



# POSGRADOS

## MAESTRÍA EN ELECTRÓNICA Y AUTOMATIZACIÓN

RPC-SO-19-No.277-2018

OPCIÓN DE  
TITULACIÓN:

PROYECTOS DE DESARROLLO

TEMA:

DESARROLLO DE UN SISTEMA DE AUTOMATIZACIÓN Y  
GESTIÓN DE LA INFORMACIÓN EN UNA LÍNEA DE  
PRODUCCIÓN DE BANANO

AUTOR:

PATRICIO JOSUÉ PARRA ABAD

DIRECTOR:

REMIGIO ISMAEL HURTADO ORTIZ

CUENCA - ECUADOR

2021

**Autor:**



***Patricio Josué Parra Abad***

Ingeniero Electrónico

Candidato a Magíster en Electrónica y Automatización,  
por la Universidad Politécnica Salesiana - Sede Cuenca

pparraa1@est.ups.edu.ec

**Dirigido por:**



***Remigio Ismael Hurtado Ortiz***

Ingeniero en Sistemas

Doctor en Ciencias y Tecnologías de la Computación para  
Smart Cities

rhurtadoo@ups.edu.ec

Todos los derechos reservados

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra para fines comerciales, sin contar con autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual. Se permite la libre difusión de este texto con fines académicos investigativos por cualquier medio, con la debida notificación a los autores.

DERECHOS RESERVADOS

©2021 Universidad Politécnica Salesiana.

CUENCA – ECUADOR – SUDAMÉRICA

PATRICIO JOSUÉ PARRA ABAD

***DESARROLLO DE UN SISTEMA DE AUTOMATIZACIÓN Y  
GESTIÓN DE LA INFORMACIÓN EN UNA LÍNEA DE  
PRODUCCIÓN DE BANANO***

# Índice general

|  |             |
|--|-------------|
| <b>Índice de Figuras</b>                       | <b>VI</b>   |
| <b>Índice de Tablas</b>                        | <b>VIII</b> |
| <b>Dedicatoria</b>                             | <b>IX</b>   |
| <b>Agradecimientos</b>                         | <b>X</b>    |
| <b>Acrónimos</b>                               | <b>XI</b>   |
| <b>Resumen</b>                                 | <b>XII</b>  |
| <b>Abstract</b>                                | <b>XIII</b> |
| <b>1. Introducción</b>                         | <b>1</b>    |
| 1.1. Descripción general del problema.....     | 3           |
| 1.2. Objetivos.....                            | 3           |
| 1.2.1. Objetivo general .....                  | 3           |
| 1.2.2. Objetivos específicos.....              | 4           |
| 1.3. Contribuciones .....                      | 4           |
| 1.4. Organización del manuscrito.....          | 4           |
| <b>2. Marco teórico</b>                        | <b>5</b>    |
| 2.1. Fundamentos de IoT.....                   | 6           |
| 2.1.1. Descripción general de IoT.....         | 6           |
| 2.1.2. Arquitectura de IoT.....                | 7           |
| 2.1.3. Aplicaciones de IoT .....               | 9           |
| 2.1.4. Beneficios del IoT.....                 | 10          |
| 2.2. Fundamentos de machine learning.....      | 11          |
| 2.2.1. Proceso del aprendizaje automático..... | 11          |

|           |   |           |
|-----------|---|-----------|
| 2.2.2.    | Métodos de aprendizaje.....                                 | 12        |
| 2.3.      | Algoritmos para problemas de clasificación y regresión..... | 13        |
| 2.3.1.    | Máquina de soporte vectorial.....                           | 13        |
| 2.3.2.    | Algoritmo clasificador Naïve de Bayes .....                 | 14        |
| 2.3.3.    | Algoritmo KNN.....  | 14        |
| 2.3.4.    | Árboles de decisión .....                                   | 15        |
| 2.3.5.    | Regresión Lineal.....                                       | 15        |
| 2.3.6.    | Redes Neuronales .....                                      | 16        |
| 2.3.7.    | Medidas de evaluación .....                                 | 17        |
| 2.3.8.    | Ejemplos de aplicación de machine leaning.....              | 17        |
| 2.3.9.    | Sumario.....  | 18        |
| <b>3.</b> | <b>Metodología del desarrollo</b>                           | <b>20</b> |
| 3.1.      | Terminología básica .....                                   | 21        |
| 3.2.      | Requisitos generales del sistema.....                       | 21        |
| 3.3.      | Arquitectura general del sistema.....                       | 22        |
| 3.4.      | Estructura física del sistema.....                          | 22        |
| 3.4.1.    | Diseño y construcción del chasis de la balanza electrónica  | 23        |
| 3.4.2.    | Subsistema para encendido y regulación de voltaje.....      | 24        |
| 3.4.3.    | Subsistema para carga de la batería.....                    | 25        |
| 3.4.4.    | Subsistema para visualización y panel de control.....       | 25        |
| 3.4.5.    | Subsistema para adquisición de datos.....                   | 26        |
| 3.4.6.    | Subsistema para comunicación inalámbrica .....              | 27        |
| 3.4.7.    | Subsistema de control.....                                  | 27        |
| 3.5.      | Desarrollo de la aplicación.....                            | 28        |
| 3.5.1.    | Ventana principal.....                                      | 28        |
| 3.5.2.    | Datos de la finca.....                                      | 29        |
| 3.5.3.    | Emparejar dispositivo .....                                 | 30        |
| 3.5.4.    | Orden de cosecha.....                                       | 31        |
| 3.5.5.    | Reporte Diario .....  | 32        |
| 3.5.6.    | Tablero de control .....                                    | 33        |
| 3.6.      | Aplicación de los algoritmos de machine learning.....       | 35        |
| 3.6.1.    | Diseño del experimento.....                                 | 35        |
| 3.6.2.    | Preprocesamiento de datos .....                             | 36        |
| 3.6.3.    | Implementación de la red neuronal.....                      | 37        |
| 3.6.4.    | Implementación del regresor de soporte vectorial.....       | 40        |
| 3.6.5.    | Implementación del algoritmo de regresión multilineal       | 41        |
| 3.6.6.    | Sumario.....  | 41        |

|  |           |
|--|-----------|
| <b>4. Resultados</b>                             | <b>43</b> |
| 4.1. Montaje Físico .....                        | 43        |
| 4.2. Sistema de gestión .....                    | 44        |
| 4.3. Algoritmos de aprendizaje supervisado ..... | 46        |
| 4.3.1. Optimización de la red neuronal .....     | 46        |
| 4.3.2. Rendimiento de los algoritmos .....       | 48        |
| 4.3.3. Sumario .....                             | 49        |
| <b>5. Conclusiones y trabajo futuro</b>          | <b>50</b> |
| 5.1. Conclusiones.....                           | 50        |
| 5.2. Trabajo futuro .....                        | 51        |

# Índice de Figuras

|  |    |
|--|----|
| 2.1. Red de conexión entre datos, personas, cosas y procesos (Figura tomada y traducida de (Firouzi, 2020))..... | 7  |
| 2.2. Modelo Arquitectónico de IoT (figura tomada y traducida de (Zhu y cols., 2017)) .....                       | 8  |
| 2.3. Arquitectura típica de una red neuronal. (Tomado y traducido de Gupta y Sehgal (2021)).....                 | 16 |
| 3.1. Estructura del sistema .....  | 22 |
| 3.2. Contorno 3D del chasis .....  | 23 |
| 3.3. Vista frontal (izq) y vista posterior del chasis .....  | 23 |
| 3.4. Esquema de implementación para encendido y regulación de voltaje.....                                       | 24 |
| 3.5. Esquema de implementación para cargar la batería.....   | 25 |
| 3.6. Esquema de implementación para el panel de control.....   | 26 |
| 3.7. Esquema de implementación para la adquisición de datos .....  | 26 |
| 3.8. Esquema de implementación para la comunicación inalámbrica .....  | 27 |
| 3.9. Esquema de implementación para el control del sistema .....   | 28 |
| 3.10. Diagrama de funcionamiento de la ventana principal del sistema para el control del dispositivo .....       | 29 |
| 3.11. Diagrama de funcionamiento del submenú de la aplicación (datos de la finca).....                           | 30 |
| 3.12. Diagrama de funcionamiento del submenú de la aplicación (emparejar dispositivo).....                       | 31 |
| 3.13. Diagrama de funcionamiento del submenú de la aplicación (orden de cosecha).....                            | 32 |
| 3.14. Diagrama de funcionamiento del submenú de la aplicación (reporte diario) .....                             | 33 |
| 3.15. Diagrama de funcionamiento del submenú de la aplicación (tablero de control).....                          | 34 |

|   |    |
|---|----|
| 3.16. Gráfica de la función de pérdida del modelo inicial .....                       | 37 |
| 3.17. Gráfica de la función de pérdida luego de la primera etapa de optimización..... | 38 |
| 4.1. Balanza electrónica.....   | 43 |
| 4.2. Ventana de ingreso a la aplicación .....   | 44 |
| 4.3. Ventana del tablero de control de la aplicación .....                            | 45 |
| 4.4. Ventana de generación de reportes de la aplicación .....                         | 46 |
| 4.5. Visualización mediante un mapa de calor .....                                    | 47 |

# Índice de Tablas

|  |    |
|--|----|
| 1. Acrónimos.....  | XI |
| 3.1. Descripción del dataset .....   | 36 |
| 3.2. Valores de prueba para la optimización por compilación de la red neuronal.....          | 38 |
| 3.3. Cantidad de neuronas en cada capa la optimización de la red neuronal .....              | 38 |
| 3.4. Porcentaje de neuronas apagadas en cada capa para la optimización por dropouts.....     | 39 |
| 3.5. Valores del peso obtenidos mediante el algoritmo de Red Neuronal Optimizado .....       | 39 |
| 3.6. Hiperparámetros para la implementación del regresor de soporte vectorial .....          | 40 |
| 3.7. Valores del peso obtenidos mediante el algoritmo Vector de soporte para regresión ..... | 40 |
| 3.8. Valores del peso obtenidos mediante el algoritmo Vector de soporte para regresión ..... | 41 |
| 4.1. Parámetros de la red neuronal original y optimizada.....                                | 47 |
| 4.2. Métricas para el rendimiento de los algoritmos. ....                                    | 48 |

# Dedicatoria

Dedico este trabajo a Dios por haberme dado fortaleza para seguir adelante. A mi esposa e hija que me estuvieron acompañando en todo el proceso dándome aliento. A mis padres que me apoyaron incansablemente en lo que era posible. A mis compañeros que hicieron cada clase y cada trabajo más fácil. Y a mi amiga Sandra Chacha que me ayudo en todo lo que estaba a su alcance para culminar a tiempo este trabajo.

# Agradecimientos

Agradezco a Dios por haberme dado la posibilidad de estudiar y la fuerza para superar las dificultades en mi vida y así poder cumplir mis metas profesionales. A mi esposa e hija que supieron afrontar todo el proceso en donde estuve ausente y a pesar de eso seguirme apoyando y dándome aliento a cada momento.

A mis padres que fueron los que me motivaron a seguir estudiando y poder ser un mejor profesional para ayudar a otras personas.

A Sandra Chacha que sin su apoyo no hubiera sido posible alcanzar la meta, demostrando una vez más que siempre se puede contar con ella.

A mi tutor Remigio Hurtado, por haberme brindado su apoyo, ideas y tiempo, además de alentarme para culminar con el posgrado.

# Acrónimos

En la primera tabla se muestra los acrónimos utilizados en el documento.

| <b>Acrónimo</b> | <b>Significado</b>  |
|-----------------|---|
| <b>IoT</b>      | <b>I</b> nternet <b>o</b> f <b>T</b> hings                              |
| <b>TIC</b>      | <b>T</b> ecnologías de la <b>I</b> nformación y la <b>C</b> omunicación |
| <b>ML</b>       | <b>M</b> achine <b>L</b> earning  |
| <b>AI</b>       | <b>A</b> rtificial <b>I</b> ntelligence                                 |
| <b>RFID</b>     | <b>R</b> adio <b>F</b> requency <b>I</b> dentification                  |
| <b>NFC</b>      | <b>N</b> ear <b>F</b> ield <b>C</b> ommunication                        |
| <b>SVM</b>      | <b>S</b> upport <b>V</b> ector <b>M</b> achine                          |
| <b>KNN</b>      | <b>K</b> -nearest <b>N</b> eighbors                                     |
| <b>EPC</b>      | <b>E</b> lectronic <b>P</b> roduct <b>C</b> ode                         |
| <b>RMSE</b>     | <b>R</b> oot of <b>M</b> ean <b>S</b> quared <b>E</b> rror              |

Tabla 1: Acrónimos

# Resumen

Los sistemas de gestión de la información permiten organizar los datos de manera que se convierta en información útil para el usuario. La forma en que se gestiona ha cambiado debido al uso de las denominadas tecnologías de la información (TIC), por ello, una parte esencial es combinar los métodos y técnicas para que el sistema sea eficiente y se adapte al campo de aplicación.

En el caso de la agricultura actualmente la tecnología se enfoca en controlar la calidad del producto, sin embargo, es importante también gestionar la información que se obtiene de los procesos, pues esto brindará un contexto de lo que ocurre en la línea de producción en general, por ejemplo control de inventario, peso del producto, datos de la cosecha, entre otros.

Un sistema de gestión de la información involucra realizar varias actividades: se inicia con la estrategia para decidir qué datos se recolectarán, luego se determina cómo se obtendrán estos datos, el siguiente paso es organizar la información recolectada para después difundirla y finalmente evaluarla. Gracias al uso de TIC, la aplicación de machine learningo aprendizaje automático, desarrollo WEB y diseños de dispositivos electrónicos es posible automatizar muchas tareas que intervienen en el sistema.

En este contexto, el presente trabajo presenta el desarrollo de un sistema automatización y gestión de la información en una línea de producción de banano; gracias al uso de técnicas de diseño de circuitos, desarrollo de aplicaciones móviles; además, se logró realizar un aprendizaje pasivo mediante la implementación de tres algoritmos de machine learning.

La propuesta de este trabajo contribuye a que una línea de producción de banano automatice la gestión de su información gracias a la recolección de datos de pesaje con una balanza electrónica que envía la información hacia una aplicación, que permite generar un reporte para el posterior análisis de datos.

# Abstract

Information management systems allow organizing data so that it becomes useful information for the user. The way how the information is managed has changed due to the use of information technology (ICT), then, an essential part is to match the methods and techniques, in a way that the system is efficient and adapts to the field of application.

In the case of agriculture, technology focuses on controlling the quality of the product; however, it is important to manage the information obtained from the processes too, because it is supported to know about what happens in the production line in general. For example, it allows knowing the stocktaking, weight of products, yield information, and others. An information management system involves many activities: it begins with the strategy to decide which data to be collected, then, it is determined how this data will be obtained, the next step is to organize the information collected and then communicate it and finally evaluate it.

Thanks to the use of TIC, knowledge of machine learning algorithms, programming, and electronic device designs, it is possible to automate the tasks involved in the system.

In this context, the present work presents the implementation of a system for automation and information management, which was applied in a banana production line. It was possible due to the use of circuit design techniques, mobile application development; also, an experiment of passive learning was made, through the implementation of three developer machine-learning algorithms.

The proposal of this work contributes to the automation of a banana production line, about the management of its information thanks to the collection of weighing data with an electronic scale that sends the data to an application, which generate a report for take decisions later.

# Capítulo 1

## Introducción

La tecnología tiene un papel protagónico en la sociedad actual pues ha generado cambios en áreas como la comunicación, los procesos de producción y en general la cotidianidad del ser humano (Pérez y Milla, 2006). Es evidente que la sociedad está cambiando vertiginosamente por ello cada sector debe adaptarse, incluyendo el sector agrícola, siendo este de vital importancia. Según el informe de la Organización para la Alimentación y Agricultura de las Naciones Unidas (FAO, por sus siglas en inglés) la implementación de nuevas tecnologías en la agricultura podría contribuir a mejorar las condiciones de hambre y de pobreza extremas (FAO, 2016). En este sentido, Ecuador es un país agrícola que produce principalmente cacao, café, arroz, flores, palma y banano (Lombeida, 2020).

Existen varias aplicaciones de tecnología en la agricultura que se enfocan principalmente de resolver problemas en prácticas de cultivos, mejorarla producción, análisis de suelos, planificación, gestión de la información meteorológica, entre otros. Además, buscan innovar y mejorar los procesos de trabajo en el sector agrícola. En (Palmieri, 2007) menciona que un elemento esencial para la innovación tecnológica es el manejo de la información de tal manera que esta pueda convertirse en conocimiento útil para empresas o las personas involucradas durante el proceso productivo, es así que surge la necesidad de tener con un sistema de gestión de información como una herramienta de automatización para manejar grandes volúmenes de información que se generan durante los procesos o actividades de una organización (De Oliveira y Sian, 2014).

La gestión de la información es considerada como todo proceso que se realiza para capturar, clasificar, preservar, recuperar, compartir y

difundir la información. La manera cómo se gestiona la información ha evolucionado debido a la implementación de las tecnologías de la información y comunicación (TIC). El manejo adecuado de la información contribuye a la innovación tecnológica, ahí radica la relevancia de la gestión de la información (Palmieri, 2007).

En (Palmieri, 2007) resume que la gestión de la información involucra los siguientes grupos de actividades principales:

- Estrategia de información
- Acceso u obtención de la información
- Organización
- Difusión de información
- Evaluación

Un sistema de gestión de información es el resultado de una sinergia entre diversos campos de conocimiento, así por ejemplo: el internet de las cosas (IoT, por sus siglas en inglés) permite la conectividad entre los elementos que intervienen en el sistema, gracias a la aplicación de algoritmos de aprendizaje automático (ML, por sus siglas en inglés) es posible implementar modelos para que los sistemas aprendan automáticamente, finalmente para realizar aplicaciones con interfaces amigables para el usuario se utilizan técnicas de desarrollo web, programación y diseño de dispositivos (Rehman y Mahmud, 2019).

Una vez que se ha presentado los beneficios que la tecnología aporta al desarrollo del sector agrícola y la importancia que tiene el cultivo de banano para el Ecuador, se puede identificar la necesidad de la automatización de procesos. Por tanto, este trabajo surge con la idea de automatizar el proceso de pesaje de los racimos del banano, de tal forma que se pueda optimizar el tiempo de recolección del datos del peso del racimo y el manejo de la información obtenida.

En este trabajo se desarrolló un sistema de automatización y gestión de la información para la línea de producción de banano en una finca ubicada en el cantón Durán de la provincia del Guayas. Las pruebas de su funcionamiento del sistema se realizaron durante el lapso de 30 días dando resultados alentadores para la implementación de sistemas relacionados y trabajos futuros. Los datos para la implementación de los algoritmos de aprendizaje automático se tomaron del trabajo de (Enriquez Tovar y Vega Caiza, 2011).

## 1.1. Descripción general del problema

Generalmente los procesos de automatización al igual que la gestión de la información en el sector de la agricultura han sido vistos como temas separados debido a que se considera un área muy rústica para implementación de tecnología, sin embargo, la demanda de producción de alimentos sumado al aumento de la población hace que se deba buscar medidas para optimizar los procesos (Rehman y Mahmud, 2019). Ecuador es un país con alta demanda de producción de banano en los meses de enero a mayo del 2020 con relación al 2019 existe un incremento de 10,33 %. En la actualidad, la capacidad de exportación del Ecuador es significativa ya que es el segundo país que más exporta banano. Según el reporte generado por la Asociación de Exportadores de Banano del Ecuador (AEBE, 2020) entre enero y mayo se exportaron 175,462.102 cajas. Por consiguiente, es necesario encontrar maneras de automatizar los procesos en este campo.

Los problemas en los procesos de producción de banano se presentan durante el pesaje pues la logística para la recolección de esta información se realiza de forma manual, así: el racimo de banano ingresa a una celda de carga entonces los datos de pesado son tomados por un operario y escritos en hojas, estas hojas son almacenadas pero los datos tomados no generan información útil para obtener un control de la producción. Uno de los problemas más importantes es el tiempo que un operador se demora en anotar el peso de manera manual en las hojas que no tienen un sistema de gestión de la información. Otro problema es el peso útil del racimo de banano ya que la estimación del peso del tallo se realiza mediante inspección visual, pues no existe un sistema automático que pueda estimar este peso lo cual genera una incertidumbre sobre el proceso de producción para la empresa.

Por ello, es necesario un sistema que permita mejorar el tiempo operativo y administrativo. En el capítulo 3 se explica con detalle cómo el sistema propuesto es capaz de realizar la recolección automática de los datos y gestionar la información recolectada para la generación de reportes. A continuación, se presentan el objetivo general y también los objetivos específicos de este trabajo.

## 1.2. Objetivos

### 1.2.1. Objetivo general

Desarrollar un sistema de automatización y gestión de la información en una línea de producción de banano.

### 1.2.2. Objetivos específicos

- Establecer un sistema prototipo para la adquisición y envío de datos hacia un dispositivo gestor de forma inalámbrica.
- Mostrar la etapa de gestión de información a través de una plataforma móvil articulando el reporte de la información obtenida.
- Estimar el peso de la cabeza del banano por medio de un algoritmo machine learning supervisado de regresión.

### 1.3. Contribuciones

El presente trabajo resuelve las necesidades de control y gestión en la línea de procesos en una hacienda productora de banano. La plataforma del sistema desarrollado posee una interfaz amigable para el usuario, el proceso de pesaje automático consiste en una balanza que se adapta a la infraestructura de pesaje actual de los racimos de banano y envía los datos hacia una aplicación Android. Además, este trabajo muestra cómo realizar la estimación de peso del tallo del racimo mediante tres diferentes algoritmos de machine learning.

El sistema es un aporte para: automatizar los procesos de producción, rápida recolección de datos, lograr eficiencia del transporte de la información entre la planta de operación y la parte administrativa que maneja los datos del cultivo, debido a su adaptabilidad brinda la posibilidad de instalar un sistema de pesaje automático en más áreas de la hacienda. Finalmente, todo esto permite que los datos recopilados se conviertan en información útil para la empresa.

### 1.4. Organización del manuscrito

El presente documento recopila la información del trabajo desarrollado, las siguientes secciones del documento están estructuradas como se describe a continuación:

El capítulo 2 presenta un marco teórico referencial, el capítulo 3 expone la metodología utilizada para el desarrollo del sistema de gestión de la información. El capítulo 4 expone los resultados obtenidos durante las pruebas de funcionamiento, finalmente el capítulo 5 expone las conclusiones y trabajo futuro.

## Capítulo 2

# Marco teórico

Este capítulo expone el fundamento teórico referencial para el desarrollo del trabajo, el capítulo aborda los siguientes puntos:

- Fundamentos de IoT
  - Descripción general de IoT
  - Arquitectura de IoT
  - Aplicaciones de IoT
- Fundamentos de aprendizaje automático (Machine Learning)
  - Proceso de aprendizaje automático
  - Métodos de aprendizaje automático
- Algoritmos para problemas de clasificación y regresión
  - Máquina de soporte vectorial
  - Algoritmo clasificador Naive de Bayes
  - Algoritmo KNN
  - Árboles de decisión
  - Regresión lineal
  - Redes Neuronales
  - Medidas de evaluación del algoritmo
  - Ejemplos de aplicación de machine learning

## 2.1. Fundamentos de IoT

### 2.1.1. Descripción general de IoT

El Internet de las cosas (IoT, por sus siglas en inglés) no posee una definición estándar sin embargo según (Lee, 2015) la base de IoT es formar una red de dispositivos conectados que se comunican y comparten información para lograr un objetivo en común. A continuación se resaltan algunas características dentro de IoT, la definiciones fueron tomadas de (Firouzi, 2020):

- Cosas inteligentes

El sistema posee dispositivos dotados de inteligencia que interactúan entre sí y con su entorno. Un dispositivo inteligente cuenta con una unidad de procesamiento, sensores/actuadores, posee una forma de conectividad con los demás dispositivo, tiene una etiqueta para ser identificado en la red y una fuente de poder.

- Conectividad

Todos los elementos de la red que participan en el sistema se mantienen comunicados.

- Datos

En IoT se recopilan datos que se convertirán en información para tomar acciones en los sistemas, por ejemplo, las condiciones climáticas son datos que se pueden utilizar en proyectos para regular el consumo de energía en una empresa.

- Inteligencia

Esta característica hace referencia a la aplicación de Inteligencia artificial (Artificial Intelligence), aprendizaje automático (Machine Learning) o análisis de datos (Data analytics) para desarrollar el sistema.

- Ecosistema

Esta característica se refiere a la importancia de considerar el entorno completo, su campo de aplicación y objetivos.

- Heterogeneidad

La heterogeneidad indica que los elementos de IoT deben ser capaces de funcionar en diferentes plataformas en diferentes redes.

- Cambios dinámicos

En un sistema de IoT los dispositivos adaptan sus modos de operación según los cambios que se presenten en la red, además, es escalable de manera que permita agregar nuevos dispositivos en función de los datos que se gestionen.

- Seguridad y privacidad

En las aplicaciones de IoT se maneja información personal y de interés según el campo de aplicación, por ello es imprescindible garantizar la confidencialidad de los datos del usuario.

### Elementos del IoT

El campo de aplicación de internet de las cosas es muy amplio desde electrodomésticos para los hogares hasta la automatización en la industria en donde se pretende mejorar los procesos de producción y logística (Yury, 2017). Según (Firouzi, 2020) en IoT intervienen varios elementos los cuales se clasifican en 4 grupos: cosas, datos, personas y procesos. La figura 2.1 ilustra de manera resumida los elementos que componen el ecosistema de IoT.

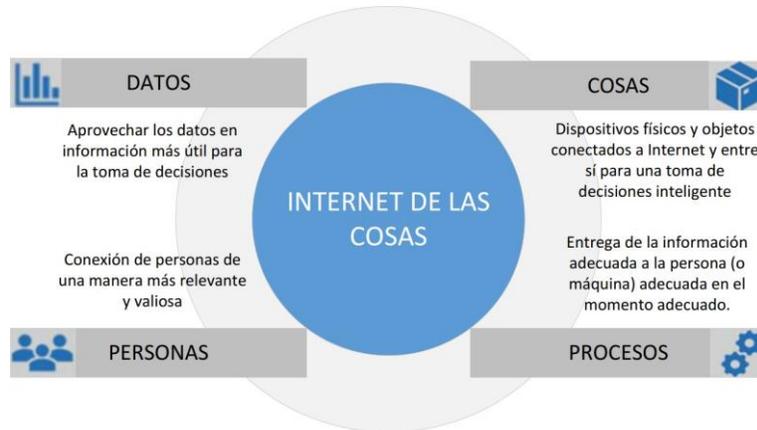


Figura 2.1: Red de conexión entre datos, personas, cosas y procesos (Figura tomada y traducida de (Firouzi, 2020))

#### 2.1.2. Arquitectura de IoT

Los componentes de la arquitectura de IoT dependen de su campo de aplicación, sin embargo una característica común es que los dispositivos

que intervienen se interconectan entre si y además se comunican de manera autónoma (Nord, 2019). La figura 2.2 muestra un resumen de los elementos que forman parte del modelo con la arquitectura IoT que contiene tres capas.



Figura 2.2: Modelo Arquitectónico de IoT (figura tomada y traducida de (Zhu y cols., 2017))

### Capa física o de dispositivo

Se encarga de la recolección de datos, es la capa básica del IoT en la cual intervienen tecnologías de comunicación como por ejemplo: la identificación por radiofrecuencia (RFID, por sus siglas en inglés), comunicaciones de campo cercano (NFC, por sus siglas en inglés). Además, en esta capa se encuentran los sensores inalámbricos y actuadores (Zhu y cols., 2017), (Wortmann, 2015).

### Capa de red

Es la capa de conectividad del IoT, manejo de software e incluye los dispositivos que intervienen en la infraestructura para lograr la comunicación (Wortmann, 2015).

### Capa de aplicación

Esta capa se encarga del manejo de los datos generados por los dispositivos que intervienen en el sistema. En esta capa también se gestiona la interacción

de los usuarios ([Wortmann, 2015](#)).

### 2.1.3. Aplicaciones de IoT

Según las búsquedas en Google y temas de interés para conversaciones en Twitter y LinkedIn, las categorías relacionadas con IoT más rankeadas son: smart home, dispositivos wearables y smart city. En cuanto a los campos de aplicación basados en proyectos reales, las tres áreas que sobresalen son: conectividad en la industria, ciudades inteligentes y energía inteligente ([Nord, 2019](#)).

#### Clasificación de las aplicaciones

Debido a que son varios los campos de aplicación de IoT, en ([Atzori, 2010](#)) se clasifican en 3 grupos de interés: transporte y logística, salud y ambientes inteligentes.

- Aplicaciones en el transporte y logística

Este campo involucra la automatización con el objetivo de agilizar los procesos de monitoreo en la cadena de suministros, para su implementación se utilizan tecnologías como la identificación por radiofrecuencia (RFID, por sus siglas en inglés) o la comunicación de campo cercano (NFC, por sus siglas en inglés).

IoT también toma un papel importante en el desarrollo de asistentes de manejo, los vehículos dotados de tecnología son una opción para brindar mayor seguridad, por ejemplo para precautelar colisiones de vehículos que transportan cargas peligrosas. Otro campo importante es la integración de IoT en vehículos que transportan mercancía pues podría ser utilizada por las empresas para mejorar la planificación de rutas en función del tráfico de carretera o conocer la ubicación del producto lo cual permitiría un monitoreo constante de la carga para estimar tiempos de entrega ([Atzori, 2010](#)).

Otro ejemplo de aplicación de IoT en la logística de productos es la implementación de etiquetas en los productos, estas etiquetas no solo son una manera de condensar información para consultar datos, también son una manera de conectar directamente a enlaces o servicios relevantes que intervienen en los procesos logísticos ([Atzori, 2010](#)).

- Aplicaciones en la salud

Según (Atzori, 2010) la aplicación de IoT en la salud se clasifica en dos usos: el transporte y el manejo de información. En cuanto al transporte se enfoca en el traslado de un paciente, al integrar IoT sería posible monitorear la ubicación del paciente y el personal que interviene durante su atención. Con respecto al manejo de la información gracias a IoT se podría acceder rápidamente a los datos del paciente como por ejemplo al historial clínico, lo cual contribuye a identificar las necesidades especiales de cada uno en menor tiempo.

Además, un beneficio de la aplicación del IoT es monitorear el estado del paciente mediante sensores que permiten conocer la información de sus indicadores de salud en tiempo real que ayudaría para proporcionar un diagnóstico más ágil basado en los datos recolectados de los sensores (Mohammad, 2021).

- Aplicación en hogares u oficinas

Aplicar IoT es convertir hogares u oficinas en espacios dotados de dispositivos inteligentes que generan un lugar más confortable para el usuario, además estos espacios son construidos para contar con eficiencia energética y seguridad electrónica (Atzori, 2010).

IoT también se aplica en el sector agrícola con el objetivo de maximizar la producción, uno de los usos más extendidos de IoT en la agricultura es el uso de sensores que permiten recopilar información para controlar la dosificación de pesticidas, monitorear el estado de los suelos, analizar el rendimiento y pronosticar posibles acciones para maximizar la producción de los cultivos (Ratnaparkhi y Khan, 2020), (Firouzi, 2020).

Los campos de interés para la aplicación de IoT en el campo empresarial son: mejorar la experiencia del cliente, el ámbito financiero y el de la gestión de activos (Wortmann, 2015).

#### **2.1.4. Beneficios del IoT**

A continuación se expone algunos de los beneficios de los sistemas de IoT que enumera (Firouzi, 2020).

- Eficiencia

El manejo ágil de los datos y el monitoreo de los parámetros en tiempo real permite tomar decisiones más inteligentes, además controlar el estado de los productos físicos durante los procesos de producción.

- Automatización y control

IoT permite la conexión entre distintos dispositivos por ello se mantiene un control dentro de los procesos sin la necesidad de la intervención de una persona, esto beneficia las tareas de supervisión. Las tareas repetitivas se automatizan lo cual disminuye la cantidad de errores en los procesos, gracias a los procesos de automatización de tareas se reduce el tiempo de operación.

- Manejo de la información

Los datos recolectados en los sistemas de IoT se convierten en información de interés para tomar decisiones, desde las tareascotidianas hasta en los procesos de la industria.

- Costo/dinero

IoT permite automatizar procesos, reducir errores y agilizar los procesos por ende se reduce el consumo de energía lo que se ve reflejado en reducción de los costos. La implementación de IoT se considera una inversión inteligente siempre que el costo del equipo de monitoreo o automatización sea menor que la cantidad que el dinero que se va ahorrar.

## 2.2. Fundamentos de machine learning

El machine learning o en español aprendizaje automático traducido al español, es una rama de la inteligencia artificial enfocada en la predicción de datos mediante la automatización del proceso de aprendizaje sin necesidad de la intervención del ser humano (Paluszek, 2019). Durante el desarrollo de este proyecto se utilizará el término machine learning en inglés debido a que es un término convencional utilizado en la bibliografía.

### 2.2.1. Proceso del aprendizaje automático

Según (Ramasubramanian, 2017) y (Vieira, 2020) el proceso de aprendizaje consta de 5 etapas que se describen a continuación:

1. Formulación del problema

Consiste en identificar el problema que vamos a resolver, el objetivo del aprendizaje, las tareas durante el aprendizaje y las variables que intervienen.

## 2. Recopilación o colección de datos

La recopilación de datos se realiza según la formulación del problema. Los datos se obtienen de encuestas, entrevistas, observación, experimentación o bases de datos online.

## 3. Procesamiento de datos

Durante esta etapa se realiza la gestión de la información, que consiste en organizar y filtrar los datos útiles para el problema. Además, todos los datos se colocan en un formato común, se escalan o agregan características.

## 4. Desarrollo del modelo

Consiste en crear el modelo de aprendizaje en base a los datos de entrada y salida por ello en esta etapa se elige el método de aprendizaje.

## 5. Evaluación del modelo

Esta etapa es la de validación del modelo que consiste en cuantificar su desempeño.

### 2.2.2. Métodos de aprendizaje

Esta sección muestra un resumen de los diferentes métodos de aprendizaje que se pueden utilizar según el problema a resolver o campo de aplicación.

#### **Aprendizaje supervisado**

Este tipo de aprendizaje cuenta con un conjunto de datos específicos o muestras etiquetadas y otro grupo de datos no etiquetados. El proceso de aprendizaje consiste en descubrir la relación entre el conjunto de datos etiquetados y no etiquetados, este proceso de clasificar las salidas para un grupo dado de datos de entrada se denomina etiquetado. El aprendizaje supervisado se usa comúnmente en problemas relacionados con regresión y clasificación ([Ramasubramanian, 2017](#)),([Malviya, 2020](#)).

#### **Aprendizaje no supervisado**

Este tipo de aprendizaje no cuenta con un grupo de variables etiquetadas por ello el aprendizaje se genera como resultado de la similitud de los datos. Las aplicaciones de este tipo de aprendizaje se utilizan en problemas de segmentación y agrupación ([Ramasubramanian, 2017](#)),([Malviya, 2020](#)).

### Aprendizaje semi-supervisado

Utiliza grupos de datos etiquetados y no etiquetados, el número de muestras etiquetadas es menor que los no etiquetados; el número de muestras menor debido a que por lo general se necesita personal cualificado para tomarlas. Este método de aprendizaje se utiliza para problemas de agrupación o asociación (Ramasubramanian, 2017),(Malviya, 2020).

### Aprendizaje por refuerzo

En este aprendizaje se utilizan datos limpios y precisos para mejorar los resultados del aprendizaje, los datos son recopilados de la interacción con el entorno, se predice en función de las experiencias. Este enfoque se aplica para resolver problemas de clasificación y control (Ramasubramanian, 2017),(Malviya, 2020).

## 2.3. Algoritmos para problemas de clasificación y regresión

Luego de conocer los diferentes métodos de aprendizaje automático se evidencia que es común utilizarlos en problemas de clasificación y regresión por ello en esta sección abordaremos algunos de los algoritmos más comunes enfocados a resolver este tipo de problemas. Cabe destacar que en el presente trabajo se utilizaron métodos de aprendizaje supervisado.

### 2.3.1. Máquina de soporte vectorial

La técnica de máquina de soporte vectorial (SVM, por sus siglas en inglés) es un método que se basa en el aprendizaje estadístico, se usa principalmente en problemas de clasificación y regresión. Para clasificar el algoritmo de SVM encuentra un hiperplano encargado de separar los conjuntos de variables que pertenecen a clases diferentes (Awad y Khanna, 2015). El hiperplano se define como una función expresada con la ecuación 2.1.

$$g(x) = w^T x + b. \quad (2.1)$$

En el trabajo de (Awad y Khanna, 2015) expresa la fórmula para definir el problema de para SVM con la ecuación 2.2

$$L(w, b, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^N \lambda_i [y_i (w^T x_i + b) - 1] \quad (2.2)$$

En donde  $w$  y  $b$  se denominan variables primarias y son multiplicadores de Lagrange.

### 2.3.2. Algoritmo clasificador Naïve de Bayes

El algoritmo clasificación de Naïve-Bayes (NB) de aprendizaje supervisado, es un método del tipo de clasificador probabilístico pues basa su funcionamiento en el Teorema de Naïve-Bayes. El modelo de aprendizaje automático predice la probabilidad de que una variable pertenezca a cada clase dado un conjunto de características. Para lo cual el algoritmo asume que la característica de una clase es independiente de cualquier otra característica. En (Gupta y Sehgal, 2021) describe el Teorema de Bayes con la ecuación 2.3.

$$PAG(A|B) = \frac{PAG(A|B)}{PAG(B)} P(A) \quad (2.3)$$

De donde:

$P(A|B)$  Probabilidad de que se produzca la hipótesis A dado el evento observado B

$P(B|A)$  Probabilidad de obtener B dada la hipótesis A verdadera.

$P(A)$  Probabilidad de la hipótesis A

$P(B)$  Probabilidad de observar el evento B

### 2.3.3. Algoritmo KNN

El algoritmo K-nearest neighbors (KNN) en español de los K vecinos más cercanos, es un método que se considera simple debido a que la solución al problema de clasificación depende de los vecinos más cercanos al dato no identificado o no etiquetado, es por ello que un mismo objeto puede ser clasificado en diferentes categorías. Para clasificar un objeto el algoritmo calcula la distancia hacia un objeto etiquetado de entrenamiento y entonces le asigna una etiqueta en base a los objetos dominantes Uddin, Khan, Hossain, y Moni (2019). La opción más usada para determinar la distancia es el cálculo de la distancia euclidiana, en Gupta y Sehgal (2021) muestra la fórmula como 2.4.

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \dots + (x_n - x'_n)^2} \quad (2.4)$$

De donde:

$d$  es la distancia

$x$  y  $x'$  son dos vectores.

### 2.3.4. Árboles de decisión

Los árboles de decisión (en inglés Decision Trees) se usan para resolver problemas de regresión y clasificación. Según (Joshi, 2020) uno de los algoritmos más utilizados para construir árboles de decisión se denomina CART. Su funcionamiento se basa en tomar un conjunto de datos a la entrada y dividirlos de manera recursiva evaluando sus características, esta división genera una estructura en forma de árbol, que consta de elementos denominados nodos de decisión y hojas. En los nodos se toma la decisión de hacia a donde se ramifica mientras que las hojas son las etiquetas o salidas (Chowdhary, 2020).

Según (Gupta y Sehgal, 2021) para tomar la decisión en cada nodo se utilizan 2 métricas: ganancia de información y el índice de Gini.

La ganancia de la información se calcula con la siguiente expresión:

**Ganancia de la información** = Entropía(S) - (promedio ponderado) \* Entropía (para cada atributo).

De donde, la entropía se calcula con la expresión 2.5 mostrada en (Gupta y Sehgal, 2021).

$$Entropía = \sum_j - P_j \log_2 P_j \quad (2.5)$$

De donde:

$P_j$  es la probabilidad de la clase  $j$  y  $s$  es el número total de datos.

El índice Gini se calcula con la expresión 2.6 mostrada en (Gupta y Sehgal, 2021)

$$índice\ de\ Gini = 1 - \sum_j P_j^2 \quad (2.6)$$

### 2.3.5. Regresión Lineal

La técnica de regresión lineal es una manera de acoplar los datos a un modelo, describe la relación que existe entre las variables que intervienen en un problema. Para resolver los problemas existen dos tipos de regresiones: la regresión lineal simple y la regresión multivariable (Malviya (2020), Haroon (2017)).

La regresión lineal simple es un modelo en donde existe la relación de uno a uno es decir intervienen máximo dos variables, sin embargo, en las aplicaciones de la vida real existen varios factores que forman parte de un sistema, por esta razón existe la segunda técnica denominada regresión lineal multivariable en la cual se puede establecer la relación entre el conjunto de variables independientes y variables dependientes, con esta técnica se

modelan problemas más complejos del mundo real pues permite realizar análisis con más de dos variables [Haroon \(2017\)](#).

En [Haroon \(2017\)](#) define matemáticamente a la regresión multivariable con la expresión 2.7.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots b_ix_i + e \quad (2.7)$$

De donde:

y: Variable dependiente

$x_1 + x_2 + x_3 \dots, x_i$  : Variables

$b_0$ : Pendiente

$b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots b_ix_i$ : Coeficientes

e: Regresión residual

### 2.3.6. Redes Neuronales

Según [Uddin y cols. \(2019\)](#) las redes neuronales son un conjunto de algoritmos que se parte del principio del funcionamiento de los sistemas biológicos como las redes neuronales del cerebro.

Los sistemas de redes neuronales se representan como la interconexión de nodos que se hallan organizadas en capas: en primer lugar, la capa de entrada, luego la capa oculta y finalmente la capa de salida. En [Gupta, 2021](#)) menciona que la implementación de una red neuronal implica una gran cantidad de datos por lo cual el entrenamiento demanda recursos computacionales, sin embargo, gracias a la evolución de las capacidades computacionales ahora es posible utilizar las redes neuronales en el aprendizaje supervisado en varios campos como el reconocimiento de imágenes o voz. La figura 2.3 muestra la arquitectura típica de una red neuronal.

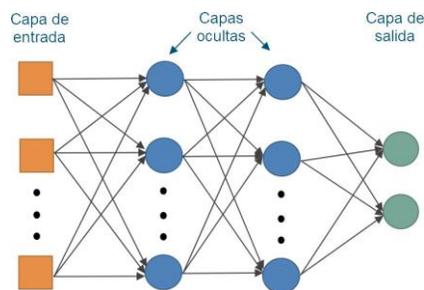


Figura 2.3: Arquitectura típica de una red neuronal. (Tomado y traducido de [Gupta y Sehgal \(2021\)](#))

### 2.3.7. Medidas de evaluación

Según (Ortiz, 2020) para evaluar una predicción las medidas de calidad que más se utilizan son: accuracy (precisión) y coverage (cobertura). Para este proyecto se evaluó mediante la medida de accuracy pues permite conocer la diferencia entre el valor de predicción y el valor real.

Para medir la precisión se utilizó el cálculo de la raíz del error cuadrático medio (RMSE de sus siglas en inglés), vale destacar que mientras el valor es más bajo la calidad del modelo es mejor. El RMSE se calcula mediante la fórmula 2.8 mostrada en (Ortiz, 2020).

$$RMSE = \frac{1}{\#U} \sum_{u \in U} \sqrt{\frac{1}{\#O_u} (p_{u,i} - r_{u,i})^2} \quad (2.8)$$

De donde:

$p_{u,i}$  es la predicción.

$r_{u,i}$  es el valor real.

Además, en (Martínez, 2020) menciona que para problemas de regresión se calcule el Coeficiente de Determinación, el cual, mientras más alto la calidad del modelo es mejor su identificador es  $R^2$ .

Se calcula con la fórmula 2.9:

$$R^2 = 1 - \frac{\text{var}(\text{error})}{\text{var}(\text{salida})} = 1 - \frac{\sum_i (y_i - t_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (2.9)$$

De donde:

$y_i$  es la salida estimada

$t_i$  es la salida esperada.

### 2.3.8. Ejemplos de aplicación de machine learning

Machine learning tiene varios campos de aplicación, a continuación, semuestran algunos ejemplos incluyendo en el área agrícola.

- Experiencia de usuario/cliente

En este punto se encuentran los sistemas de recomendación que son una herramienta que permite mejorar la experiencia del usuario mediante la predicción de recomendaciones en base a sus preferencias. Los sistemas de recomendación más actuales se basan en el filtrado colaborativo el cual generalmente se implementa mediante el algoritmo KNN de

machine learning, en donde se utiliza las calificaciones para realizar las predicciones (Hurtado, Bobadilla, Bojorque, Ortega, y Li, 2018).

- Salud

En la era digital las redes sociales son un medio que influyen en la cotidianidad de las personas, convirtiéndose en un campo de estudio de interés pues contienen datos de salud que los usuarios comparten en sus redes. El trabajo de (Trifan, Antunes, y Oliveira, 2021) muestra cómo los algoritmos de machine learning pueden utilizarse para estudiar el comportamiento de los usuarios en redes sociales y enfocarse en situaciones específicas como detectar individuos deprimidos en redes sociales. Este trabajo realiza la comparación de tres algoritmos (máquinas de soporte vectorial, bayes multinomial y pasivo agresivo).

- Agricultura de precisión

El trabajo de (Vishwanath, Upendra, y Ahmed, 2021) muestra una revisión de las aplicaciones en el campo agrícola mediante principios de IoT y algoritmos de machine learning. Un ejemplo es la detección de la enfermedad de tizón tardío en las papas y tomates, para ello se tiene un sistema inteligente de recolección y procesamiento de los datos como humedad, y temperatura, luego, mediante el algoritmo de máquina de soporte vectorial se determina la gravedad de la enfermedad.

Otra aplicación de machine learning la presenta (González, 2020) en el cual realiza un clasificador de productos agrícolas para el control de calidad mediante un modelo de red neuronal y un algoritmo de entrenamiento KNN. En este caso el trabajo se centró en la clasificación de bananas, naranjas, plátano verde y manzanas; bajo las condiciones del experimento, las naranjas fueron el producto más fácil de identificar mediante el algoritmo KNN.

### 2.3.9. Sumario

En este capítulo se abordó una breve descripción de las características principales, aplicaciones y arquitectura del internet de las cosas (IoT). Conocer estas definiciones es imprescindible para entender como una red compuesta de dispositivos interconectados trabajan en conjunto para lograr un objetivo común (Lee, 2015). Los objetivos están definidos según el campo de aplicación, así por ejemplo: IoT puede aplicarse en campos de la salud, construcción, transporte y por su puesto la agricultura (Atzori, 2010).

Además, en el capítulo se abordó la temática de machine learning, de lo cual es importante recalcar sus métodos, algoritmos de aprendizaje, medidas de evaluación y ejemplos de aplicaciones.

Existen cuatro métodos de aprendizaje: supervisado, no supervisado, semi-supervisado y por refuerzo ([Ramasubramanian, 2017](#)), ([Malviya, 2020](#)). En cuanto al estudio de los algoritmos de aprendizaje este capítulo se enfocó en los algoritmos de aprendizaje supervisado que resuelven problemas de regresión y clasificación.

Finalmente, luego de conocer algunas definiciones generales y aplicaciones de IoT, machine learning, el siguiente capítulo muestra el desarrollo del trabajo propuesto.

## Capítulo 3

# Metodología del desarrollo

Este capítulo muestra el proceso metodológico de desarrollo del presente trabajo para diseñar y construir un prototipo capaz de gestionar la información del peso de los racimos de banano, adicionalmente predecir el peso del tallo del racimo, por lo cual el capítulo abarca los siguientes puntos:

- Terminología básica
- Requisitos generales del sistema
- Arquitectura del sistema
- Diseño del hardware
  - Subsistema de encendido y regulación de voltaje
  - Subsistema para carga de la batería
  - Subsistema para visualización y panel de control
  - Subsistema para adquisición de datos
  - Subsistema para comunicación inalámbrica
  - Subsistema de control
- Desarrollo de la aplicación para Android
  - Ventana principal
  - Menús y submenús de las aplicaciones.
- Aplicación de los algoritmos de machine learning.
  - Diseño del experimento y pre procesamiento de datos

- Implementación de la red neuronal
- Implementación del regresor de soporte vectorial
- Implementación de regresión lineal multivariable

### 3.1. Terminología básica

Esta sección introduce la terminología básica que se utiliza en las fincas bananeras y sus procesos de producción. Durante el desarrollo del trabajo se hace referencia a estos términos que se muestran en el grupo de datos tomados para la experimentación de los algoritmos de machine learning.

- Racimo: estructura que se conforma de manos de banano que se sostienen de un tallo principal.
- Raquis: conocido también como tallo principal del racimo de banano.
- Mano: conjunto formado por bananos individuales, dichos conjuntos constituyen la estructura general del racimo de banano.
- Convoy: conjunto de racimos de banano sujetos y organizados en forma lineal.
- Garrucha: pieza generalmente de metal de donde se sostiene un racimo de banano.

### 3.2. Requisitos generales del sistema

Los requisitos para implementar el sistema de automatización y gestión de la información en una línea de producción de banano de este trabajo son:

- Capacidad de pesaje, recolección automática de información de los racimos de banano pesados.
- Mantener la comunicación inalámbrica entre el dispositivo de pesaje y aplicación para la gestión de la información.
- Aplicación con interfaz de usuario amigable para la gestión de la información y generación de reportes.
- Determinación de un algoritmo de aprendizaje automático para estimar el peso del raquis.

- Dispositivo para agregar parámetros específicos de cada lote de forma manual.
- Sistema de fácil instalación con una interfaz amigable con el operador.

### 3.3. Arquitectura general del sistema

El sistema de gestión integra la construcción de una balanza digital que se encarga de filtrar y digitalizar los datos del peso neto del racimo de banano, luego esta información es enviada hacia una tableta que contiene una aplicación Android. El esquema general del presente trabajo se resume en la figura 3.1.

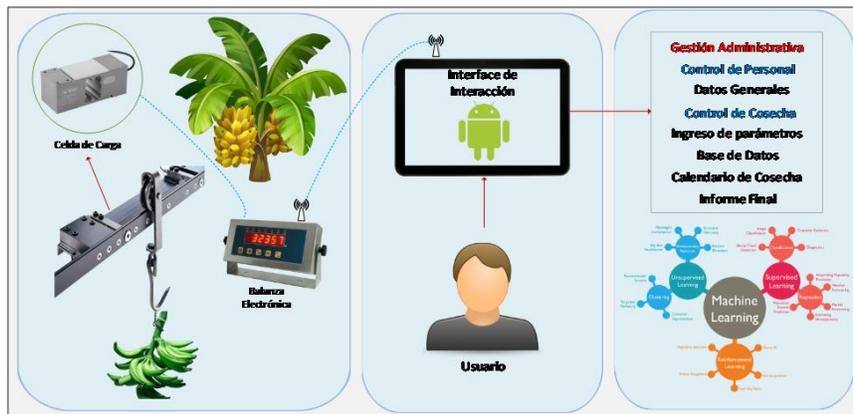


Figura 3.1: Estructura del sistema

La transmisión de datos desde la balanza hacia la aplicación se realizó mediante protocolo bluetooth debido a que permite una comunicación inalámbrica a bajo costo y la transmisión de datos a corta distancia para uso del operador, además, no requiere infraestructura extra para la implementación.

### 3.4. Estructura física del sistema

Luego de haber determinado las características y funcionalidad del sistema de gestión, esta sección muestra los principales componentes para la construcción del prototipo de la balanza digital, además indica sus principales características y circuitos electrónicos.

La balanza electrónica sobre la cual se realizan todos los ensayos y toma de datos de este trabajo se conforma de: chasis, circuitos electrónicos, paneles de control y panel de conexionado. Las secciones 3.4.1 y 3.4.2 muestran la descripción de la elaboración del prototipo.

### 3.4.1. Diseño y construcción del chasis de la balanza electrónica

El chasis alberga en su interior los circuitos electrónicos. El diseño del chasis se elaboró en el software AutoCAD y para su construcción se utilizó el método de impresión 3D con filamento PLA, este método permite crear un objeto tridimensional personalizado. La figura 3.2 muestra la vista en perspectiva del contorno del chasis.

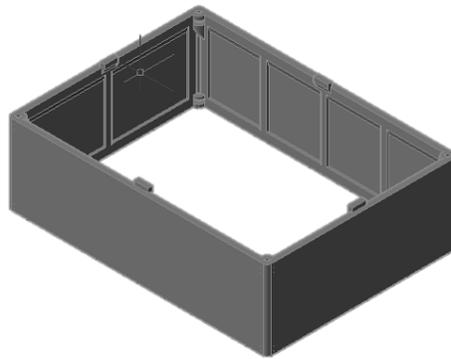


Figura 3.2: Contorno 3D del chasis

La figura 3.3 muestra la vista frontal y posterior del chasis de la balanza, en la parte frontal se aloja el panel de control del usuario, mientras que en la parte posterior se localiza la zona de conexión y alimentación de la balanza.

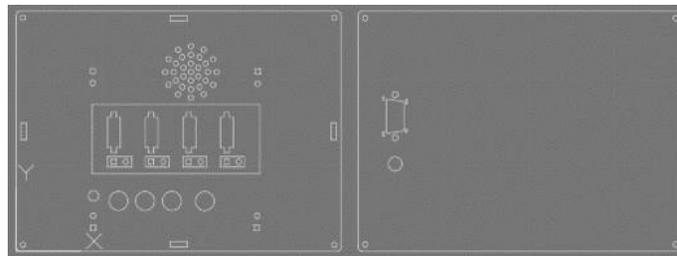


Figura 3.3: Vista frontal (izq) y vista posterior del chasis



El circuito integrado LM2931T es un regulador de baja caída de tensión y bajo ruido que es el encargado de suministrar 5 voltios desde la batería de 6V.

### 3.4.3. Subsistema para carga de la batería

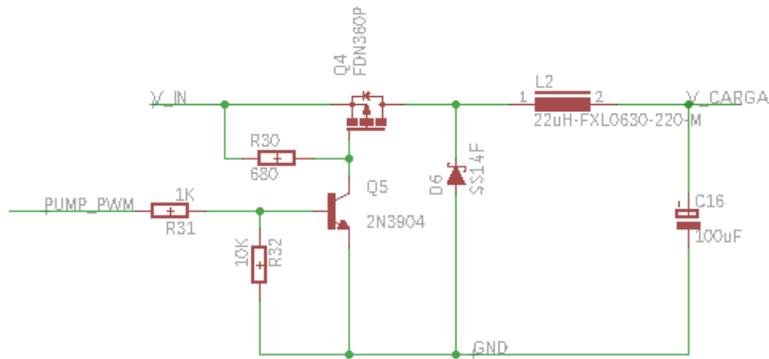


Figura 3.5: Esquema de implementación para cargar la batería

Este sistema es un regulador conmutado tipo buck o de potencia. El transistor Q4 es el encargado de trocear la corriente que pasa hacia el inductor L2 y carga al condensador C16. El diodo de libre circulación D6 conduce la corriente mientras el transistor Q4 está desactivado asegurando una corriente en el inductor para que el circuito trabaje en modo continuo.

El condensador C16 ayuda a mantener constante el voltaje de salida con un mínimo rizado. El transistor Q5 es el encargado de conmutar al transistor Q4 con una señal de control desde el MCU a 100Khz.

### 3.4.4. Subsistema para visualización y panel de control

Este subsistema cuenta con los botones de menú con resistencias de pull-up, un buzzer para generar sonido al momento de presionar una teclay una LCD 4x20 encargada de mostrar toda la interface, los menús e indicaciones, en esta parte del circuito consta con un filtro pasa bajo de un polo con R9 y C5 para controlar el contraste de la pantalla y un transistor T2 para el control de la luz de fono mediante modulación de ancho de pulso (PWM).

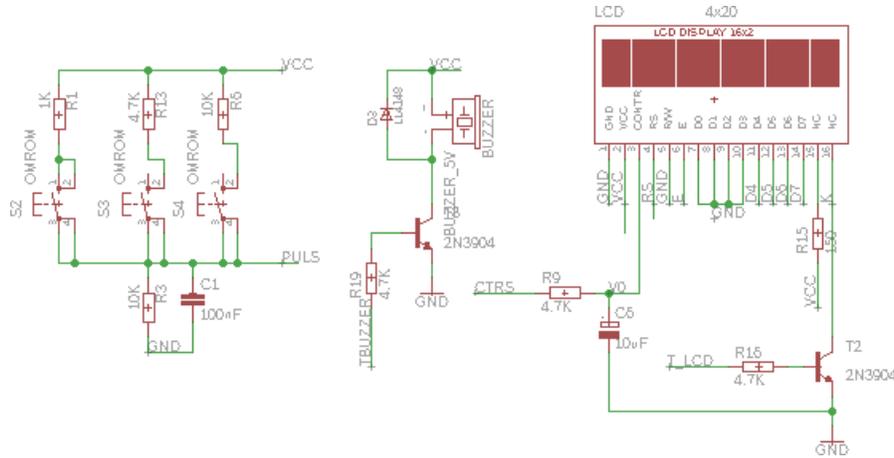


Figura 3.6: Esquema de implementación para el panel de control

### 3.4.5. Subsistema para adquisición de datos

Para adquirir los datos se utiliza un HX711 que es un convertidor ADC diferencial con una resolución de 24bits y salida digital SPI síncrona, este esquema se realizó siguiendo las recomendaciones del fabricante proporcionadas en la hoja de datos.

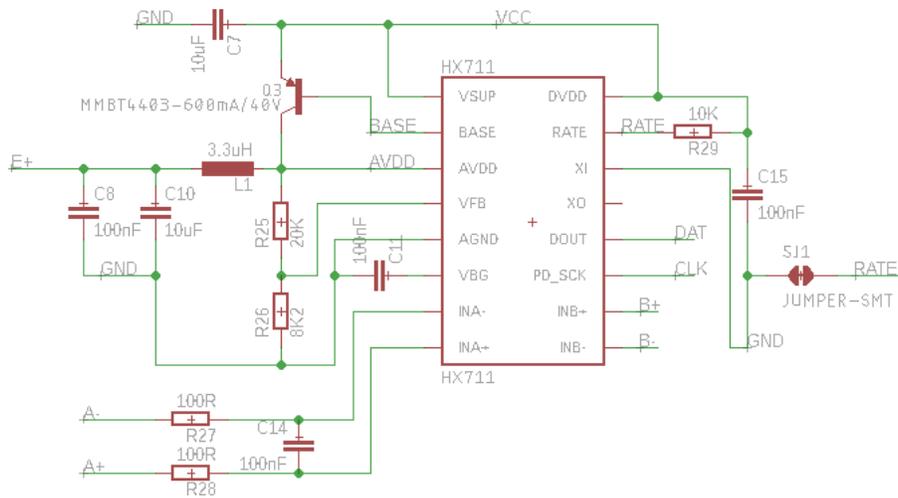


Figura 3.7: Esquema de implementación para la adquisición de datos

### 3.4.6. Subsistema para comunicación inalámbrica

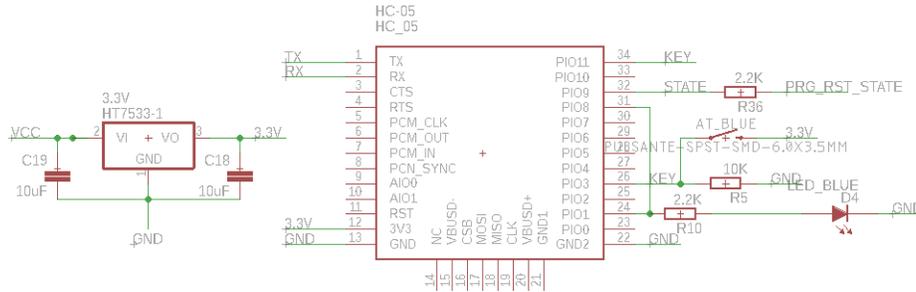


Figura 3.8: Esquema de implementación para la comunicación inalámbrica

La comunicación inalámbrica se realizó mediante el módulo bluetooth HC-05 que requiere una alimentación de 3.3V y 100mA por lo que se utilizó un regulador de baja caída de tensión HT7533.

### 3.4.7. Subsistema de control

La figura 3.9 muestra el esquema del circuito implementado. El sistema está gobernado por un microcontrolador AVR de 8bits Atmega328P corriendo a 16Mhz de frecuencia del reloj principal. Este microcontrolador se encarga de gestionar seis tareas generales:

- El encendido y apagado de todo el sistema.
- El sistema de carga de la batería que incluye la lectura de los valores de voltaje junto con el control de la fuente conmutada buck.
- La adquisición de datos del pesaje mediante el bus SPI síncrono.
- La transmisión de datos inalámbrica mediante el bus UART asíncrono.
- La interface humano máquina (botones de configuración y pantalla de visualización LCD).
- Cuenta con memoria EEPROM para el almacenamiento de la configuración de los parámetros del sistema (Brillo, Contraste, Sonido, Peso Patrón, entre otros).

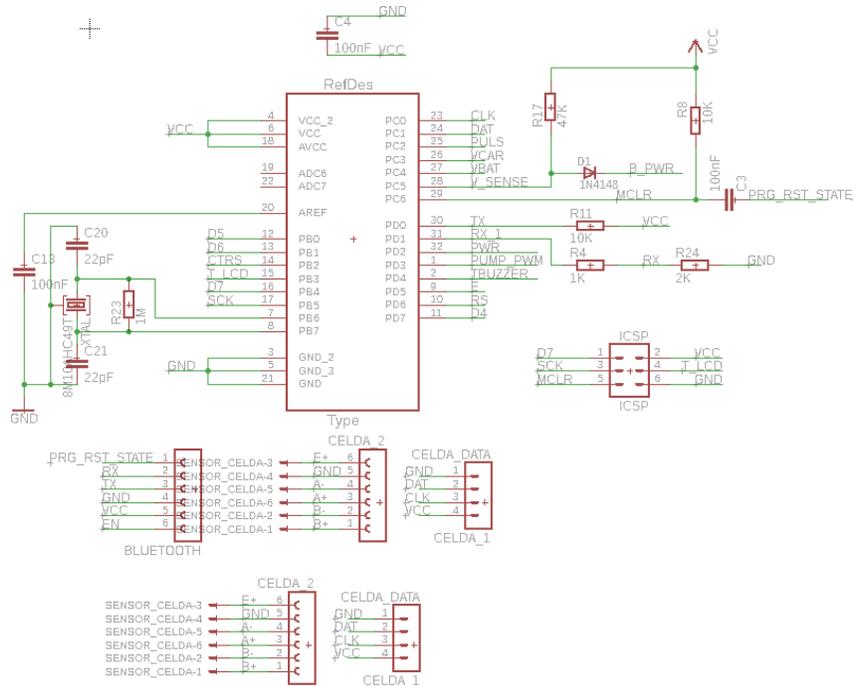


Figura 3.9: Esquema de implementación para el control del sistema

### 3.5. Desarrollo de la aplicación

Esta sección describe la funcionalidad de la aplicación para la gestión de la información, la cual se desarrolló en Android debido a su adaptabilidad incluso con dispositivos electrónicos de gama baja. La interfaz de la aplicación consta de seis partes o submenús, que se muestran en las secciones 3.5.1 a 3.5.6.

#### 3.5.1. Ventana principal

La ventana principal de la interfaz, es aquella que se visualiza al iniciar el proceso de pesaje. Este menú gestiona 5 tareas organizadas en submenús, el flujo de funcionamiento se muestra en la figura 3.10.

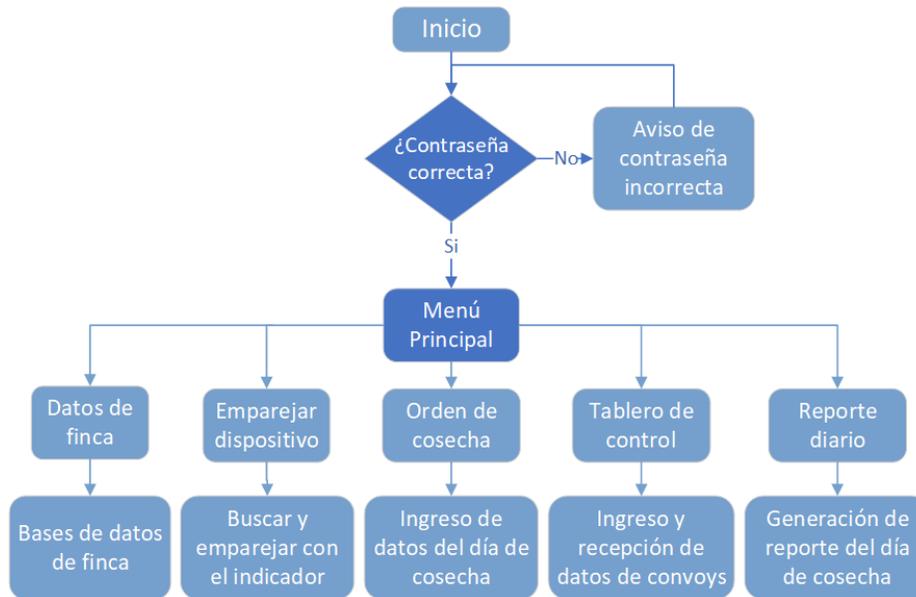


Figura 3.10: Diagrama de funcionamiento de la ventana principal del sistema para el control del dispositivo.

### 3.5.2. Datos de la finca

Este submenú permite gestionar la información de la finca en donde se realizará el proceso de pesaje. En caso de que los datos de la finca estén registrados, entonces se podrán cargar los datos antes de iniciar el pesaje.

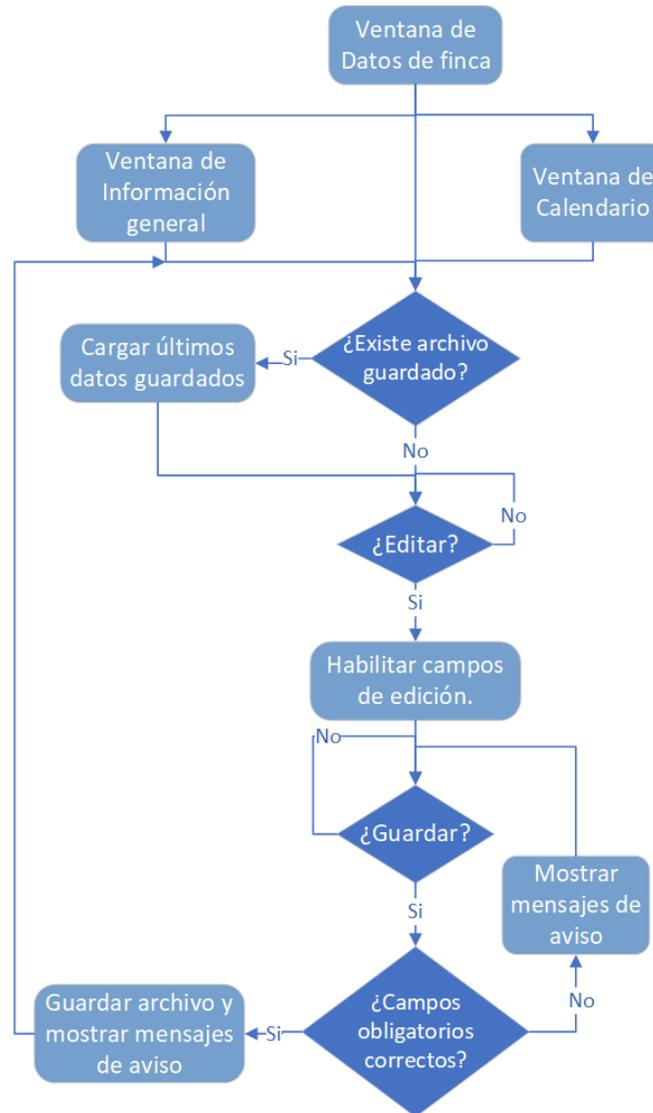


Figura 3.11: Diagrama de funcionamiento del submenú de la aplicación (datos de la finca)

### 3.5.3. Emparejar dispositivo

El submenú emparejar dispositivo se encarga de iniciar la comunicación inalámbrica mediante bluetooth entre la balanza digital y la aplicación instalada en la tablet. El flujo de funcionamiento del submenú de emparejar

dispositivo se muestra en la figura 3.12.

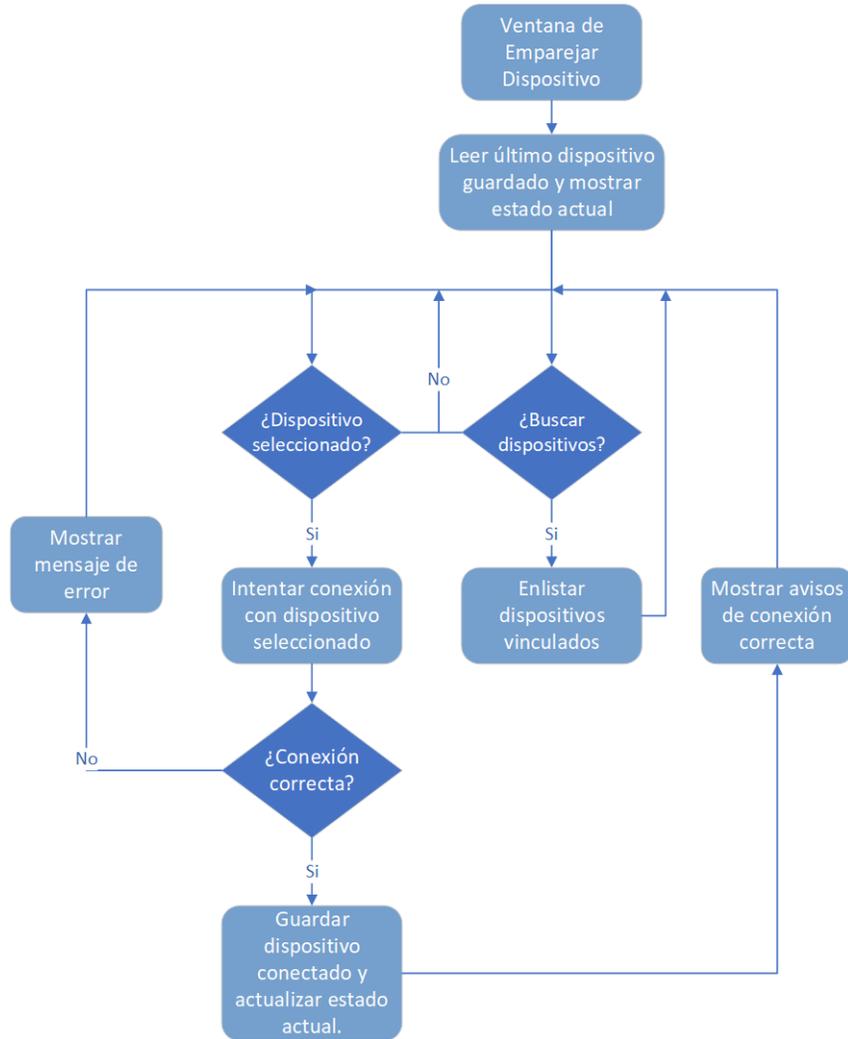


Figura 3.12: Diagrama de funcionamiento del submenú de la aplicación (emparejar dispositivo)

### 3.5.4. Orden de cosecha

Este submenú permite cargar la información del lote de cosecha de banano que comprende: la fecha de pesaje, la semana de cosecha, el número de orden, los datos de la hacienda, el número de la empacadora, el peso bruto, el peso

base. El diagrama de funcionamiento del se muestra en la figura 3.13

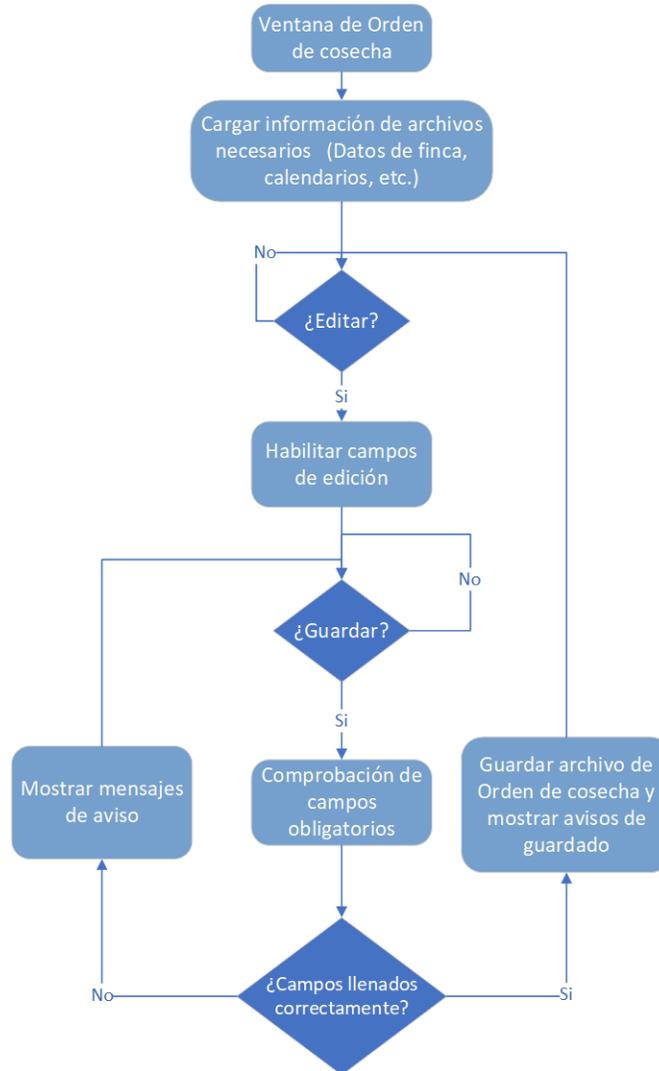


Figura 3.13: Diagrama de funcionamiento del submenú de la aplicación (orden de cosecha)

### 3.5.5. Reporte Diario

Esta ventana permite visualizar la información de pesaje de los racimos de banano y de la finca. En este submenú podemos descargar los datos en

un archivo xls. La funcionalidad se muestra en la figura 3.14.

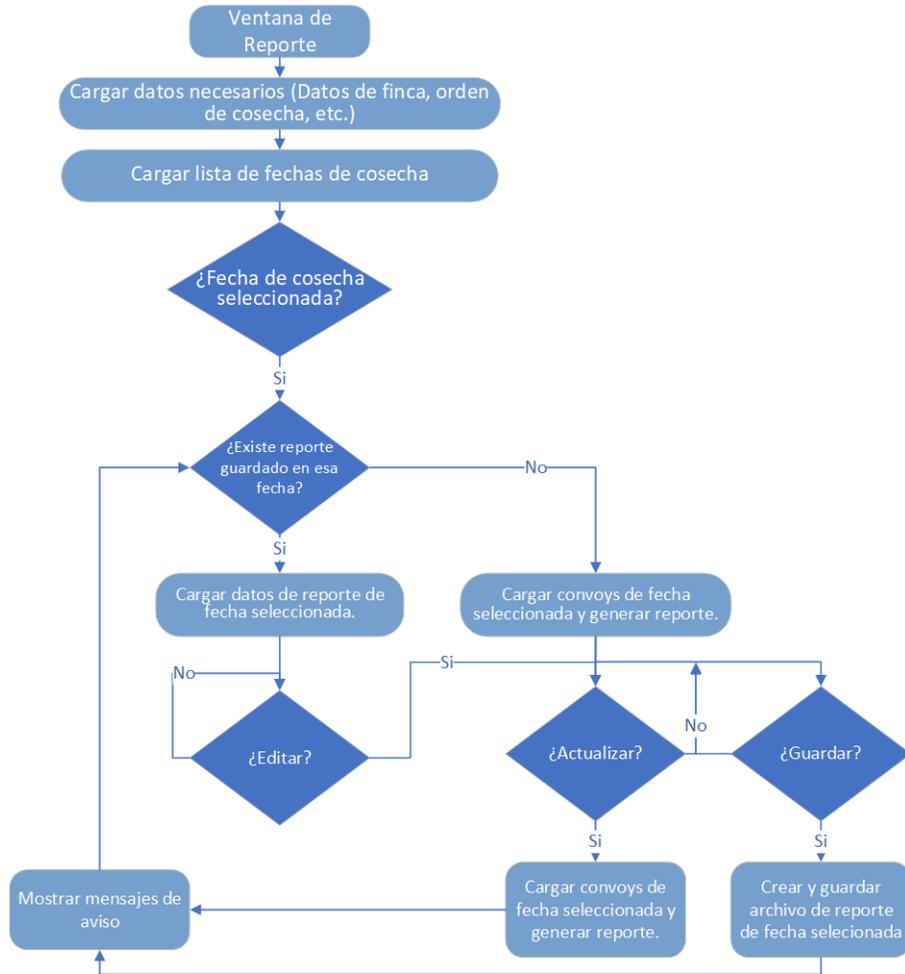


Figura 3.14: Diagrama de funcionamiento del submenú de la aplicación (reporte diario)

### 3.5.6. Tablero de control

Muestra la información del peso de los racimos de banano, datos de la finca y permite descargar los datos en un archivo xls.

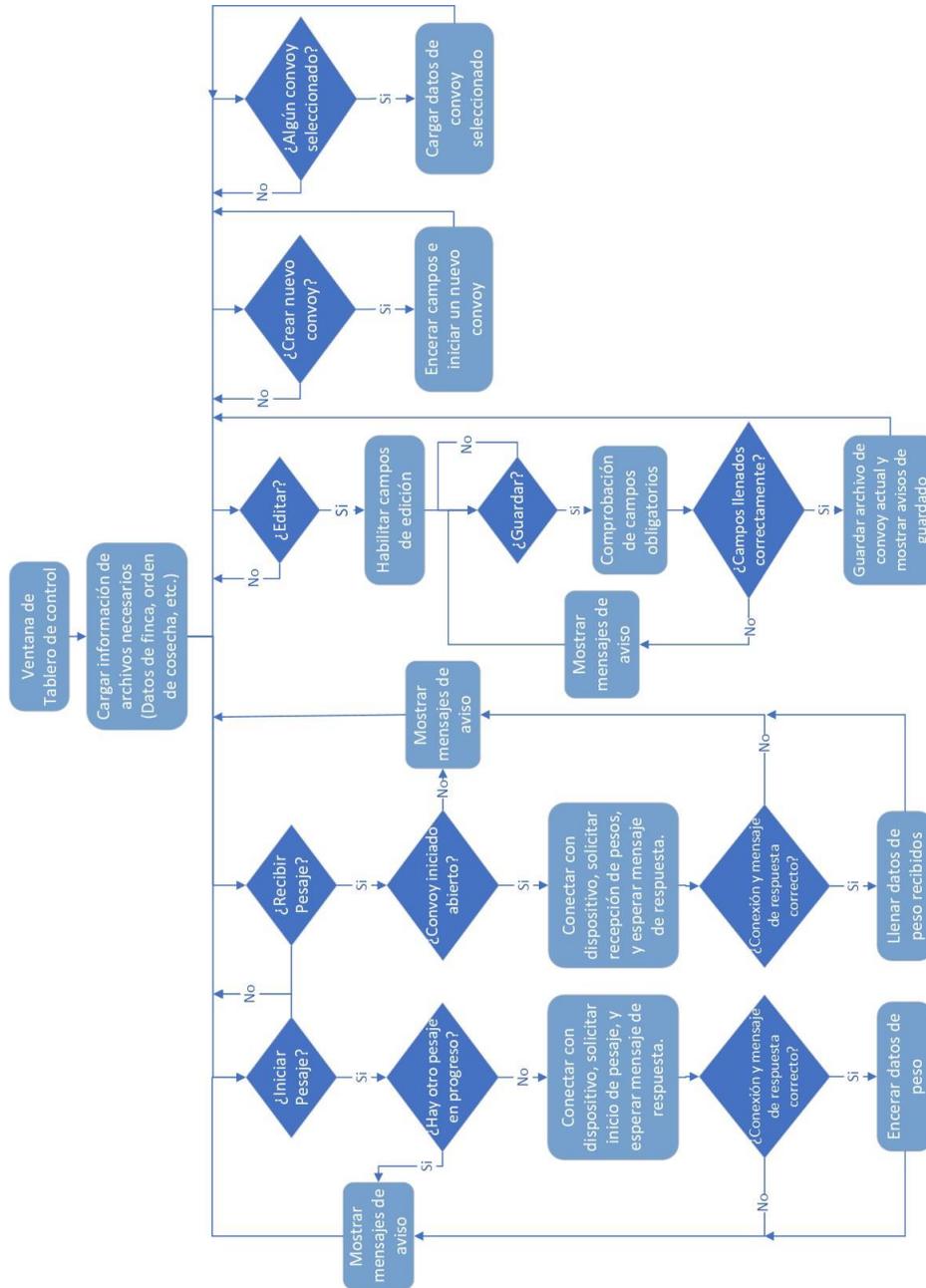


Figura 3.15: Diagrama de funcionamiento del submenú de la aplicación (tablero de control)

### 3.6. Aplicación de los algoritmos de machine learning

En esta sección se realizó la comparación del desempeño de 3 algoritmos de machine learning. El problema del presente trabajo es una tarea de regresión por tanto se eligió tres algoritmos para su desarrollo: redes neuronales, regresión de soporte vectorial y regresión lineal multivariable. Para los experimentos que se muestra en la siguiente sección se ha utilizado el producto de Google Reseach conocido como “Colaboratory” el cual permite programar en phyton.

#### 3.6.1. Diseño del experimento

Para la implementación se consideraron las características de los datos tomados de (Enriquez Tovar y Vega Caiza, 2011), enlace: <http://repositorio.utc.edu.ec/handle/27000/3541>

Nombre del dataset: "Datos\_Racimos.csv"

Número de variables o atributos: 8 debido a la naturaleza de los datos tomados del trabajo de (Enriquez Tovar y Vega Caiza, 2011), las cuales se detallan a continuación:

- Peso del Racimo en kilogramos
- Número de Manos del Racimo
- Promedio de Grado de dedos de última mano
- Promedio de Grado de dedos de segunda mano
- Promedio de longitud de dedo última mano
- Promedio de longitud de dedo segunda mano
- Peso del Rechazo (Kg)
- Peso del Raquis (Kg)

Salida: predicción peso del racimo. La tabla 3.1 muestra el tamaño y la composición de la base de datos para el experimento.

|   | <b>Número de muestras</b> | <b>Media</b> | <b>Máximo</b> | <b>Mínimo</b> |
|---|---------------------------|--------------|---------------|---------------|
| <b>Peso del Racimo (Kg)</b>                       | 190                       | 25.96        | 13            | 44            |
| <b>Número de Manos del Racimo</b>                 | 190                       | 7.04         | 5             | 9             |
| <b>Promedio de Grado de dedos de última mano</b>  | 190                       | 39.34        | 37            | 44            |
| <b>Promedio de Grado de dedos de segunda mano</b> | 190                       | 43.45        | 39            | 47            |
| <b>Promedio de longitud de dedo última mano</b>   | 190                       | 8.23         | 7             | 10            |
| <b>Promedio de longitud de dedo segunda mano</b>  | 190                       | 10.05        | 1.03          | 11.5          |
| <b>Peso del Rechazo (Kg)</b>                      | 190                       | 0.71         | 0             | 3.3           |
| <b>Peso del Raquis (Kg)</b>                       | 190                       | 2.34         | 1.2           | 3.9           |

Tabla 3.1: Descripción del dataset.

### 3.6.2. Preprocesamiento de datos

Estandarización y preprocesamiento de datos: en este punto las variables de entrada numéricas se escalan a un rango estándar con lo cual se busca mejorar el aprendizaje automático que utilizan los algoritmos de regresión, de manera que la media de los datos en cada variable sea igual a cero y la varianza sea uno. Esta diferencia de escala para las variables de entrada afecta a los algoritmos de aprendizaje automático. El preprocesamiento consiste en un cambio de escala lineal de las variables de entrada mediante el objeto `MinMaxScaler` en el rango  $[0,1]$ .

Para evitar fluctuaciones en los resultados de los experimentos se ha utilizado una validación cruzada `KFolds`, en donde  $K=5$ , debido a que utilizó la clase `GridSearchCV` que permite seleccionar de manera automática los

parámetros del modelo.

### 3.6.3. Implementación de la red neuronal

#### Diseño de la red neuronal

El modelo inicial de la red neuronal consta de la capa de entrada que recibe 8 parámetro, 2 capas ocultas, la primera capa tiene una densidad de 8 neuronas y la segunda una densidad de 32 neuronas mientras que la capa de salida tiene una neurona. La primera red se creó con 100 iteraciones (“epochs”), un tamaño de lote por muestras (batch 35). La figura 3.16 muestra la función de pérdida del primer modelo sin optimizar de la red neuronal.

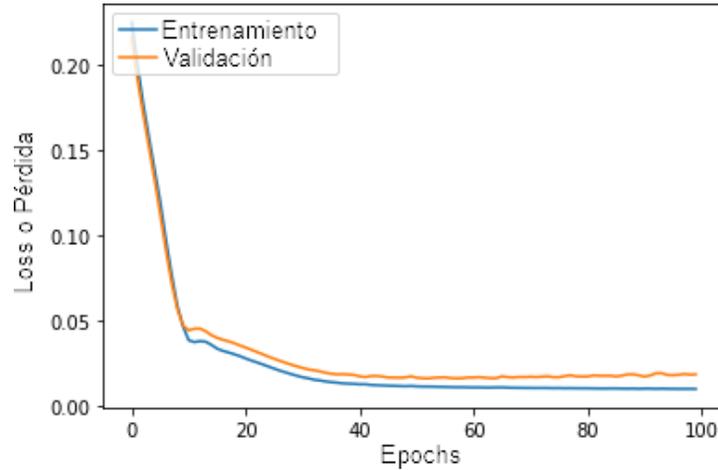


Figura 3.16: Gráfica de la función de pérdida del modelo inicial

#### Optimización

Partimos del objetivo del entrenamiento de redes neuronales es encontrar los pesos adecuados para los cuales la red trabaje de manera más generalizada. La optimización se realizó en tres etapas, este proceso de

- **Primera etapa:** optimización en la compilación

En esta etapa se evaluó el comportamiento de la red modificando los parámetros del tamaño de batch, epochs y el algoritmo de optimización Adam y Adadelta. La tabla 3.2 muestra los valores de los parámetros

que se evaluaron, la mejor combinación fue de lote 30, épocas 300 y el algoritmo Adam.

|               |    |     |     |     |
|---------------|----|-----|-----|-----|
| <b>Epochs</b> | 50 | 100 | 200 | 300 |
| <b>Batch</b>  | 5  | 20  | 30  | 40  |

Tabla 3.2: Valores de prueba para la optimización por compilación de la red neuronal.

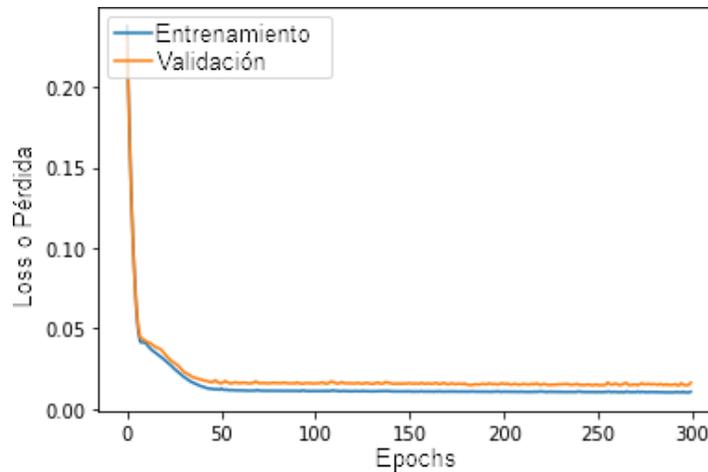


Figura 3.17: Gráfica de la función de pérdida luego de la primera etapa de optimización

- **Segunda etapa:** optimización de las capas neuronales.

En esta etapa se modifica la cantidad de neuronas de la capa oculta. El número de neuronas a evaluarse se muestran en las capas de entrada y oculta se muestran en la tabla 3.3 y el resultado de la combinación para la optimización fue de 64 neuronas en la primera capa y 8 en la segunda capa.

|                            |   |    |    |     |     |
|----------------------------|---|----|----|-----|-----|
| <b>Primera capa oculta</b> | 8 | 16 | 64 | 128 | 256 |
| <b>Segunda capa oculta</b> | 8 | 16 | 32 | 64  | 256 |

Tabla 3.3: Cantidad de neuronas en cada capa la optimización de la red neuronal

- **Tercera etapa:** proceso con dropouts.

En esta etapa se redujo la densidad en las capas ocultas, la tabla 3.4 muestra el porcentaje de neuronas apagadas, el resultado de la optimización fue un 10 % de neuronas desactivadas en la segunda capa oculta.

|                            |     |      |      |      |
|----------------------------|-----|------|------|------|
| <b>Primera capa oculta</b> | 0 % | 10 % | 20 % | 30 % |
| <b>Segunda capa oculta</b> | 0 % | 10 % | 20 % | 30 % |

Tabla 3.4: Porcentaje de neuronas apagadas en cada capa para la optimización por apagado de neuronas.

### Valores de predicción y reales.

| Valor real | Valor de predicción |
|------------|---------------------|
| 2.00       | 1.96                |
| 2.30       | 2.17                |
| 1.60       | 2.06                |
| 2.50       | 2.37                |
| 2.60       | 2.88                |
| 1.90       | 1.87                |
| 3.70       | 3.63                |
| 3.60       | 3.52                |
| 1.70       | 1.89                |
| 2.20       | 2.00                |

Tabla 3.5: Valores del peso obtenidos mediante el algoritmo de Red Neuronal Optimizado

Luego de realizar el entrenamiento, la tabla 3.5 expone una muestra de 10 valores de los pesos de los raquis tomados de la base de datos [Enriquez Tovar y Vega Caiza \(2011\)](#) y los valores de predicción de la red neuronal optimizada. El error cuadrático RMSE del algoritmo fue de 0.01867 y el coeficiente de determinación R<sup>2</sup> de predicción 0.4776. En el capítulo 4 se aborda un análisis de los resultados obtenidos de los 3 algoritmos implementados en el presente trabajo.

### 3.6.4. Implementación del regresor de soporte vectorial

#### Diseño de SVR

En el trabajo se implementó un regresor de soporte vectorial, por ello en primer lugar se incluye los parámetros del algoritmo de la librería Scikit Learn. La tabla 3.6 muestra un resumen de los parámetros que ajustan el modelo, es decir controlan el proceso de aprendizaje se conocen con el nombre de hiperparámetros. Luego se llama al método fit, en este caso para el entrenamiento se tomó el 80 % de la información de la base de datos. Los datos tuvieron una etapa de preprocesamiento para escalar las variables de entrada mediante la función MinMaxScaler.

|                       |       |
|-----------------------|-------|
| <b>Tipo de Kernel</b> | poly  |
| <b>Degree</b>         | 3     |
| <b>Gamma</b>          | 10    |
| <b>C</b>              | 0.001 |

Tabla 3.6: Hiperparámetros para la implementación del regresor de soporte vectorial

#### Valores de predicción y reales

| Valor real | Valor de predicción |
|------------|---------------------|
| 2.80       | 2.57                |
| 1.70       | 1.89                |
| 3.60       | 3.44                |
| 1.90       | 1.95                |
| 2.60       | 2.33                |
| 3.80       | 3.67                |
| 2.80       | 2.21                |
| 2.00       | 2.14                |
| 3.30       | 3.11                |
| 1.80       | 1.81                |

Tabla 3.7: Valores del peso obtenidos mediante el algoritmo Vector de soporte para regresión

Luego de realizar el entrenamiento, la tabla 3.7 expone una muestra de 10 valores de los pesos de los raquis tomados de la base de datos [Enriquez Tovar y Vega Caiza \(2011\)](#) y los valores de predicción según el algoritmo de vector

de soporte para regresión. El error cuadrático RMSE del algoritmo fue de 0.0082 y el coeficiente de determinación R2 de predicción 0.7969. En el capítulo 4 se aborda un análisis de los resultados obtenidos de los 3 algoritmos implementados en el presente trabajo.

### 3.6.5. Implementación del algoritmo de regresión multilínea

El modelo de regresión intenta encontrar los coeficientes más óptimos para todos los atributos. En el trabajo se implementó el algoritmo de regresión lineal multivariable, en primer lugar, se utilizó la clase LinearRegression de la librería Scikit Learn. Luego llamamos al método fit para ajustar realizar el entrenamiento son el 80 % de la información de la base de datos. Se obtuvo un RMSE de 0.011246 y un R2 de 0.755332. La tabla 3.8 muestra 10 valores de predicción del algoritmo en comparación con los valores reales.

| Valor real | Valor de predicción |
|------------|---------------------|
| 3.20       | 2.69                |
| 1.80       | 1.97                |
| 1.40       | 1.43                |
| 3.80       | 3.72                |
| 2.30       | 2.38                |
| 2.20       | 2.42                |
| 1.80       | 2.08                |
| 2.00       | 1.58                |
| 2.50       | 2.37                |
| 1.90       | 2.59                |

Tabla 3.8: Valores del peso obtenidos mediante el algoritmo Vector de soporte para regresión

### 3.6.6. Sumario

El capítulo 3 abordó los temas para que permiten conocer la metodología para desarrollar el sistema de gestión, además muestra los componentes que lo integran. En cuanto a la construcción del prototipo de balanza digital fue posible conocer su chasis y circuitería interna.

Este capítulo orienta a conocer las múltiples funcionalidades de la aplicación de Android, los diferentes menús y modo de operación, por ejemplo: este posee un menú de ingreso, de registro de datos de la cosecha, el manejo para el pesaje automático y la generación de reportes en formato xls.

Finalmente, el capítulo mostró el diseño del experimento para implementar 3 algoritmos de machine learning. Aquí fue posible la optimización en tres etapas de la red neuronal. En el siguiente capítulo se mostrarán los resultados de la metodología implementada a nivel físico, del desarrollo de la aplicación y algoritmos de machine learning para el problema de regresión de este trabajo.

## Capítulo 4

# Resultados

En este capítulo se muestra la interfaz del tablero de control de la aplicación para la gestión de la información, junto con prototipo de balanza digital para recolección de datos y el diseño de reportes que se visualizan en la aplicación. Además, se exponen la comparación de rendimiento de los algoritmos de aprendizaje supervisado para estimar el peso del raquis o tallo del racimo de banano.

### 4.1. Montaje Físico

La figura 4.1 muestra el prototipo de balanza electrónica ensamblado y funcional.



Figura 4.1: Balanza electrónica

## 4.2. Sistema de gestión

La aplicación desarrollada para el sistema de gestión nos muestra una interfaz amigable que inicia con la ventana de ingreso al sistema que se muestra en la figura 4.2



Figura 4.2: Ventana de ingreso a la aplicación

La figura 4.3 muestra la ventana de control del sistema de control que permite manejar el inicio de pesaje, que comprende la adquisición y envío de datos en forma inalámbrica entre la balanza digital y la aplicación.



Figura 4.3: Ventana del tablero de control de la aplicación

En cuanto a la sección de reportes la aplicación organiza la información recopilada e ingresada en ocho tipos de reportes: de cosecha por lote, de cosecha por jornada, de recobro, de parámetros de producción, de racimos rechazados, de control de personal, de convoy, de garruchas. Estos reportes pueden ser exportados a un documento xls. La figura 4.4 muestra la interfaz de la ventana para el manejo de los reportes.



Figura 4.4: Ventana de generación de reportes de la aplicación

El desarrollo del prototipo de pesaje integrado a la aplicación permitió el manejo de la información de la línea de producción de banano.

### 4.3. Algoritmos de aprendizaje supervisado

#### 4.3.1. Optimización de la red neuronal

En primer lugar esta sección muestra la tabla 4.1 que expone un resumen de los parámetros iniciales de la red neuronal y parámetros resultantes de la optimización detallada en la sección 3.5.3 del documento.

|  | <b>Red neuronal inicial</b> | <b>Red neuronal optimizada</b> |
|--|-----------------------------|--------------------------------|
| Epochs                                       | 100                         | 300                            |
| Tamaño del batch                             | 35                          | 30                             |
| Optimizador                                  | Adam                        | Adam                           |
| Densidad de la capa de entrada               | 8 neuronas                  | 8 neuronas                     |
| Densidad primera capa oculta (I1)            | 8 neuronas                  | 64 neuronas                    |
| Densidad de la capa segunda capa oculta (I2) | 32 neuronas                 | 8 neuronas                     |
| Densidad de la capa de salida                | 1 neurona                   | 1 neurona                      |

Tabla 4.1: Parámetros de la red neuronal original y optimizada

La figura 4.5 muestra la visualización mediante un mapa de calor (heatmap) con la técnica Grid Search para establecer el parámetro de puntuación verificando las métricas del coeficiente de determinación ( $R^2$ ).

Mientras este coeficiente es más alto la calidad es mejor. En el mapa se demuestra que los valores óptimos son: 64 neuronas en la capa I1 y 8 neuronas en la capa I2.

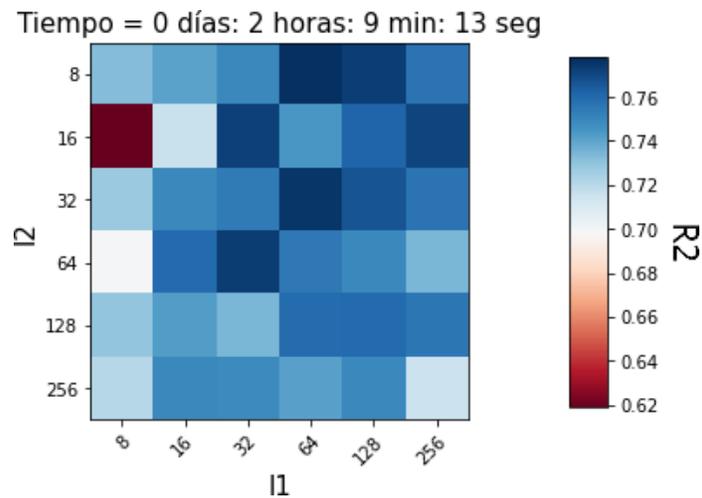


Figura 4.5: Visualización mediante un mapa de calor

### 4.3.2. Rendimiento de los algoritmos

Para evaluar el rendimiento de los algoritmos implementados en el sistema se utilizó el cálculo del error cuadrático medio (RMSE) que permitió comparar los valores de predicción y los reales, también se calculó el coeficiente de determinación (R2) y el método convencional (MCO) que es una métrica usada en la logística de la empresa que estima que el peso del raquis es aproximadamente el 10 % del peso total del racimo. Los resultados de muestran en la tabla 4.2

| <b>Algoritmo</b>                         | <b>RMSE</b> | <b>R2</b> |
|--|-------------|-----------|
| Algoritmo de redes neuronales optimizado | 0.0117      | 0.5901    |
| Regresión de soporte vectorial           | 0.0082      | 0.7969    |
| Regresión multivariable                  | 0.0112      | 0.7553    |
| Método convencional                      | 0.2130      | 0.5726    |

Tabla 4.2: Métricas para el rendimiento de los algoritmos.

El primer algoritmo utilizado fue una red neuronal artificial, una desventaja de las redes neuronales es el tiempo que toma el entrenamiento y ajuste; para el problema de regresión el tiempo utilizado en el entrenamiento fue de 5.15 segundos dada la baja cantidad de datos; se utilizó una estrategia de optimización con un algoritmo que facilita las diferentes combinaciones de batch\_size, epochs, optimizer, densidad de cada capa, dropouts, logrando un tiempo de 10 minutos, esto ayudó a reducir el tiempo de optimización que manualmente tomaría entre 3 a 5 veces más dependiendo de la habilidad del usuario; luego de haber superado este obstáculo se puede observar que el RMSE para la red optimizada es de 0.0117 comparado con el método convencional donde se obtiene un RMSE de 0.212 por lo que el algoritmo ha mejorado significativamente la precisión y puede realizar mejores predicciones.

El segundo algoritmo utilizado fue un regresor de soporte vectorial, este algoritmo demostró un costo de tiempo de entrenamiento mucho menor a la red neuronal y fue el mejor en predicción con un RMSE de 0.0082 comparado con el RMSE 0.0117 de la red neuronal.

El tercer algoritmo utilizado fue una regresión lineal múltiple consiguiendo un RMSE de 0.0112, lo cual la deja en tercer lugar con respecto a la red neuronal y la máquina de soporte vectorial.

De la tabla 4.2 se puede identificar que el algoritmo idóneo para aplicar a nuestro trabajo fue el de Regresión de soporte vectorial, debido a que el

coeficiente  $R^2$  es 0,79 siendo lo ideal 1, mientras que el error cuadrático medio es el menor obtenido de los tres algoritmos.

### 4.3.3. Sumario

El capítulo 4 permitió visibilizar los resultados que se obtuvieron de la implementación del sistema de gestión y la experimentación para los algoritmos de machine learning para el problema de regresión del presente trabajo. La sección 4.1 mostró el resultado del ensamblaje de la balanza electrónica con sus respectivos circuitos, la sección 4.2 presentó la interfaz y funcionalidades del sistema; la sección 4.3 se enfocó en mostrar los resultados de un modelo optimizado de la red neuronal incluyendo construcción de métricas y heatmap, finalmente se presentó el rendimiento de los algoritmos implementados en el capítulo 3.

Gracias a estos puntos fue posible conocer las funcionalidades del sistema propuesto para la gestión de la información, a nivel físico fue posible la recolección de datos de manera automática mediante la balanza digital que envía los datos hacia la aplicación desarrollada, estos datos son procesados y convertidos en información útil mediante la generación de reportes en archivos de tipo xls.

Al observar el Error cuadrático medio (RMSE) de la tabla 4.2 es un valor bajo, esto se debe a que la predicción del peso del raquis es un aprendizaje pasivo pues se recolectan la información de una base de datos y no en tiempo real utilizando el prototipo de balanza construido, además los datos de muestra del problema son pocos debido a la base de datos utilizada para el experimento.

Finalmente al inspeccionar los resultados de la tabla 4.2 que corresponden a las métricas para el rendimiento de los algoritmos se pudo notar el algoritmo más conveniente para el problema de predicción abordado y basado en la base de datos es el de regresor de soporte vectorial.

## Capítulo 5

# Conclusiones y trabajo futuro

Este capítulo muestra como concluye las tareas realizadas para cumplir los objetivos del trabajo, además resalta las contribuciones logradas con el desarrollo del trabajo: la automatización en el proceso de pesaje en la línea de producción de banano, la gestión de la información y la comparación del rendimiento de 3 algoritmos de machine learning. Al final del capítulo se plantea el trabajo futuro como siguientes pasos a partir de este trabajo.

### 5.1. Conclusiones

En este trabajo se presentó el desarrollo de la implementación de un sistema de gestión de automatización y gestión de la información en una línea de producción de banano, que permite la adquisición y envío de datos de forma automática desde un prototipo de balanza electrónica hacia un dispositivo que posee una plataforma móvil. Se mostró la estructura física, desarrollo de la aplicación móvil para el sistema, gracias a esto se automatizó el proceso de pesaje.

En la aplicación se gestionan los datos recolectados e ingresados por el usuario, posteriormente los datos se convierten en información útil gracias a la generación de reportes. Los reportes se exportan a un archivo en formato xls, de esta manera se logró cumplir el objetivo general.

En cuanto a la estimación del peso de la cabeza de banano, gracias a la implementación de tres algoritmos supervisados de machine learning se determinó la predicción del raquis o tallo del racimo de banano, vale destacar que la predicción fue un aprendizaje pasivo pues se utilizó información de una base de datos y no en tiempo real en la planta de producción de banano.

De la sección de 4.3.2 referente al rendimiento de los algoritmos se concluye

que: las redes neuronales pueden trabajar mejor cuando estamos frente a una gran cantidad de datos en donde necesitamos realizar un aprendizaje profundo; por otro lado, las máquinas de soporte vectorial funcionan muy bien en problemas donde hay altas o bajas cantidades de datos sobre todo en la clasificación de imágenes, para el problema de regresión planteado en un ambiente de pocos datos las máquinas de soporte vectorial han superado a las redes neuronales dando como resultado un buen desempeño con pocos datos; el algoritmo de regresión Multivariable es el que peor se desempeñó en este problema aunque muy por encima del método convencional por lo que aplicar cualquier algoritmo de machine learning trae consigo un beneficio para la predicción del peso del raquis y por consiguiente la mejor aproximación logística y beneficio económico ya que este es uno de los parámetros que más influye en el cálculo de las cosechas.

## 5.2. Trabajo futuro

El uso de la tecnología en la agricultura podría contribuir para mejorar las condiciones de vida de las personas en relación a combatir la pobreza extrema (FAO, 2016) por ello es necesario automatizar los procesos en el área de producción.

El presente trabajo es un ejemplo de cómo automatizar el proceso de pesaje en una línea de producción de banano, sin embargo, se puede trabajar en otras líneas de desarrollo planteadas a continuación:

- Determinar el material y técnicas adecuadas para construir la estructura física de la balanza.
- Desarrollar aplicaciones de gestión de la información para dispositivos móviles con sistema operativo iOS.
- Implementar algoritmos de machine learning que permitan la predicción del peso de la cabeza o racimo de banano mediante recolección de datos en tiempo real.
- Integración de los datos de predicción mediante machine learning del peso del racimo a la plataforma móvil.
- Realizar un aprendizaje profundo mediante redes neuronales con un gran volumen de datos (big data).

## Referencias

- AEBE. (2020). Exportación de banano, mayo 2020. *Revista Bnanotas*(143), 7.
- Atzori, A. y. M. G., Luigi y Iera. (2010). The internet of things: A survey. *Computer networks*, 54 (15), 2787–2805.
- Awad, M., y Khanna, R. (2015). Support vector machines for classification. En *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (pp. 39–66). Berkeley, CA: Apress. Descargado de [https://doi.org/10.1007/978-1-4302-5990-9\\_3](https://doi.org/10.1007/978-1-4302-5990-9_3) doi: 10.1007/978-1-4302-5990-9\_3
- Chowdhary, K. R. (2020). Machine learning. En *Fundamentals of artificial intelligence* (pp. 375–413). New Delhi: Springer India. Descargado de [https://doi.org/10.1007/978-81-322-3972-7\\_13](https://doi.org/10.1007/978-81-322-3972-7_13) doi: 10.1007/978-81-322-3972-7\_13
- De Oliveira, M. y. S. V., Tiago H Moreira y Painho, y Sian, A., Otávio y Barriguinha. (2014). Development of an agricultural management information system based on open-source solutions. *Procedia Technology*, 16, 342–354.
- Enriquez Tovar, L. E., y Vega Caiza, F. H. (2011). Evaluacion de seis tipos de protectores en el enfunde de racimo de banano (musa sapientum.)en la finca manguila del canton la mana año 2010-2011.
- FAO. (2016). El estado 2016 la agricultura y mundial de la alimentación agricultura y seguridad cambio climático, alimentaria.
- Firouzi, B. y. W. M. y. D. G. y. A. F. S., Farshad y Farahani. (2020). Iot fundamentals: Definitions, architectures, challenges, y promises. En *Intelligent internet of things* (pp. 3–50). Springer.
- González, V., Carlos Andrés Guailazaca y Hernandez. (2020). Clasificador de productos agrícolas para control de calidad basado en machine learning e industria 4.0. *Revista Perspectivas*, 2(2), 21–28.
- Gupta, P. (2021). *Introduction to machine learning in the cloud with python: Concepts and practices*. Springer Nature.

- Gupta, P., y Sehgal, N. K. (2021). Machine learning algorithms. En *Introduction to machine learning in the cloud with python: Concepts and practices* (pp. 23–77). Cham: Springer International Publishing. Descargado de [https://doi.org/10.1007/978-3-030-71270-9\\_2](https://doi.org/10.1007/978-3-030-71270-9_2) doi: 10.1007/978-3-030-71270-9\_2
- Haroon, D. (2017). Regression. En *Python machine learning case studies* (pp. 45–94). Springer.
- Hurtado, R., Bobadilla, J., Bojorque, R., Ortega, F., y Li, X. (2018). A new recommendation approach based on probabilistic soft clustering methods: A scientific documentation case study. *IEEE Access*, 7, 7522–7534.
- Joshi, A. V. (2020). Decision trees. En *Machine learning and artificial intelligence* (pp. 53–63). Cham: Springer International Publishing. Descargado de [https://doi.org/10.1007/978-3-030-26622-6\\_6](https://doi.org/10.1007/978-3-030-26622-6_6) doi: 10.1007/978-3-030-26622-6\_6
- Lee, D.-W., Seong-Hoon y Lee. (2015). Review on current situations for internet of things. En *2015 7th international conference on multimedia, computer graphics and broadcasting (mulgrab)* (pp. 19–21).
- Lombeida, J. V. A. y. o., María Dolores Quintana y Herrera. (2020). Desarrollo de cultivos sostenibles de vainilla en ecuador. *Revista de Investigación Talentos*, 7(1), 71–79.
- Malviya, A. (2020). Machine learning: An overview of classification techniques. *Computing Algorithms with Applications in Engineering*, 389–401.
- Martínez, L. F. (2020). Aplicación de métodos de deep learning a un sistema de evaluación de estudiantes.
- Mohammad, S., Gouse Baig y Shitharth. (2021). Wireless sensor network and iot based systems for healthcare application. *Materials Today: Proceedings*.
- Nord, A. y. P. J., Jeretta Horn y Koohang. (2019). The internet of things: Review and theoretical framework. *Expert Systems with Applications*, 133, 97–108.
- Ortiz, R. H. (2020). *Recomendación a grupos de usuarios usando el concepto de singularidades*. Doctoral dissertation, UNIVERSIDAD POLITÉCNICA DE MADRID.
- Palmieri, L., Viviana y Rivas. (2007). Gestión de información para la innovación tecnológica agropecuaria. *ComunIICA*, 3, 17–25.
- Paluszek, S., Michael y Thomas. (2019). An overview of machine learning. *MATLAB Machine Learning Recipes*, 1–18.
- Pérez, A., y Milla, M., M y Mesa. (2006). Impacto de las tecnologías de la

- información y la comunicación en la agricultura. *Cultivos tropicales*, 27 (1), 11–17.
- Ramasubramanian, A., Karthik y Singh. (2017). Machine learning theory and practices. En *Machine learning using r* (pp. 219–424). Springer.
- Ratnaparkhi, S., y Khan, C. y. K. S. y. S. P. y. D. M. y. S. A., Suvaid y Arya. (2020). Smart agriculture sensors in iot: A review. *Materials Today: Proceedings*.
- Rehman, T. U., y Mahmud, Y. K. y. J. J. y. S. J., Md Sultan y Chang. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and electronics in agriculture*, 156, 585–605.
- Trifan, A., Antunes, R., y Oliveira, J. L. (2021). Machine learning for depression screening in online communities. En G. Panuccio, M. Rocha, F. Fdez-Riverola, M. S. Mohamad, y R. Casado-Vara (Eds.), *Practical applications of computational biology & bioinformatics, 14th international conference (pacbb 2020)* (pp. 102–111). Cham: Springer International Publishing.
- Uddin, S., Khan, A., Hossain, M. E., y Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1–16.
- Vieira, W. H. L. y. M. A., Sandra y Pinaya. (2020). Main concepts in machine learning. En *Machine learning* (pp. 21–44). Elsevier.
- Vishwanath, Y., Upendra, R. S., y Ahmed, M. R. (2021). A review on advent of iot, cloud, and machine learning in agriculture. En J. S. Raj (Ed.), *International conference on mobile computing and sustainable informatics* (pp. 595–603). Cham: Springer International Publishing.
- Wortmann, K., F y Flüchter. (2015). *Internet of things: Technology and value added. business and information systems engineering*, 57 (3), 221–224.
- Yury, E., Shaev y Samoylova. (2017). The internet of things as socio-technological institution of civil society in post-informational era. En *2017 2nd international conference on computer and communication systems (icccs)* (pp. 142–145).
- Zhu, R.-G., Zhen y Huang, y cols. (2017). Study on the iot architecture and access technology. En *2017 16th international symposium on distributed computing and applications to business, engineering and science (dcabes)* (pp. 113–116).