímicos de transformación de hidratos de carbono no edulcorantes a edulcorantes, están estrechamente relacionados con las condiciones ambientales, que aceleran o retardan la maduración, sumado a esto, el escogimiento de las muestras de banano, deberán ser recolectadas de un mismo racimo, así como también, que sus unidades estén al mismo nivel de su “*cluster*”, debido a que, bananos de diferentes ubicaciones en un mismo racimo, su maduración no es homogéneo y, ello va a incidir, en la exactitud del modelo de predicción.

CONCLUSIÓN

Esta investigación demostró que, la madurez del banano cultivado al suroeste del Ecuador, puede predecirse como inmaduro, maduro o sobre maduro, a través de un modelo predictivo usando algoritmos de *Machine Learning*, con un promedio de exactitud del 91 %.

BIBLIOGRAFIA

- AOAC (Association Of Official Analytical Chemists). (1990). Official methods of analysis of the association of official analytical chemists. 11a edition. Washington.
- Beaudry, R. M., Severson, R. F., Black, C. C., & Kays, S. J. (1989). Banana ripening: implications of changes in glycolytic intermediate concentrations, glycolytic and gluconeogenic carbon flux, and fructose 2, 6-bisphosphate concentration. *Plant physiology*, 91(4), 1436-1444.
- Bengio Y. (2009). Learning Deep Architectures for AI. *Foundation and Trends in Machine Learning*, 2(1), 1-27.
- Bielza, C., Barreiro, P., Rodriguez-Galiano, M., & Martin, J. (2003). Logistic regression for simulating damage occurrence on a fruit grading line. *Computers and Electronics in Agriculture*, 39(2), 95-113.
- Casallas Malaver, L. F. (2010). Evaluación del análisis fisicoquímico del banano común (*musa sapientum l*) transformado por acción de la levadura *Candida guilliermondii*. Grado de Microbióloga industrial y Nutricionista dietista. Pontificia Universidad Javeriana – Colombia.
- Corporación Financiera Nacional. (20 noviembre 2020). Ficha sectorial banano y plátano. Recuperado el 15 de enero 2022 de <https://www.cfn.fin.ec/wp-content/uploads/downloads/biblioteca/2020/ficha-sectorial-4-trimestre-2020/FS-banano-4T2020.pdf>.
- Farkas, I., Remenyi, P., & Biró, A. (2000). A neural network topology for modelling grain drying. *Computers and Electronics in Agriculture*, 26(2), 147-158.
- Jonathan Victor Aguilar-Alvarado; Milton Alfredo Campoverde-Molina (2020). Clasificación de frutas basadas en redes neuronales convolucionales. *Pol. Con.* 41 (5).
- MUCHERINO, A., P. J. PAPAJORGJI, & P. M. PARDALOS. (2009). *Data Mining in Agriculture*. Vol. 34, Springer Optimization and Its Applications. Edited by Panos M. Pardalos. New York, NY: Springer Özsu. 532-538. New York: Springer US.

- Periasamy, A.P., Chang, Y.-J., & Chen, S.-M (2011). Amperometric glucose base don glucose oxidase immobilized on gelatin-multiwalled carbón nanotube modief glassy carbón electrode. *Bioelectrochemistry*, 80(2), 114-120.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, (5), 532-538.
- Romero G. (2016). Detección de vinos mediante una red de Biosensores de Pirrol y Nanopartículas de Oro. Grado de Ingeniería Mecánica. Universidad de Valladolid – España.
- Roussel, S., Bellon-Maurel, V., Roger, J.-M., & Grenier, P. (2003). Authenticating White grape must variety with classification models base don aroma sensors, FT-IR and UV spectrometry. *Journal of food engineering*, 60(4), 407-419.
- Shahin, M.A., Tollner, E. W., & McClendon, R.W. (2001). AE- Automation and Emergen Technologies: Artificial Intelligence Classifiers for sorting Apples base don Watercore. *Journal of Agricultural Engineering Reseach*, 79 (3), 265-274.
- Torres, R., Montes, E. J., Pérez, O. A., & Andrade, R. D. (2013). Relación del color y del estado de madurez con las propiedades fisicoquímicas de frutas tropicales. *Información tecnológica*, 24(3), 51-56.
- Villalba, A., Requena, T., Solanilla, F., y Rangel, J. C. (2020). Prototipo de un sistema que determine el estado de madurez de un plátano utilizando Deep Learning y Visión Artificial. *Revista de Iniciación Científica*, (6). 3p.
- Yáñez Bustamante, W. D., Quevedo Guerrero, J. N., García Batista, R. M., Herrera Reyes, S. N., y Luna Romero, Á. E. (2020). Determinación de la relación carga química grados brix en hojas y frutos de banano clon Williams (*Musa paradisiaca*). *Revista Universidad y Sociedad*, 12(5), 421-430.