DISEÑO DE UN MODELO DE PREDICCIÓN PARA DETECTAR SIGATOKA (MYCOSPHAERELLA) EN LAS PLANTACIONES DE BANANO

AUTOR (A): MILTON R. QUIÑONEZ P.

CO AUTOR: MSC. PEDRO F. ECHEVERRIA B

UNIVERSIDAD ECOTEC

FACULTAD DE POSGRADO, CARRERA DE SISTEMAS

.

RESUMEN

Debido la hipótesis planteada de la investigación en la cual se integra el sector agrícola y el sector tecnológico, se han implementado miles de proyectos o aplicaciones desarrolladas con el fin de mejorar la producción de la agricultura, la investigación fue realizada en una empresa bananera ubicada en Ecuador La Provincia de Los Ríos Babahoyo, conocida como la más grande a nivel nacional en base a experiencias y conocimiento surgió la idea de poder implementar un proyecto que me permita identificar la sigatoka negra o amarilla para proponer una solución enfocada al modelos de negocio y los procesos de deshoje y fertilización siendo una labor muy demandada en este sector, se realizó un estudio de campo para la toma de muestras de imágenes para realizar un procesamiento de clasificación de sigatoka, la cual dio resultados favorables en cada proceso de la metodología de extracción y procesamiento de datos ayudando a tener un orden de aprendizaje y de evaluación de resultados.

**PALABRAS CLAVES:** Aprendizaje profundo, sigatoka, redes neuronales, Aprendizaje automático

**INTRODUCCIÓN**

En la actualidad existen muchas empresas productoras de banano a nivel nacional e internacional, “el Ecuador tiene alrededor de 162.236 hectáreas sembradas de banano y cuenta con 4.473 productores de la fruta“ (Ministerio de Comercio Exterior, 2017)

Tomando como ejemplo la Provincia de Los Ríos existe un aproximado de 50.000 hectáreas de banano lo que significa en el país una producción del 31% de esta distribución tomaremos como base la Hacienda COOPROCLEM donde “se localiza geográficamente en la provincia de Los Ríos, ciudad de Babahoyo, específicamente en la parroquia rural La Unión; es la hacienda más grande del país teniendo a su haber 12 mil hectáreas de tierra fértil de las cuales el 21% de las mismas (2500 has) son utilizadas exclusivamente para la producción de banano de exportación contribuyendo así de considerable manera a la economía nacional y siendo una de las mayores fuentes de empleo de la localidad.” (Jordano, 2019)

Esta información nos sirve para poder tener en cuenta el impacto que puede llegar a tener una aplicación o diseño de aplicación para clasificar la sigatoka utilizando machine learning, dentro del mundo bananero existen muchas variantes donde se podría aplicar el machine learning uno de ellos puede ser la merma o fruta rechazada.

Existen muchas aplicaciones móviles y web que se utilizan en el área de banano, tanto como un control de calidad y un registro de inventario de producción, además existen variantes de programas para controlar el avance de aplicación de fertilizantes entre otros. Un ejemplo práctico es un sistema de producción agrícola de banano para el control de inventario y procesos de empaque, utilizando como un servicio adicional una aplicación móvil para el registro de merma registrado por un usuario asignado, este control funciona capturando una foto del daño en la fruta describiendo o seleccionando en la aplicación una vez identificados los ítems de registro se procede a guardar este proceso puede tiene dos variantes área de racimos y área de cajas.

El modelo detectará mediante video o imágenes el nivel de sigatoka que tiene un área y el tipo de sigatoka (negra o amarilla), esto para poder realizar un control más detallado de la aplicación de fertilizantes y mejorar las labores de deshoje como tal que son básicas en el control de esta infección de la planta.

El control será de vital importancia durante las etapas de cada ciclo antes de aplicar fertilizantes, cabe recalcar que el proceso es una muestra de la infección de sigatoka, y verificación de la misma para de esta manera enfocarse en la corrección de la planta infectada y tomar medidas ya sea en labores de deshoje (lamineo, despunte, corte completo) o aplicar mayor herbicida en base a la cantidad que se pueda mostrar en el modelo de predicción.

Como base en tecnología nos enfocaremos en la recolección de información necesaria para realizar un entrenamiento de datos efectivo y obtener una base en el cual los ingresos se vayan incrementando automáticamente.

MATERIALES Y MÉTODOS

Las metodologías son de mucha relevancia al momento de realizar una investigación, proyecto u ensayo de cualquier índole, ya que es la principal base de adquisición de información, en esta investigación de que tiene como principal objetivo describir el nivel de sigatoka en las plantaciones de banano mediante un modelo de predicción basado en datos de entrenamiento (imágenes capturadas y clasificadas).

Una de las principales características del modelo de predicción es que se basa en datos que describen el objeto a analizar, estos datos pueden ser históricos y actuales, partiendo de esta manera en este capítulo se describirá lo siguiente; la metodología de investigación, el tipo de investigación y su diseño, contexto (lugar o áreas de obtención de datos) y también los instrumentos y técnicas de investigación en el desarrollo del modelo de predicción.

Metodología Descriptiva

La investigación descriptiva responde a las preguntas: ¿cómo son?, ¿dónde están?, ¿cuántos son?, ¿quiénes son?, etc..; es decir nos dice y refiere sobre las características, cualidades internas y externas, propiedades y rasgos esenciales de los hechos y fenómenos de la realidad en un momento y tiempo histórico concreto y determinado. (Castillo, 2020). También se puede definir como una manera de interpretar la situación actual y permite correlacionar variables investigativas.

Otra definición de la investigación descriptiva según (Castillo, 2020) define que “*Se refiere a aquella orientación que se centra en responder la pregunta acerca de cómo es una determinada parte de la realidad objeto de estudio*”.

Para nuestro proyecto del modelo de predicciones trabajaremos con la investigación antes mencionada y utilizaremos la técnica de observación donde (Guzmán, 2019) da la siguiente definición “es una técnica de investigación que consiste en recolectar datos e información con ayuda de los sentidos para analizar los hechos, realidades sociales y personas en su contexto real.”

En la recolección de información que implica nuestra investigación se desarrollara en las plantaciones de banano de la Provincia de Los Ríos, Babahoyo Hacienda COOPROCLEM siendo esta la hacienda más grande a nivel nacional con una extensión de 12 mil hectáreas en las que se cultiva limón, café, pastizales para la ganadería y teca. La actividad principal es la producción de banano con un total de 2671,51 hectáreas. (Cooproclem, 2021).

Metodología Correlacional

La metodología correlacional se define con una forma de interpretar dos variables para determinar el resultado de una investigación, también se puede definir por (Olleta, 2021) como “*Cuando una teoría científica postula la existencia de una correlación entre dos*[*variables*](https://www.e-torredebabel.com/Psicologia/Vocabulario/Variable.htm)*que no pueden ser analizadas*[*experimentalmente*](https://www.e-torredebabel.com/Psicologia/Vocabulario/Verificacion-Experimental.htm)*, su presencia en una población o grupo puede ser detectada con el recurso del método de verificación correlacional*”. De esta forma podemos definir las variables a relacionarse y clasificar por dependiente e independiente.

**La Variable Dependiente**

**Categoría de Sigatoka. -** Es la base de nuestro proyecto que se desea realizar por un método de observación de campo, siendo esta una de las principales enfermedades de infección en la plantación de banano.

Para el desarrollo del proyecto se contará con una fuente de información en formato imágenes clasificadas de la siguiente manera:

* ENVEZ con sigatoka
* ENVEZ sin sigatoka
* HAZ con sigatoka
* HAZ sin sigatoka

Esta clasificación me permitirá conocer de manera eficaz en base a la información que será obtenida de la Hacienda Clementina Cooproclem. Las imágenes obtenidas, deberán pasar por tres etapas descritas a continuación:

* Preprocesamiento
* Entrenamiento
* Prueba

**La Variable Independiente**

**Pixeles de las imágenes de una fotografía. -**  Es la manera en cómo el modelo de predicción detectara si tiene o no sigatoka y que clasificación le otorgara, esto mediante un reconocimiento de patrones previo al preprocesamiento y entrenamiento de la data.

Metodología CRISP-DM

Existen tres metodologías dominantes para el proceso de minería de datos son KDD, SESMA Y CRISP-DM siendo esta última con la cual se implementará el proyecto de investigación, dentro de cada metodología se diferencias por las fases o procesos a seguir para la minería de datos. Según el autor (Jesús, 2019) define el modelo como “*un estándar empleado a nivel mundial tanto en la industria como en la academia para proyectos de minería de datos*”

CRISP-DM (Cross Industry Standard Process for Data Mining) inicia en Sept. De 1996 Financiado por la Comisión Europea. Esta metodología identifica las diferentes fases de la implementación de un proyecto de data mining o minería de datos, siendo aplicable a cualquier industria, y analizar cualquier tipo de datos (comerciales, financieros, de recursos humanos, producción industrial, producción agrícola, etc.). Esta metodología es también muy 15 flexible y adaptable para muchos estudios, esta divide el proceso de minería de datos en seis fases principales las cuales son: Comprensión del negocio, Comprensión de Datos, Preparación de datos, Modelado, Evaluación y Despliegue de las cuales se muestra en la Ilustración 1. (GUACHAMÍN, 2020).

Fases de la metodología CRISP-DM

1. Entendimiento del Negocio
2. Entendimiento de los datos
3. Preparación de los datos
4. Modelamiento
5. Evaluación
6. Despliegue

Resultados y Discusiones

Resultados de negocio

Para comprender este resultado hay que tener claro el modelo del negocio al cual se está aplicando, existen negocios aplicados a la producción como aplicados a la fertilización prestando servicios al sector bananero u otro sector agrícola, la investigación parte de la idea de mejorar el proceso de clasificación de imágenes para detectar el nivel de sigatoka que hay en una plantación de banano, este se divide en dos que son; Sigatoka negra y Sigatoka Amarill.

También se puede observar que el modelo fue probado en una Hacienda de Ecuador ubicado en la Provincia de Los Ríos – Babahoyo – Clementina, existe un mapa y detalles de la superficie de la hacienda donde se encontraran más información de la metodología de recopilación de información.

Resultados de Obtención de datos

Esta muestra se basa principalmente en la captura de los datos y clasificación de los mismos mediante un proceso de aprendizaje para realizar una distribución más efectiva de los datos, podemos indicar que la obtención fue clasificada en 4 clases y adquirimos lo siguiente:

* Imágenes de entrenamiento: 539
* Imágenes de prueba: 53

En el entendimiento de las imágenes también utilizamos herramientas que nos permita obtener mejores detalles de nuestra extracción de datos, donde se puede comprender con imágenes de cada clase como se distribuye la cantidad de pixeles y la combinación de colores RGB para la lectura posterior de los datos.

Resultados Aplicación tecnológica

Esta muestra un listado de las diferentes tecnologías y herramientas aplicada para la minería de datos y generación del resultado final, utilizamos tecnología basada en Machine Learning y Deep Learning trabajando principalmente con redes neuronales convolucionales.

Como herramientas tecnológicas de desarrollo utilizamos las siguientes:

* Teachable Machine
* R-Studio
* TensorFlow
* Android Studio

Cada herramienta con un propósito específico. R-Studio es un programa especializado en la aplicación de Ciencias de datos y minería de datos con la cual realizamos la medición de datos de imágenes, Techable Machine que utiliza el modelo AutoML el cual indica que no se necesita ser un experto para poder crear un modelo de clasificación de imágenes, detección de objetos entre otros ejemplos, mediante la herramienta anterior se puede generar con el modelo entrenado un versión de TensorFlow Lite especializado para trabajar con plataformas de Android y finalmente el desarrollo de la aplicación mediante Android Studio siendo IDE de programación especializado para aplicaciones móviles.

Conclusiones

Tras el análisis de nuestra investigación para el desarrollo de un modelo de predicción para la clasificación de imágenes, se detalló la importancia de conocer el modelo del negocio y cómo funciona una de las enfermedades más impactantes del banano como es la sigatoka, se puede concluir que actualmente existen muchas investigaciones y de las dos clases de sigatoka que existen existe más información enfocada a la sigatoka negra que por lo general es la que más afecta en las producción de banano a nivel nacional e internacional.

Ahora que entendemos la importancia de la tecnología en la actualidad, esta hipótesis se enfocó en la idea de poder obtener un proceso de observaciones detalladas, no solo a través del ojo humano sino también a través de la tecnología que en la actualidad la inteligencia artificial y machine learning, están implementando soluciones en las múltiples áreas de negocios no solo de agricultura si no de cada campo posible de aplicar.

Mediante nuestra investigación nos enfocamos en Deep Learning siendo un modelo de machine learning muy eficaz para la propuesta planteada ya que como objetivo principal tenemos de poder clasificar la sigatoka en las diferentes clases que existen, sabemos que para muchos modelos especialmente cuando se trabaja con redes neuronales e imágenes Deep Learning utiliza Redes Neuronales Convolucionales que se especializan en la clasificación y detección de objetos de video o imágenes estáticas.

Como complemento se pueden utilizar herramientas que facilitan el trabajo a los desarrolladores, es muy importante entender el manejo de cada proceso para obtener el resultado final que se espera en un proceso de desarrollo de Data Science.

# Bibliografía

Castillo, B. (14 de Octubre de 2020). *6 tipos de métodos de investigación.* Obtenido de DIRECTORIO UNIVERSITARIO: https://guiauniversitaria.mx/6-tipos-de-metodos-de-investigacion/

Cooproclem. (04 de 07 de 2021). *DATOS GENERALES DE LA EMPRESA*. Obtenido de clementina.ec: http://clementina.ec/index.php/noticias/14-informacion-con-video

GUACHAMÍN, M. E. (2020). *Repositorio Universidad Central del Ecuador*. Obtenido de DISEÑO E IMPLEMENTACIÓN DE UNA APLICACIÓN DE MINERÍA DE DATOS QUE PERMITA SUGERIR PERSONAS CON INTERESES AFINES MEDIANTE EL ANÁLISIS DE SUS PREFERENCIAS EN MEDIOS SOCIALES: http://www.dspace.uce.edu.ec/bitstream/25000/7371/1/T-UCE-0011-289.pdf

Guzman, J. (2019). *Unidades de Apoyo para el Aprendizaje. CUAED/Facultad de Contaduría y Administración* . Obtenido de Técnicas de Investigación de Campo: https://uapa.cuaieed.unam.mx/sites/default/files/minisite/static/0fecd888-6a3f-4b31-b704-a2d94e3eed72/U000308176506/index.html

Jesús, E. Z. (Noviembre de 2019). *Aplicación de metodología CRISP-DM para segmentación geográfica de una base de imagenes.* Obtenido de Revista Ingeniería: https://www.revistaingenieria.unam.mx/numeros/2020/v21n1-08.pdf

Jordano, M. N. (2019). *Cumplimiento de Metas y Objetivos en el Departamento de Auditoría de la Hacienda Cooproclem.* Obtenido de Repositorio Digital Universidad Tecnica de Babahoyo: http://dspace.utb.edu.ec/bitstream/handle/49000/5688/-E-UTB-FAFI-INGCOM-000311.pdf?sequence=1&isAllowed=y

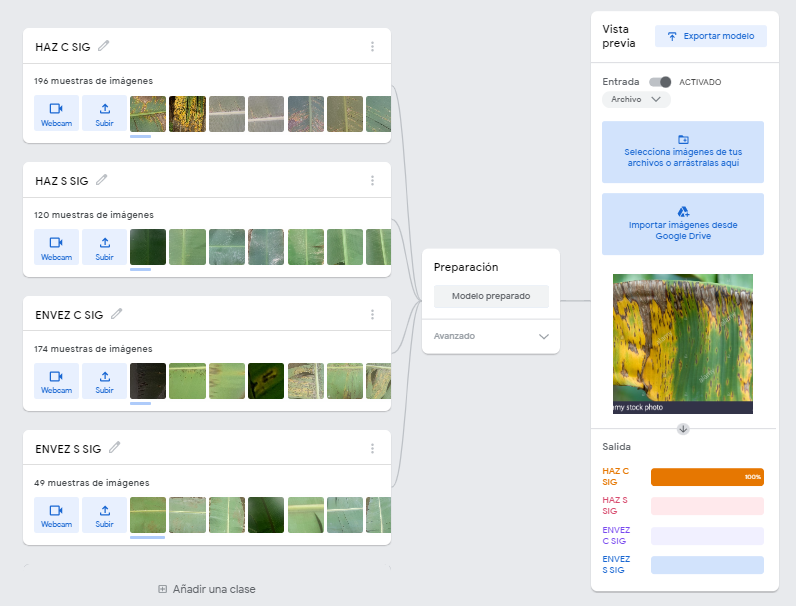
Ministerio de Comercio Exterior. (2017). www.produccion.gob.ec. *Informe Sector Bananero Ecuatoriano*, 50.

Olleta, J. E. (2021). *MÉTODO CORRELACIONAL*. Obtenido de DICCIONARIO DE PSICOLOGÍA CIENTÍFICA Y FILOSÓFICA: https://www.e-torredebabel.com/Psicologia/Vocabulario/Metodo-Correlacional.htm

ANEXOS



*Ilustración 1: Trabajo de labores de deshoje*



*Ilustración 2: Clasificacion y Entrenamiento del modelo de prediccion*



*Ilustración 3: Ejemplo de la aplicación con el modelo de entrenamiento*