

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ingeniería en Electricidad y Computación**

**DISEÑO Y DESARROLLO DE UNA HERRAMIENTA DE  
VISUALIZACIÓN PARA EL MANEJO Y MONITOREO  
AGRÍCOLA EN HACIENDAS DE PRODUCCIÓN DE CACAO**

**PROYECTO DE TITULACIÓN**

Previo la obtención del Título de:

**Magister en Ciencias de Datos**

Presentado por:

Christian Roberto Vergara Marcillo

GUAYAQUIL - ECUADOR

Año: 2022

## **DEDICATORIA**

Dedico este proyecto especialmente a mi familia, por ser el pilar fundamental de mi vida.

## **AGRADECIMIENTOS**

Agradezco al equipo de Investigación y Operaciones, quienes hicieron posible el desarrollo de este proyecto. Gracias por contribuir con todos los aspectos agronómicos presentes en cada uno de los módulos.

Quiero agradecer también a todos quienes de alguna u otra forma han sido parte de mi camino, no solo en el ámbito profesional, sino en el ámbito personal.

## DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, me corresponden conforme al reglamento de propiedad intelectual de la institución, y doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

Christian R. Vergara M.  
Christian Vergara M.

# COMITÉ EVALUADOR

---

**Miguel Realpe**

PROFESOR TUTOR

---

**José Córdova**

PROFESOR EVALUADOR

## RESUMEN

Ecuador es uno de los principales países latinoamericanos con mayores niveles de producción y exportación de cacao a nivel mundial. Las condiciones climáticas, el capital natural y variedades altamente productivas con las que cuenta el país, permiten el desarrollo de este cultivo. La agricultura, al igual que otros sectores productivos, se encuentra en un proceso de transformación digital, debido al surgimiento de nuevas tecnologías y a la cantidad de información que se genera a diario, lo cual hace posible el desarrollo de metodologías y soluciones basadas en datos que permiten la automatización de procesos. El presente proyecto tiene como objetivo desarrollar una herramienta de fácil acceso para la administración de datos en empresas dedicadas a la producción de cacao en Ecuador, mostrando la información a través de visualizaciones interactivas, que permitirán a los usuarios conocer el estado y avance de actividades agrícolas, productividad y monitoreo de frutos. De la misma forma se ha implementado un módulo de visualización para indicadores climáticos, los cuales provienen directamente de sensores instalados en la zona de producción, y de los que se obtiene la información para realizar predicciones de series temporales. Un análisis de los beneficios que representa la implementación de este proyecto en una hacienda de producción de cacao es realizado, el cual demuestra que el uso de este tipo de herramientas permite optimizar y automatizar procesos que contribuyen en gran medida a la toma de decisiones por parte del área operativa, con el objetivo de administrar apropiadamente los recursos.

**Palabras Clave:** Cacao, Periodo, Jornales, Visualización de datos, Aplicación web, Series temporales, Pronósticos

## **ABSTRACT**

*Ecuador is one of the main Latin American countries with the highest levels of cocoa production and exports worldwide. The climate conditions, the natural capital and high productivity genotypes in the country enables the development of this crop. Agriculture, like other productive sectors, is undergoing a process of digital transformation, because of the emergence of new technologies and the amount of data generated daily which make possible the development of methodologies and data-driven solutions that enable processes automation. The objective of this project is to develop an easy access tool for data management in cocoa production companies in Ecuador, displaying the information through interactive visualizations that will allow to the users to know the status and progress of agricultural activities, productivity, and fruit monitoring. In the same way, a visualization module has been implemented for the climate indicators data, which come directly from sensors installed in the production area, and from which the information is obtained to make time series predictions. An analysis of the benefits of implementing this project in a cocoa production farm is done, which shows that the use of this type of tools allows to optimize and automate processes that contribute greatly to decision making by the operational area in order to manage the resources properly.*

**Keywords:** *Cocoa, Period, Wages, Data visualization, Web application, Time series, Forecasts*

## ÍNDICE GENERAL

COMITÉ EVALUADOR.....	5
RESUMEN.....	I
<i>ABSTRACT</i> .....	II
ÍNDICE GENERAL.....	III
ABREVIATURAS .....	V
SIMBOLOGÍA .....	VI
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS .....	VIII
CAPÍTULO 1 .....	9
1.    Introducción .....	9
1.1    Descripción del problema .....	9
1.2    Justificación del problema.....	12
1.3    Solución Propuesta.....	13
1.4    Objetivos.....	14
1.4.1    Objetivo General .....	14
1.4.2    Objetivos Específicos .....	15
1.5    Metodología .....	16
1.6    Resultados esperados .....	17
1.7    Dataset .....	18
CAPÍTULO 2 .....	21
2.    Introducción .....	21
2.1    Fundamentos del problema .....	21
2.2    Soluciones basadas en datos relacionadas al problema .....	22



2.2.1	Sistemas de Información de Gestión Agrícola (FMIS) y Agricultura de precisión	22
2.2.2	Sistemas de Información Geográfica en Agricultura	24
2.2.3	Sistemas de Visualización de datos	26
2.2.4	Sistemas de Visualización de datos en agricultura	27
2.3	Librerías y software a utilizar	30
2.3.1	Visualización de datos en un entorno web	30
2.3.2	Pronóstico de series temporales con ARIMA	31
CAPÍTULO 3		34
3.	DISEÑO E IMPLEMENTACIÓN DE LA HERRAMIENTA	34
3.1	Análisis, exploración y validación de datos y fuentes.	34
3.2	Arquitectura y diseño de la aplicación web	38
3.3	Prototipos de algoritmos, modelos y módulos del sistema	47
3.4	Infraestructura para procesamiento y almacenamiento	51
3.5	Plataformas y prototipos de visualización	52
3.6	Métricas y comunicación de resultados	53
CAPÍTULO 4		55
4.	ANÁLISIS DE LOS RESULTADOS	55
4.1	Recolección de datos y estrategias para validación del proyecto	55
4.2	Puesta en marcha y funcionamiento	63
4.3	Pruebas de funcionalidad	64
4.4	Análisis costo/beneficio	67
CONCLUSIONES Y RECOMENDACIONES		72
BIBLIOGRAFÍA		<b>Error! Bookmark not defined.</b>

## **ABREVIATURAS**

ESPOL	Escuela Superior Politécnica del Litoral
FMIS	Farm Management Information System
FMS	Farm Management System
GIS	Geographic Information System
ARIMA	Autoregressive Integrated Moving Average

## SIMBOLOGÍA

TM	Toneladas Métricas
° C	Grado Centígrado
mm	Milímetros
MJ	Mega Joules
m <sup>2</sup>	Metro cuadrado

## ÍNDICE DE FIGURAS

Figura 3.1 Modelo de flujo de datos en Django .....	<b>Error! Bookmark not defined.</b>
Figura 3.2 Diseño general de la herramienta de visualización	<b>Error! Bookmark not defined.</b>
Figura 3.3 Diseño del módulo de visualización de actividades operativas.....	42
Figura 3.4 Diseño del módulo de visualización de producción	<b>Error! Bookmark not defined.</b>
Figura 3.5 Diseño del módulo de visualización de conteo de mazorcas.....	<b>Error! Bookmark not defined.</b>
Figura 3.6 Diseño de la visualización para indicadores climáticos	<b>Error! Bookmark not defined.</b>
Figura 3.7 Diagrama de módulos de la herramienta de visualización.....	48
Figura 3.8 Pronósticos de modelos STL y ARIMA en R	<b>Error! Bookmark not defined.</b>
Figura 3.9 Visualización de temperaturas promedio (verde) y pronóstico a 30 días .....	<b>Error! Bookmark not defined.</b>
Figura 4.1 Proceso de ingesta de datos, procesamiento y visualización .....	<b>Error! Bookmark not defined.</b>
Figura 4.2 Flujo de datos para generar visualizaciones de actividades operativas .....	<b>Error! Bookmark not defined.</b>
Figura 4.3 Visualización de actividades operativas.....	<b>Error! Bookmark not defined.</b>
Figura 4.4 Flujo de datos para generar visualizaciones de producción .....	<b>Error! Bookmark not defined.</b>
Figura 4.5 Visualización de producción por periodos...	<b>Error! Bookmark not defined.</b>
Figura 4.6 Visualización para el monitoreo de mazorcas periódico	<b>Error! Bookmark not defined.</b>
Figura 4.7 Visualización de temperaturas promedio ....	<b>Error! Bookmark not defined.</b>
Figura 4.8 Tabla de análisis de indicadores climáticos	<b>Error! Bookmark not defined.</b>
Figura 4.9 Visualización de temperaturas promedio y pronósticos	<b>Error! Bookmark not defined.</b>

Figura 4.10 Resultado del pronóstico en una tabla en la interfaz web..... **Error!**

**Bookmark not defined.**

Figura 4.11 Gráfico de actividades con botón para agregar datos**Error!   Bookmark**

**not defined.**

Figura 4.12 Formulario de registro de actividades operativas**Error!   Bookmark   not**

**defined.**

## ÍNDICE DE TABLAS

Tabla 1.1 Dataset de actividades operativas .....	19
Tabla 1.2 Dataset de producción .....	19
Tabla 1.3 Dataset de conteo de mazorcas .....	20
Tabla 1.4 Dataset de condiciones climáticas .....	20
Tabla 3.1 Funciones en la vista Index.....	46
Tabla 3.2 Funciones en la vista Actividades operativas.....	46
Tabla 3.3 Funciones en la vista Condiciones Climáticas .....	46
Tabla 3.4 Funciones en la vista Producción .....	46
Tabla 3.5 Funciones en la vista Conteo de Mazorcas .....	47
Tabla 3.6 Resultado de diferentes modelos de pronóstico para series temporales...	49
Tabla 4.1 Resumen de actividades de un periodo obtenido con RStudio.....	58
Tabla 4.2 Funciones utilizadas en el procesamiento de los datos .....	65
Tabla 4.3 Análisis de tiempo requerido para cada funcionalidad del sistema.....	69
Tabla 4.4 Personal por áreas que utilizarían la herramienta.....	70
Tabla 4.5 Valores de 2 proveedores de servicios de hosting web .....	71

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1 Descripción del problema

Ecuador es uno de los países latinoamericanos con mayor producción y exportación de cacao a nivel mundial, contribuyendo el sector cacaotero al 5% de la población económicamente activa (PEA) y el 15% de la PEA rural del país, de acuerdo con el Instituto Nacional de Estadísticas y Censos (INEC). El cacao es una de las bases económicas de familias del sector costero, andino y amazónico del país (ANECACAO, 2021). Las provincias con mayor producción en toneladas métricas (Tm) en Ecuador son: Guayas (81.94 Tm), Los Ríos (39.26 Tm), Cañar (23.80 Tm) y Manabí (18.26 Tm) (Macías Barberán et al., 2019).

La producción de cacao en Ecuador depende de diversos factores, entre ellos, las condiciones climáticas, las cuales permiten el crecimiento del árbol que produce mazorcas en cuyo interior se encuentran las almendras que, luego de un proceso de postcosecha y secado, permiten la elaboración del licor de cacao y productos de consumo como el chocolate. Otro de los factores de su amplio desarrollo en el país es la disponibilidad de variedades de alta productividad que han sido desarrolladas genéticamente durante varias décadas. Las dos variedades principales que se cultivan en Ecuador son: CCN-51 (Colección Castro Naranjal) y fino de aroma o Nacional (León Villamar et al., 2016). Por este motivo, el cacao ecuatoriano se encuentra entre uno de los cultivos del sector agrícola de mayor relevancia, no solo a nivel local debido al consumo interno, sino a nivel internacional, ya que forma parte de la materia prima utilizada por industrias dedicadas a la producción masiva de productos alimenticios derivados del cacao.

En este sentido, las haciendas o granjas de producción de este cultivo a gran escala, además de formar parte del sector productivo del país, manejan una estructura comercial que requiere la administración adecuada de los recursos necesarios para su correcta operatividad, conforme a los diferentes factores y retos asociados con el proceso de cultivar cacao. Debido a esto, es importante implementar soluciones tecnológicas basadas en datos, que permitan a los

agricultores el manejo eficiente de la información que se registra a diario, con la finalidad de elaborar planes operativos conforme el cultivo se desarrolla y de esta forma contribuir no solo a la tecnificación de la agricultura, sino al desarrollo económico de este sector.

La información referente a las diferentes labores agrícolas, personal operativo, recursos y cosecha se registran a lo largo del año con la finalidad de monitorear el cultivo y asegurar que las principales actividades para su apropiado crecimiento se realicen de manera oportuna. De la misma forma es importante visualizar y comprender la información obtenida de fuentes externas como sensores meteorológicos, que proporcionan indicadores climáticos en la zona de producción y como este se relaciona con el desarrollo de los cultivos.

Actualmente, en muchas haciendas de producción a gran escala de cacao, esta información se registra mediante el uso de computadoras utilizando hojas de cálculo. En el mejor de los casos, estos archivos son compartidos en la nube para que las diferentes personas interesadas en la empresa y que participan en el registro de los datos puedan acceder a una misma fuente de información; sin embargo, esto implica varias consideraciones que deben ser tomadas en cuenta:

1. La información, si bien se almacena en archivos en la nube, al momento de requerir del procesamiento para realizar un análisis y generar reportes, será necesario descargar una copia del archivo, creando respaldos locales en el computador del usuario. Esto podría producir una inconsistencia si no se dispone de un control de versiones de los archivos utilizados.
2. El proceso para obtener información con resultados requiere de la participación de varias personas de forma diaria, semanal y mensual: Supervisores de campo que registran las actividades de un determinado sector productivo, Asistente operativo que toma los datos y los almacena en una hoja de cálculo, la cual posteriormente es compartida a los Técnicos agrícolas, quienes son los que finalmente procesarán la información para obtener métricas y resultados. Todo este proceso involucra del análisis manual



y minucioso de los datos para generar gráficos, tablas e información relevante para tomar decisiones y planificar de recursos.

3. Los gráficos y visualizaciones generadas se realizan utilizando únicamente MS Excel como software principal para analizar y procesar los datos, que, si bien es una herramienta ampliamente conocida, tiene limitaciones en los tipos de gráficos y en la forma de filtrar la información, ya que requiere de un conocimiento en el uso de fórmulas para poder obtener resultados finales.
4. Los datos de producción en su mayoría pueden ser georreferenciados o pertenecer a una zona geográfica determinada dentro de la hacienda, por lo tanto, estos podrían ser presentados en mediante mapas interactivos para su mejor comprensión; sin embargo, al momento no es posible, ya que se utiliza únicamente la suite de Office para visualizar los resultados.
5. Los análisis son realizados de forma independiente, es decir, se cuenta con varios datasets: uno para la producción, uno de actividades agrícolas y uno de condiciones climáticas. En la mayor parte del tiempo no es posible combinar esta información con la finalidad de realizar una interpretación más profunda de los datos.
6. En el manejo de los registros existe información relacionada a las actividades operativas que al momento y debido a su frecuencia de registro (diario), es poco analizada y no se dispone de una herramienta que permita, de forma automática, generar visualizaciones.

Debido a la problemática analizada, surge la necesidad de disponer de una herramienta única que centralice la información, permitiendo a los encargados del área operativa, registrar los datos necesarios; para que de esta forma la herramienta los procese automáticamente y genere visualizaciones periódicas.

Los datos considerados en el desarrollo del proyecto descritos en este documento serán:

- Información de actividades operativas
- Producción periódica
- Conteo periódico de frutos

- Indicadores climáticos (Temperatura, Humedad relativa, Radiación solar, Lluvia, Evapotranspiración referencial y Grados días)

## **1.2 Justificación del problema**

El manejo apropiado del cultivo de cacao durante todas sus etapas de desarrollo inicia con una planificación previa, considerando todos los factores que influirán en la producción esperada. Estos factores están relacionados a la motivación económica, el retorno de la inversión, el lugar en donde se sembrará el cultivo y sus condiciones climáticas; y a lo largo de su crecimiento, todas las labores agrícolas que serán necesarias para asegurar su desarrollo y posterior producción. Por este motivo, proporcionar a las empresas y haciendas de producción a gran escala de este cultivo, una solución basada en datos posibilitaría un manejo adecuado de la información que es recopilada, almacenada y visualizada; ya que al momento muy pocas empresas del sector agrícola en Ecuador trabajan con este tipo de herramientas tecnológicas, las cuales facilitan la tarea de procesar y automatizar la generación de resultados y contribuyen a una apropiada planificación operativa a lo largo del año.

Esta herramienta permitiría centralizar el almacenamiento de los datos que son utilizados para la gestión de recursos, ya que actualmente esta información se encuentra en la nube o en archivos locales de los usuarios encargados. De igual forma y al trabajar mayormente con datos temporales, la herramienta será capaz de generar gráficos dinámicos en intervalos de tiempo que pueden ser escogidos por el usuario (rango de tiempo específico) o en su defecto se proporcionará un listado de periodos predefinidos dependiendo de la información requerida, ya que las visualizaciones utilizadas en la actualidad por la empresa son estáticas debido a que son generadas únicamente con MS Excel.

La solución propuesta permitirá la generación de visualizaciones efectivas ya sea en gráficos de líneas, gráficos de barras o mapas interactivos, a partir de los datos ingresados al sistema, para permitirle a la empresa optimizar el tiempo que actualmente es utilizado en el procesamiento de toda la información registrada en hojas de cálculo.

Debido a que el riego en las plantaciones de cacao es una de las actividades agrícolas de mayor importancia, ya que requiere de la instalación de sistemas de suministro de agua que provean la cantidad necesaria para que el cultivo crezca apropiadamente y en condiciones óptimas, es importante que dentro de la herramienta exista un componente que indique cuánto regar. Actualmente este indicador se obtiene, en algunos casos, utilizando la Evapotranspiración Referencial (ET<sub>o</sub>). La ET<sub>o</sub> es un indicador que combina información de temperatura, humedad relativa, radiación solar, velocidad del viento, presión atmosférica y ubicación geográfica del lugar en donde se realiza la medición, y es una medida en milímetros de cuánta agua se requiere en una determinada área (Hectáreas).

Por este motivo, las condiciones climáticas son un componente importante que funcionará como un módulo en la herramienta descrita en el presente documento, ya que permitirá a lo largo del año monitorear el desempeño en la producción y de las diferentes actividades agrícolas en conjunto con los indicadores climáticos de la zona, la cual es obtenida mediante sensores instalados en una estación meteorológica automatizada Campbell Scientific localizada en la hacienda. Esta información permitirá gestionar de una forma más apropiada los recursos y el personal disponible, ya en determinadas épocas del año se evita realizar determinadas labores porque las condiciones lo impiden (como en la época de lluvia), y se priorizan otras dentro de la planificación operativa. De la misma forma y debido a que los indicadores climáticos como la temperatura, radiación solar, y evapotranspiración referencial tienen las características de series de tiempo, será posible obtener estimaciones (predicciones) para periodos de tiempo próximos.

### **1.3 Solución Propuesta**

La solución descrita en el presente documento plantea el desarrollo de una herramienta de fácil acceso que permita a los encargados de la producción de cacao monitorear, mediante visualizaciones efectivas en un entorno web, las diferentes actividades agrícolas y operativas asociadas al proceso y etapas de desarrollo del cultivo. Esta información junto con la producción real obtenida se

actualizará periódicamente (cada 4 semanas) en el sistema. De la misma forma se incluirá el conteo de frutos necesarios para estimar la cosecha que se obtendrá en periodos de tiempo próximo en la hacienda. El clima como factor asociado al desarrollo del cultivo también será incluido en la plataforma, facilitando la generación automática de gráficos de los diferentes indicadores climáticos.

Este proyecto ha sido diseñado considerando tres de los aspectos más importantes para la gestión operativa de una hacienda dedicada al cultivo y producción de cacao, los cuales son: Planificación y ejecución de actividades agrícolas esenciales para el desarrollo del cultivo, Producción en toneladas métricas cuantificada de forma periódica e Indicadores climáticos obtenidos en la hacienda.

Estos factores proveen información que permite elaborar planes para la gestión apropiada de recursos y métricas en base a la productividad del cultivo. Existen otros factores asociados al manejo y producción como, por ejemplo: el riego, la fertilización, y la incidencia de plagas y enfermedades, entre otros; que, a pesar de no haber sido considerados en el desarrollo de esta herramienta de visualización, pueden ser agregados como módulos dentro del sistema para obtener de esta forma una plataforma integral de manejo agrícola.

## **1.4 Objetivos**

### **1.4.1 Objetivo General**

El objetivo principal de este proyecto es integrar la información del manejo operativo de una hacienda de producción de cacao a gran escala en una plataforma de fácil acceso, con la finalidad de generar visualizaciones periódicas de las principales actividades agrícolas asociadas a la producción y manejo del cultivo, así como la descripción de la producción en periodos de tiempos próximos y su relación con la estimación del clima basado en datos de sensores.

### **1.4.2 Objetivos Específicos**

Implementar una herramienta que permita a los usuarios a cargo de las áreas productivas de la empresa registrar los diferentes datos relacionados a las actividades agrícolas, así como la producción obtenida de forma periódica. De la misma forma, esta herramienta se conectará a la base de datos de la estación meteorológica de la empresa para obtener la información climática. La aplicación se encargará del preprocesamiento de la información y los almacenará en una base de datos que posteriormente será utilizada para las respectivas visualizaciones y predicciones.

Desarrollar un módulo que permita la generación de visualizaciones interactivas a partir de los datos ingresados de las diferentes actividades agrícolas por periodos de tiempo (que pueden ser escogidos por el usuario), producción obtenida a la fecha e información de factores climáticos por medio de grupos en graficas multivariantes. Además de las visualizaciones, la herramienta proveerá de información descriptiva de los datos tales como su distribución, promedio, varianza y desviación estándar.

Relacionar la estimación manual de producción de cacao en la hacienda para periodos de tiempo próximos con predicciones de series de tiempo en base a los datos históricos y conforme se incorporen nuevos valores del muestreo de frutos periódico. De igual forma, los indicadores climáticos de temperatura promedio, radiación solar y evapotranspiración serán modelados como series de tiempo para que los usuarios a cargo del desarrollo de las actividades operativas dispongan de información climática para la planificación anual de las respectivas labores agrícolas.

## **1.5 Metodología**

Para el presente proyecto se diseñará y desarrollará una aplicación web como herramienta o plataforma principal para el monitoreo de actividades agrícolas, producción e indicadores climáticos, mediante la generación automática de gráficos interactivos de la información que es gestionada por los encargados del área operativa. La herramienta tomará inicialmente la información histórica de la empresa que será almacenada en una base de datos diseñada para esta solución; sin embargo, se dispondrá de formularios para el ingreso de datos en cada uno de los datasets considerados para este proyecto y cuya información depende del registro de información por parte del equipo operativo. El sistema contará con las siguientes características:

### **a) Visualización de la información de producción y actividades operativas**

La información relacionada a las actividades operativas permite realizar planificaciones del personal y los recursos del área operativa de la empresa, con el propósito de asegurar el manejo apropiado del cultivo durante el año. Esta información será visualizada tanto en mapas que muestran la información geográfica de los sectores productivos y gráficos de barras con el detalle de las actividades operativas, así como el personal requerido para las mismas. Esta información se registra de forma diaria, pero es analizada de forma periódica (cada 4 semanas). Para este propósito se ha realizado un levantamiento de requerimientos con los usuarios encargados, con la finalidad de desarrollar dos módulos: uno que permita el ingreso de la información necesaria y uno que gestione las visualizaciones en el sistema.

### **b) Visualización el estado del cultivo en términos de fructificación y monitoreo del clima**

La información del conteo de frutas es registrada cada 4 semanas y es almacenada en una base de datos conectada a una aplicación móvil, ya que para esta actividad se utilizan tablets. Esta información será mostrada en un formato de gráfico de barras que permita visibilizar la cantidad de las mazorcas de cacao en sus diferentes etapas, y para el cual se ha diseñado un

módulo de visualización dentro del sistema. Con el conteo de mazorcas se estima la cantidad de mazorcas que serán cosechadas en el siguiente periodo (luego de 4 semanas).

La información del clima es almacenada con una frecuencia de 60 minutos en la base de datos de la estación meteorológica principal Campbell Scientific y para el presente proyecto se establecerá una conexión entre la base de datos del sistema y la estación meteorológica. Dependiendo del indicador climático, estos podrán ser modelados como series de tiempo en gráficos de líneas, o como gráficos de barras, dentro de un módulo de visualización diseñado para este propósito. La importancia del clima en el cultivo de cacao radica en que algunos indicadores están relacionados directamente con actividades operativas, como la evapotranspiración referencial que permite establecer planes de riego para el cultivo.

### **c) Estimación de la producción del cultivo e indicadores climáticos**

Tanto la información de la producción del cultivo registrada por el personal operativo, como la información del clima (temperatura, radiación solar y evapotranspiración) registrada por la estación meteorológica serán modeladas en series de tiempo en base a los datos históricos, con la finalidad de realizar estimaciones o predicciones para periodos de tiempo próximos (días o meses). La información de predicción permitirá al equipo operativo disponer de un recurso adicional para la planificación de actividades y evaluación de resultados.

Debido a que la herramienta ha sido desarrollada utilizando el framework de Django, los algoritmos de predicción de series de tiempo se realizarán con librerías para este objetivo en el lenguaje de programación Python.

## **1.6 Resultados esperados**

Al finalizar el desarrollo del proyecto se espera proporcionar una herramienta para la gestión y monitoreo de la producción del cultivo de cacao que permita la visualización de los factores necesarios para su manejo y que permita a los

encargados de áreas operativas una solución para la toma de decisiones basadas en los datos. De la misma forma, se espera proveer a la empresa de una herramienta que permita la gestión de la información necesaria para evaluar el estado del cultivo y su producción optimizando el tiempo que actualmente es requerido para realizar el análisis de la información y las visualizaciones de forma manual.

## 1.7 Dataset

En la descripción de los diferentes datasets que se utilizarán en el proyecto se manejan los siguientes términos asociados al contexto del sector agrícola y que son utilizados por la empresa:

**Periodo:** Un periodo comprende de 4 semanas y el primero empieza en la primera semana de cada año, por lo tanto, anualmente se cuenta con 13 periodos.

**Sector:** La hacienda se encuentra dividida en 3 sectores productivos principales y que son administrados por el Equipo Operativa de la Hacienda. Cada sector tiene características geográficas y tipos de suelo específicos. Un sector esta conformado de varios lotes. Cada sector tiene su planificación propia de acuerdo a los recursos disponibles y a las métricas establecidas.

**Lote:** Un lote es un área de terreno delimitada en donde se cultiva cacao, con características específicas tales como: diseño de plantación, tipo de suelo, sistemas de riego y variedad del cultivo.

### **Dataset 1: Actividades operativas**

Contienen la información de las actividades operativas realizadas por un equipo de personas, tiempo dedicado, tipo de actividad y lote en donde se realiza. Los tipos de actividad contienen información de la labor agrícola realizada, esta puede ser: cosecha, control de malezas, aplicación de fertilizantes, riego, entre otras. Esta información tiene un nivel de registro diario para cada actividad y es



administrada por los encargados de cada sector productivo, en total se disponen de 5547 registros. Esta información actualmente se encuentra en hojas de cálculo almacenadas en los computadores del equipo operativo y en la nube en OneDrive, los tipos de datos y rangos se describen en la tabla a continuación:

**Tabla 1.1**  
**Dataset de actividades operativas**

<b>Variable</b>	<b>Tipo</b>	<b>Descripción</b>	<b>Rango</b>
<b>Fecha</b>	Fecha	Fecha de la actividad	Desde 01/01/2021
<b>Periodo</b>	Entero	Periodo del año (4 semanas)	1 - 13
<b>Año</b>	Entero	Año de la actividad	2021
<b>Mes</b>	Entero	Mes del año	1-12
<b>Semana</b>	Entero	Semana del año	1 - 52
<b>Día</b>	Entero	Día de la semana	1 - 7
<b>Sector</b>	Entero	Sector productivo	1 - 3
<b>Lote</b>	Entero	Subárea del sector	1 - 46
<b>Personas</b>	Entero	Cantidad de personas requeridas	1 - 20
<b>Actividad</b>	Texto	Tipo de actividad	6 tipos de actividad
<b>Tiempo</b>	Entero	Tiempo asignado a la actividad (horas)	1 - 8

### **Dataset 2: Producción**

Contiene la información de la producción obtenida en toneladas métricas de los diferentes lotes durante cada periodo (cada 4 semanas). Esta información se encuentra disponible en OneDrive y en total se disponen de 345 registros con las siguientes características de los datos son:

**Tabla 1.2 Dataset de producción**

<b>Variable</b>	<b>Tipo</b>	<b>Descripción</b>	<b>Rango</b>
Año	Fecha	Año en curso	Desde 2021

Lote	Entero	Área productiva	1 - 46
Sector	Entero	Sector productivo	1 - 3
Periodo	Entero	Periodo del año (4 semanas)	1 - 13
Producción	Float	Producción en toneladas métricas	-

### **Dataset 3: Conteo de mazorcas**

Contiene la información del estado del cultivo en cantidad representativa de frutos, la misma que se registra cada periodo y que permite estimar la cosecha en periodos de tiempo próximos. Esta información se encuentra alojada en una base de datos en OneDrive, implementada para almacenar los datos que se registran desde una aplicación móvil en una tablet. En total se disponen de 43662 registros y las características de los datos son:

**Tabla 1.3 Dataset de conteo de mazorcas**

<b>Variable</b>	<b>Tipo</b>	<b>Descripción</b>	<b>Rango</b>
Fecha	Fecha	Fecha del muestreo	Desde 01/01/2017
Periodo	Entero	Periodo del año (4 semanas)	1 - 13
Sector	Entero	Sector productivo	1 - 3
Lote	Entero	Subárea del sector	1 - 36
Bloque	Entero	División del lote	1 - 3
Árbol	Entero	Número del árbol a muestrear	1 - 15
Estado_fruto	Texto	Estado de la mazorca de cacao	9 tipos

### **Dataset 4: Condiciones climáticas**

La información del clima tomado de la estación meteorológica principal de la empresa contiene datos registrados por sensores, tales como: temperatura, humedad, radiación, evapotranspiración, lluvia, velocidad y dirección del viento. Esta información se encuentra en la base de datos del sistema y su frecuencia de registro es automática cada 60 minutos. Para el presente proyecto se cuenta con 159934 registros con datos tomados desde mayo de 2018. Las características de los datos son:

**Tabla 1.4 Dataset de condiciones climáticas**

<b>Variable</b>	<b>Tipo</b>	<b>Descripción</b>	<b>Rango</b>
Fecha_Hora	Fecha	Fecha y hora de la toma de datos	Desde 05/01/2018

# CA PÍT ULO 2

BP_mbar_Avg	Flotante	Presión atmosférica	0.0 - 1020.0
AirTC_Avg	Flotante	Temperatura del aire (Promedio)	0.0 - 40.0
AirTC_Max	Flotante	Temperatura del aire (Máxima)	0.0 - 40.0
AirTC_Min	Flotante	Temperatura del aire (Mínima)	0.0 - 30.0
RH_Avg	Flotante	Humedad Relativa (Promedio)	0.0 - 99.9
RH_Max	Flotante	Humedad Relativa (Máxima)	0.0 - 99.9
RH_Min	Flotante	Humedad Relativa (Mínima)	0.0 - 99.9
WindDir	Flotante	Dirección del viento	0.0 - 360.0
rain_Tot	Flotante	Precipitación total	0.0 - 35.0
SlrMJ_Tot	Flotante	Radiación total	0.0 - 25.0
Dew_Point_C	Flotante	Punto de rocío	0.0 - 25.0
ET_mm	Flotante	Evapotranspiración referencial	0.0 - 6.0

## 2. INTRODUCCIÓN

### 2.1 Fundamentos del problema

En la actualidad, se ha convertido en una prioridad para las empresas el uso de herramientas y aplicaciones que permitan una gestión apropiada de datos abundantes y diversos, y que permitan obtener información que conduzca a la toma de decisiones y optimización de recursos. Por lo cual, es importante disponer de mecanismos y técnicas que faciliten el análisis de datos y que permitan visualizar la información de forma correcta y efectiva con herramientas de analítica desarrolladas para este propósito. Desde esta perspectiva, las haciendas o granjas de producción de cultivos, vinculadas al sector agrícola, presentan una oportunidad para el desarrollo de tecnologías de software que les permita optimizar y gestionar las operaciones y actividades relacionadas a la producción deseada. Estos sistemas, denominados Sistemas de Información de Gestión Agrícola (FMIS, por sus siglas en inglés), o Sistemas de Gestión Agrícola (FMS) permiten el monitoreo y analítica de actividades agrícolas, manejo, registro y almacenamiento de datos, y revisión de la producción y horarios de trabajo (Top 9 Farm Management Software in 2021, 2021).

Al igual que en muchas empresas, en el ámbito de entornos de producción agrícola es común presenciar el uso de herramientas o programas utilitarios para

la gestión de los datos, sin embargo, es importante resaltar las características generales de un Sistema de Gestión Agrícola que permita el manejo apropiado de la información generada durante el proceso y desarrollo del cultivo, ya que mediante este, será posible establecer una buena planificación operativa a lo largo del año y por lo tanto incrementar las ganancias, basado en la optimización del uso de recursos y promoviendo prácticas agrícolas más sostenibles (Pechter, 2021). Un sistema de gestión agrícola permite diferentes opciones para optimizar los resultados, en términos generales es posible contar con:

- Mapeo de la hacienda o granja
- Gestión de personal y medición de actividades agrícolas
- Generación de reportes
- Análisis agrícola y financiero
- Condiciones climáticas
- Gestión de inventario

## **2.2 Soluciones basadas en datos relacionadas al problema**

### **2.2.1 Sistemas de Información de Gestión Agrícola (FMIS) y Agricultura de precisión**

Los Sistemas de Gestión de Información Agrícola son sistemas que integran los componentes necesarios para la implementación de una agricultura de precisión en granjas (Stafford, 2000). El término “Agricultura de precisión” fue desarrollado por Johnson et al. (1983) y describe una “labranza (del suelo) prescrita y personalizada” y ya desde esta época se contemplaba, de forma visionaria, que la automatización y los sistemas de detección y localización transformarían la producción agrícola, conforme la tecnología sea puesta en marcha. De acuerdo con Stafford (2000) la agricultura de precisión es computacionalmente más completa y es descrita como una práctica de “Información intensa” ya que involucra el procesamiento de grandes cantidades de datos generados por un mapeo de suelos, factores ambientales y manejo del cultivo, siendo posible la incorporación de datos externos como información del clima o de mercados. Debido a esto, surge la necesidad del desarrollo de

herramientas de integración de datos que soporten la sobrecarga de la información, que finalmente será presentada a los administradores o gerentes de haciendas de producción de cultivos para la toma de decisiones (Sigrimis et al., 1999).

De acuerdo con Nikkilä et al. (2010) la implementación de un sistema completo para la gestión de información agrícola es complejo e implica un desarrollo de software integral, por lo cual, muchas investigaciones en este campo sugieren que el enfoque de implementación debe ser realizado en un componente individual de un FMIS. Por este motivo, muchos de los sistemas FMIS están asociados a soluciones que contribuyen al soporte en la toma de decisiones o a modelos biológicos computacionales orientados en la estimación de cosecha de un cultivo mediante modelos de predicción. Sin embargo, promover la implementación y posterior adopción de este tipo de herramientas, o sistemas de visualización de datos, para el manejo de la información en una hacienda de producción de cacao podría representar un reto debido diversos factores que han sido estudiados en otros países. La educación y experiencia de los agricultores, la disposición a utilizar computadoras, seguridad en la gestión del cultivo e ingresos debido a la producción esperada debido al uso de tecnología son indicadores determinantes en el impacto en la adopción de este tipo de soluciones de software (Carrer et al., 2017).

En la actualidad existen una variedad de soluciones que se enfocan en un componente particular de un FMIS y que tienen como objetivo asistir a propietarios o agricultores en la gestión de recursos y elaboración de planes operativos que aseguren el desarrollo apropiado de sus cultivos. Estos productos son herramientas de software diseñadas para ser utilizadas en dos ambientes de trabajo distintos; soluciones como AgriSoft, Hecterra y AgroWin, son herramientas de software que se instalan en un sistema operativo como Windows (aplicaciones de

escritorio), y soluciones como Agrosoft o FarmLogic han sido diseñadas para trabajar en un entorno web, en donde la información es almacenada en servidores remotos. Ambos entornos de trabajo disponen de características que deben ser consideradas antes de ser implementadas. Los sistemas que son instalados en un computador, dependiendo de su implementación, podrían funcionar sin conectividad a internet, siendo este un factor importante antes de decidir por una solución, ya que muchas áreas de producción se encuentran en localidades en donde el acceso a internet es limitado. Este tipo de herramientas disponen también de la posibilidad de interactuar directamente con dispositivos externos conectados al computador como registradores de datos y equipos de control, que a su vez toman información de sensores. Sin embargo, este tipo de herramientas centralizan la información en un solo computador, a diferencia de las soluciones basadas en entornos web, que permiten la adquisición de los datos localmente y su almacenamiento tanto local como remoto, a su vez que hacen posible el acceso a la información desde cualquier ubicación en el mundo, ya que la centralizan en servidores remotos o en la nube. Para este tipo de herramientas es importante considerar que será necesaria una conexión a internet.

### **2.2.2 Sistemas de Información Geográfica en Agricultura**

Los Sistemas de Información Geográfica o GIS, por sus siglas en inglés, son sistemas que cuentan con las características de un sistema de información y son utilizados para asistir en la toma de decisiones, pero con la diferencia de que los datos son georreferenciados. Los datos almacenados y gestionados en un GIS además de contar con información espacial, representada por la ubicación mediante un sistema de coordenadas con latitud y longitud (Xiao et al., 2007), permiten la asociación con características temporales e información no espacial; debido a esto, este tipo de sistemas permiten un análisis tanto espacial como temporal de la información (Liu et al., 2017). De esta forma, los GIS son utilizados por agricultores para gestionar y analizar información

georreferenciada, que ha sido registrada en una o múltiples granjas para verificar de forma localizada como progresan las condiciones del área en donde crece el cultivo, el suelo, su biodiversidad, salinidad, etc. (Pettit et al., 2007).

El uso de estos sistemas en la agricultura se debe en gran medida a la posibilidad de inspeccionar y monitorear aspectos del manejo de granjas en lugares específicos, ya que es posible visualizar la información como capas sobre mapas, en donde se obtienen detalles focalizados en un solo gráfico. Esta ventaja de los GIS posibilita la realización de planes operativos en áreas determinadas, ya sea para suministrar fertilizantes, aplicar una cantidad específica de riego o asignar mayor personal para la actividad de cosecha en sectores de mayor productividad.

Los Sistemas de Información Geográfica también permiten conocer aquellas áreas con mayor incidencia de plagas y enfermedades y en las cuales es necesario aplicar insecticidas o fungicidas para proteger a todo el cultivo o solo a segmentos particulares que están siendo afectados. En este sentido, los GIS son herramientas poderosas que contribuyen no solo al análisis de los datos georreferenciados, sino a la presentación de los resultados de forma efectiva y focalizada con el propósito de gestionar acciones inmediatas ya que permiten, por ejemplo, conocer aquellas áreas en donde existe variabilidad en la producción, aquellas en donde la cosecha es mayor a la esperada o aquellas en donde existe un déficit.

Existen soluciones de GIS que son utilizadas por los agricultores para gestionar la información del negocio por todas las ventajas previamente descritas, sin embargo, muchas de ellas requieren de licencias para su uso y de un entrenamiento preliminar; entre las más conocidas están QGIS y ArcGIS, siendo QGIS de código abierto. Actualmente en el mercado se dispone de varias soluciones de software que incorporan el componente GIS con la finalidad de analizar la información ingresada como visualizaciones espaciales, siendo este una característica

importante, ya que facilita la administración de recursos y de personal al simplificar la comprensión de las qué áreas o lugares en donde se asigna mayor o menor recurso dependiendo de la naturaleza de la actividad. A su vez, los GIS permiten registrar y analizar la información del suelo en donde se cultiva, debido a que sus propiedades tendrán un impacto directo en la producción futura.

### **2.2.3 Sistemas de Visualización de datos**

La ciencia de datos y la analítica de datos se han convertido en un factor clave para resolver problemas complejos en diferentes áreas y disciplinas, ya que combinan varios aspectos de las ciencias computacionales, estadística, matemáticas aplicadas y visualizaciones, permitiendo analizar grandes cantidades de información con la finalidad de extraer conocimiento. La visualización de datos es una área extensiva y esencial en la ciencia de datos, importante para tomar decisiones y analizar información mediante representaciones gráficas (Dzemyda et al., 2019).

A diferencia de la visualización de datos estadísticos, la visualización en ciencia de datos permite a los usuarios seleccionar formas de representación gráfica de acuerdo con la naturaleza de la información y el objetivo que se desea transmitir, ya que estos son generados mediante el uso de computadoras que trabajan con patrones y tendencias. En la actualidad existen herramientas disponibles para la visualización de datos masivos, siendo estos, programas que se instalan localmente en un computador, aplicaciones web o librerías en lenguajes de programación específicos, diseñados para la analítica y ciencia de datos. Estas herramientas y aplicaciones cuentan con un conjunto de opciones predefinidas que en general incluyen los tipos de gráficos más básicos (gráficos de dispersión, gráficos de líneas, gráficos de barras, gráficos de burbujas y gráficos de radar), y de forma menos común visualizaciones más avanzadas tales como diagrama de acordes, gráficos de horizonte, entre otros (Sadiku et al., 2016).



De acuerdo con el estudio de Schmidt (2020), realizado con la finalidad de analizar diferentes herramientas y aplicaciones para la visualización de datos masivos, se puede obtener la siguiente clasificación según la disponibilidad de diferentes tipos de gráficos:

**De código abierto:** Python: Plotly y Seaborn, R: GGPlot2, Vega-Lite, D3.js, Google Charts, Charts.js, Apexcharts, dygraphs, Bokeh, RAW Graphs, .Net LiveCharts, Qt Charts

**Comercial:** Microsoft Power BI, Tableau, SAS Visual Analytics, Highcharts.js, Quadrigram, Matlab.

#### **2.2.4 Sistemas de Visualización de datos en agricultura**

Las características de soluciones FMIS desarrolladas para agricultura deben estar alineadas a las diferentes actividades operativas de las que depende el crecimiento óptimo del cultivo, por este motivo, en la mayoría de los escenarios, una solución en particular FMIS requiere de una capacitación previa y una inversión económica anual ya que operan mediante una licencia y soporte técnico remoto. Por esta razón, y por la cantidad de opciones disponibles en el mercado, es posible que los administradores de haciendas de producción agrícola decidan continuar utilizando herramientas de software como MS Excel. Uno de los objetivos de este proyecto, es el desarrollo de una herramienta que permita generar visualizaciones automáticas y efectivas, que contribuyan al monitoreo de la plantación y la producción obtenida con la finalidad de asistir a la toma de decisiones en la empresa. En este sentido, es importante mencionar que existen en el mercado herramientas de visualización de datos y analítica como Tableau (Tableau, 2021) y Microsoft Power BI (Power BI, 2021), que podrían ser utilizados con un propósito similar.

#### **Tableau**

Tableau es un software que empezó como un proyecto de ciencias computacionales en Stanford, su enfoque es hacer accesible la información y los resultados obtenidos a partir de los datos ingresados mediante la generación de visualizaciones. Incluye características como estadística y lenguaje natural, en una plataforma que integra la analítica de los datos que son conectados al sistema (Tableau, 2021). La interfaz permite la conexión directa a una base de datos, al igual que el uso de diferentes conectores para trabajar con hojas de cálculo de MS Excel. La plataforma ofrece un número de visualizaciones posibles de acuerdo con la naturaleza de la información, en este punto Tableau realiza un preprocesamiento para ofrecer alternativas alineadas directamente al tipo de datos, por ejemplo, si detecta registros relacionados con datos geográficos o localidades, es probable que sugiera el uso de mapas. También dispone de opciones para customizar las visualizaciones y combinar tipos de gráficos de acuerdo con las necesidades del negocio y herramientas del tipo drag and drop con las variables presentes en el dataset.

### **Microsoft Power BI**

Power BI es la herramienta principal de Microsoft para el análisis, inteligencia de negocios y visualización, que permite generar gráficos a partir de los datos que son conectados a la plataforma, mediante una interfaz intuitiva, desde el ingreso de la información hasta la generación de reportes con gráficos personalizados (Power BI, 2021). El sistema dispone de una variedad de opciones predefinidas para las visualizaciones y una de las ventajas frente a otras soluciones, es su integración robusta con otras plataformas del entorno de Microsoft, tales como, Azure, OneDrive for Business, MS Excel, entre otras.

El uso apropiado de Power BI requiere de un entrenamiento o capacitación previa al personal que lo utilizará, con la finalidad de conocer todas las posibilidades que la herramienta ofrece y cómo manejarlas de

forma correcta. Esto podría, al igual que en el caso de Tableau, ser una limitante para que una empresa o negocio decida utilizar este software. El uso de Power BI requiere también de una licencia de pago luego del periodo de prueba otorgado por Microsoft.

### **RStudio**

RStudio es una herramienta que permite trabajar con el lenguaje de programación R y su orientación es el análisis estadístico de datos. RStudio dispone de un número amplio de librerías desarrolladas para trabajar en proyectos de ciencia de datos; de la misma forma RStudio permite generar visualizaciones con los datos que son cargados al entorno mediante el uso de librerías como GGPlot2. Esta librería trabaja en integración con otros paquetes y funcionalidades del entorno de R, lo cual la ha convertido en una de las principales opciones al momento de generar visualizaciones durante el flujo de trabajo en la analítica y ciencia de datos.

Una de las ventajas de R, es que es un lenguaje de código abierto y dispone de documentación estable y una amplia comunidad de desarrolladores. RStudio, a pesar de disponer de una opción de pago, también dispone de una versión gratuita, lo que la convierte en el complemento perfecto para utilizar R mediante una interfaz gráfica. El uso de R es amplio en el área de estadística y proyectos científicos, ya que su aprendizaje es bastante accesible, llegando a áreas como biociencias, genética y agricultura. La librería GGPlot2 de R permite la creación de gráficos altamente personalizables, además de que, al trabajar con un lenguaje de programación interpretado, permite optimizar procesos y automatizar tareas.

Uno de los requisitos para utilizar RStudio, es disponer de un conocimiento en programación en R, lo cual, para un usuario final, que desea ver resultados mediante visualizaciones, lo hace bastante limitado; sin embargo, su uso en la ciencia de datos sigue en aumento en conjunto

con lenguajes de programación como Python y librerías de visualización como Seaborn.

## **2.3 Librerías y software a utilizar**

### **2.3.1 Visualización de datos en un entorno web**

Debido a que la herramienta a desarrollar será una aplicación web, la implementación dispondrá de una interfaz gráfica con opciones propias de un Sistema de Gestión de Información Agrícola o FMIS, ya que permitirá registrar información y descargar reportes, siendo su componente principal la generación de visualizaciones. De la misma forma, y al trabajar con datos georreferenciados en la información de cosecha tendrá un componente GIS mediante el uso de la librería Leaflet.js, en donde la información de producción obtenida será presentada en mapas interactivos con el objetivo de facilitar la comprensión de los resultados.

Las visualizaciones serán el eje central de esta herramienta, si bien la información podrá ser manejada en tablas de resultados, gran parte de los datos será posible representarlos ya sea en gráficos de barras, para las actividades operativas, conteo de frutos y producción, y en grafico de líneas para las series temporales asociadas al clima.

Esta herramienta ha sido creada siguiendo la arquitectura cliente-servidor, con la finalidad de separar las tareas que serán necesarias en cada una de las funcionalidades y módulos diseñados para el sistema. Del lado del servidor se ha escogido trabajar con Django como framework principal que gestiona todo el flujo de información en el entorno y la generación de visualizaciones. El motivo del uso de Django radica principalmente en que trabaja fundamentalmente con el lenguaje de programación Python, el cual permite, a su vez, trabajar con librerías ya conocidas como Pandas y Numpy. Del lado del cliente se trabajará principalmente con Javascript y librerías para generación de gráficos como D3.js con los datos facilitados por Django.

De forma similar, y con la finalidad de separar procesos y tareas, esta solución sigue los lineamientos del patrón de diseño Modelo-Vista-Template de Django, ya que permite una gestión apropiada de los datos que son ingresados por el usuario o adquiridos de fuentes externas como sensores. A continuación, se describe brevemente el propósito de los componentes del framework:

**Modelo:** Encargado de gestionar el modelo de base de datos relacional diseñado para este proyecto y de proveer los datos necesarios a las vistas cuando se realicen visualizaciones periódicas gestionadas por las plantillas. El modelo permite la representación de cada tabla con información de la base de datos como clases en Python y cuyos atributos se relacionan directamente con las columnas.

**Vistas:** Gestiona las funciones necesarias para la comunicación con los datos y que permite procesar y almacenar la información de cada uno de los datasets considerados. Las vistas son el componente intermediario entre el modelo (datos) y los templates (interfaz de usuario)

**Plantillas o Templates:** Se encarga de la programación de las interfaces de usuario con HTML, JavaScript y CSS, utilizando librerías de visualización desarrolladas específicamente para generar los gráficos de acuerdo con los datos. De igual forma en esta sección se trabajará en el diseño visual de la aplicación web. La librería considerada para el propósito de visualización es D3.js y los mapas interactivos mediante la librería Leaflet.js

### 2.3.2 Pronóstico de series temporales con ARIMA

La predicción del clima se encuentra dentro de los estudios de pronóstico con mayor importancia, debido a que muchos sectores en la industria, incluyendo la agricultura, dependen de las condiciones climáticas y

ambientales (Paras et al., 2009). Las mediciones de temperatura, humedad relativa, radiación solar, dirección y velocidad del viento, precipitación, etc. son observaciones ordenadas cronológicamente asociadas a momentos puntuales o a intervalos de tiempo específicos, lo que permite su tratamiento como series temporales.

En escalas de tiempo, desde periodos cortos a extensos, pensamos en el clima como un fenómeno persistente ya que, por ejemplo, encontramos épocas secas y épocas húmedas, periodos fríos y cálidos. Por este motivo cuando trabajamos con observaciones meteorológicas, a menudo los datos son suavizados o promediados con la finalidad de remover ruido y obtener una percepción del comportamiento de la serie o indicador climático (Katz & Skaggs, 1981).

El uso de métodos estadísticos u otros medios para pronóstico de series temporales de indicadores climáticos es útil debido a las formas en que las observaciones son registradas y gestionadas. Muchos de estos métodos están basados en el uso de datos históricos con el propósito de encontrar tendencias en el pasado que puedan predecir comportamientos o eventos futuros; si bien en los últimos años se han desarrollado métodos que utilizan técnicas de inteligencia artificial, como redes neuronales artificiales, muchos de los estudios han analizado el uso de métodos estadísticos como ARIMA para pronósticos (Murat et al., 2018).

Los modelos ARIMA, introducidos por primera vez por Box y Jenkins (1976), son buenos para este tipo de análisis ya que su metodología combina tres modelos estadísticos: Autorregresivo, Integrado y Media Móvil y consta de 3 parámetros  $p$ ,  $d$  y  $q$ . El valor  $p$  es el orden autorregresivo,  $e$  indica que la variable de la serie es explicada por las observaciones pasadas o de periodos anteriores, el parámetro  $d$  es el valor de diferenciación o cuantas veces el método debe ser diferenciado para ser estacionario y el valor  $q$  indica el orden de medias móviles (De

Arce & Mahía, 2003). En general el objetivo de los modelos ARIMA es pronosticar o predecir un valor futuro en la serie de tiempo a partir de la combinación lineal de las observaciones previas y sus errores (Mehrmolaei & Keyvanpour, 2016).

El modelo ARIMA ha sido utilizado en diversas aplicaciones relacionadas a la economía, medicina, negocios, finanzas e ingeniería, entre otros; sin embargo, su uso se ha extendido en numerosas aplicaciones meteorológicas con el propósito de comprender diferentes fenómenos relacionados con la temperatura del aire y la precipitación (Murat et al., 2018), considerando a la temperatura como uno de los indicadores climáticos esenciales para la sostenibilidad de la agricultura, el agua, y los recursos (Brath et al., 1999).

Diferentes estudios han sido realizados utilizando modelos ARIMA con la finalidad de evaluar su utilidad en aplicaciones agroclimáticas, por ejemplo: Balyani et al. (2012) trabajó con datos de alrededor de 50 años del sur de Irán, con la finalidad de modelar temperaturas utilizando ARIMA como modelo estadístico. Khedhiri (2016) realizó un estudio sobre las propiedades estadísticas de los datos históricos de temperaturas en Canadá con datos de un periodo de 100 años, desarrollando un modelo ARIMA para pronosticar datos de temperatura en tiempos futuros. Akpanta, Okorie, & Okoye (2015) utilizaron SARIMA como modelo de pronóstico para precipitaciones mensuales en Umuahia, Nigeria basados en su frecuencia. Murat et al. (2018) utilizaron ARIMA para modelar series univariadas de temperaturas del aire y precipitación con información de 30 años en cuatro ciudades europeas (Jokioinen, Dikopshof, Lleida and Lublin), a este modelo se le incorporó la tendencia como una función polinomial y la estacionalidad con series de Fourier.

Los parámetros  $p$ ,  $d$  y  $q$  del modelo ARIMA, utilizado para el pronóstico de temperaturas promedio en el módulo de clima para el desarrollo del

presente proyecto fueron generados de forma automática utilizando el software R con la librería fpp2 y forecast, y su implementación en Python mediante el uso de la librería statsmodel.

## CAPÍTULO 3

### 3. DISEÑO E IMPLEMENTACIÓN DE LA HERRAMIENTA

#### 3.1 Análisis, exploración y validación de datos y fuentes.

El objetivo principal de la solución basada en datos descrita en el presente documento es proporcionar una herramienta o aplicación web, que permita a los encargados de las actividades operativas en una hacienda de producción del cultivo de cacao conocer, de acuerdo con la información ingresada al sistema, la distribución del personal en las diferentes labores agrícolas, la dinámica en la producción obtenida y su relación con un muestreo de mazorcas, así como la visualización de las condiciones climáticas obtenidas en la zona en la que se desarrolla el cultivo. Por este motivo y debido a que en la agricultura existen diferentes factores que contribuyen al manejo apropiado durante el crecimiento de un determinado cultivo, fue importante definir aquellos que, dentro de una aplicación web, aportarían con información que permita monitorear y evaluar tendencias en la producción en una etapa preliminar. En este caso, la herramienta permite conocer la distribución del personal para la realización de diferentes actividades agrícolas y de esta forma optimizar y gestionar apropiadamente los recursos necesarios con los que cuenta la empresa; de la misma forma, esta información hace posible la realización de ajustes en la planificación a lo largo del año.



Durante el proceso de diseño y previo al desarrollo de la herramienta fue importante realizar un levantamiento exhaustivo de requerimientos, con el propósito de conocer todos los detalles y variables involucradas en el registro de datos, así como las tareas necesarias que son ejecutadas actualmente por cada una de las personas que participan en el proceso de generación de reportes periódicos y que contribuyen a la evaluación de métricas y resultados. En este sentido y de acuerdo con el diseño de la aplicación web se cuenta con tres fuentes de información que provienen de registros que requieren de la participación del personal operativo, y una fuente que se obtiene de forma automatizada ya que proviene de sensores meteorológicos. A continuación, se describe el proceso de obtención de esta información y los atributos de cada uno de los datasets que fueron considerados en el desarrollo de este proyecto:

### **Actividades agrícolas**

Las actividades operativas en una hacienda de producción de cacao están relacionadas con el manejo requerido para el desarrollo apropiado del cultivo, las mismas que deben ser planificadas de acuerdo con los requerimientos propios de la agricultura, la labranza del suelo, el uso del agua, uso de fertilizantes, el clima, etc. Actualmente este proceso es realizado semanalmente por los encargados de cada sector productivo, con el detalle de cuantas personas y a que labores se dedicarán en la siguiente semana; esta información puede ser acumulada en un cada 4 semanas (periodo) y tener así un detalle de la distribución del personal durante los 13 periodos del año y como esto cambia en el tiempo; de forma similar, la información puede ser traducida en indicadores que se relacionan directamente con el tamaño o área de producción del cultivo, en este caso podría reflejar la cantidad de personas para realizar una actividad específica en un tamaño de suelo o hectáreas. Los atributos que serán utilizados en el proyecto para el módulo de visualización son: Fecha, Periodo, Dia, Semana, Mes, Año, Sector, Lote, Jornales y Actividad.

### **Conteo de mazorcas**

El conteo de mazorcas es una actividad que se realiza cada periodo (cada 4 semanas) por el equipo de investigación, con la finalidad de evaluar la cantidad de mazorcas en sus diferentes estados de desarrollo, en los diferentes lotes de producción, ya que el proceso desde que la flor es fecundada hasta que se encuentra lista para su cosecha toma alrededor de 6 meses. El proceso consiste en muestrear la cantidad de mazorcas en los diferentes estados clasificados en 9 tipos, en árboles previamente seleccionados. Los estados posibles de las mazorcas son: Cherule, Pequeña 1, Pequeña 2, Pequeña mediana, Mediana grande, Grande, Madura, Madura pequeña y Enferma.

Los atributos que serán considerados para el módulo de visualización son: Fecha, Periodo, Sector, Lote, Estado de mazorca.

### **Producción**

La información de producción es manejada diariamente por los encargados de cada uno de los 3 sectores productivos. Esta información se compila en periodos de tiempo semanal y periódicos (cada 4 semanas). Para los gráficos de producción general en la hacienda y por cada uno de los sectores se utilizarán los siguientes atributos con la información periódica: Año, Periodo, Sector, Lote y Producción obtenida.

### **Condiciones Climáticas**

La información de las condiciones climáticas es obtenida mediante sensores que están conectados a un registrador de datos principal, el cual envía los datos mediante radio a una computadora que los registra en una base de datos MySQL, estos componentes en conjunto forman la estación meteorológica principal de la empresa. Los datos climáticos son procesados y distribuidos diariamente mediante correo electrónico a los diferentes departamentos o áreas que utilizan esta información con fines operativos y de investigación.

La información si bien es obtenida cada segundo, se almacena en el registrador de datos o datalogger y en la base de datos del computador cada 15 minutos. Con el propósito de facilitar esta información en una herramienta web de fácil acceso, los datos serán obtenidos de forma automática desde esta fuente, ya que se ha configurado un proceso para que la información sea enviada desde el

computador de la estación meteorológica hacia el servidor web en donde se aloja la base de datos de la aplicación.

La estación meteorológica cuenta con sensores que registran los datos descritos a continuación:

- Sensor de temperatura y Humedad relativa: Registra datos de temperatura promedio, máxima y mínima, así como la humedad relativa promedio, máxima y mínima.
- Sensor de presión atmosférica: Provee la información de presión atmosférica promedio en el área en donde se encuentra instalada la estación meteorológica.
- Sensor de velocidad y dirección del viento o Anemómetro: Este sensor provee la información de la dirección y velocidad del viento máxima y promedio.
- Pluviómetro: Sensor que mide la cantidad de lluvia o precipitación obtenida en una cantidad de tiempo determinada.
- Sensor de radiación solar: Provee la información de radiación solar por metros cuadrados, las mediciones proporcionadas son valores de potencia (Watt) y energía (MegaJoules).

Existen tres indicadores adicionales que son calculados combinando los datos obtenidos de los sensores, estos indicadores son: Punto de rocío y Evapotranspiración referencial (ET<sub>o</sub>) y Grados días.

La información que ha sido considerada para el módulo de visualización en la aplicación web es: Temperatura promedio, máxima y mínima, Humedad relativa promedio, Radiación solar en valores de energía acumulado por metro cuadrado (MJ), Precipitación o lluvia acumulada, Evapotranspiración referencial y Grados días. Estos indicadores fueron escogidos debido a su relevancia y uso frecuente en la empresa para diferentes actividades agrícolas, así como su uso con fines investigativos.

Con la finalidad de registrar la información de los diferentes datasets hacia la base de datos principal de la herramienta, se realizó el siguiente procedimiento utilizando un notebook de Jupyter.

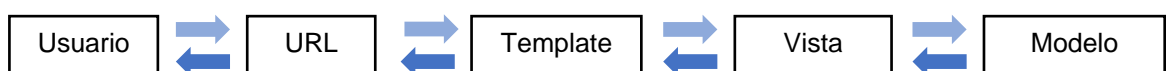
- Los datasets al encontrarse completamente en hojas de cálculo de MS Excel fueron primero convertidos al formato csv.
- Mediante la librería Pandas se cargaron estos datasets en un notebook.
- Se creó la base de datos con las tablas que contendrían la información en el formato deseado (tablas vacías).
- Mediante la función `to_sql` de un dataframe en Pandas se envió toda la información a cada una de las tablas.

### 3.2 Arquitectura y diseño de la aplicación web

#### Arquitectura de la herramienta de visualización: Django y MVC

La herramienta de visualización o aplicación web descrita en el presente documento utiliza Django como framework principal de desarrollo y el patrón de arquitectura de software MVC o Modelo Vista Controlador (MTV o Modelo Template Vista en Django). El motivo de la elección de Django radica en que este framework trabaja fundamentalmente con el lenguaje de programación Python, lo cual hace posible el uso de librerías para ciencia de datos directamente en el entorno web con la información ingresada a la base de datos desde la aplicación; a su vez, Django permite crear funcionalidades para la visualización de esta información con librerías de Javascript en las interfaces web del lado del cliente.

El patrón de diseño Modelo Vista Controlador es una arquitectura de software que tiene como objetivo separar en capas el código necesario para el funcionamiento de una aplicación web, de acuerdo con las distintas responsabilidades y manteniendo un propósito concreto para cada una. En la Figura 3.1 se detalla el flujo de trabajo de las diferentes capas en el MVC.



**Figura 3.1 Modelo de flujo de datos en Django**

Al utilizar este patrón de diseño se asegura la calidad en las diferentes aplicaciones, así como una mejora en la estructura del código fuente, debido a que separa las funciones y rutinas permitiendo su reutilización y mantenimiento. Dado que el patrón permite desglosar y asignar responsabilidades, uno de sus beneficios es la mejora en la interacción de trabajo entre profesionales encargados del desarrollo de un determinado software. Django trabaja con la estructura del patrón MVC, pero con ligeros cambios en su denominación, ya que para Django los Templates son las Vistas y las Vistas son los Controladores, en este sentido se podría decir que Django sigue un patrón de diseño MTV. A continuación, se describen cada una de las capas.

### **Modelo**

El modelo se encarga de la gestión de los datos, así como de las funciones necesarias el acceso, validación, creación, modificación o eliminación. Esta capa puede contener funciones que interactúan directamente con la base de datos mediante consultas y sentencias propias del motor seleccionado. De la misma forma el modelo es el encargado de conocer las relaciones entre las diferentes tablas y gestionar el mapeo relacional de objetos (ORM), el cual permite y facilita la interacción con la base de datos mediante instrucciones de Python similar al uso de sentencias SQL.

El archivo principal que contiene todas las funcionalidades de los modelos es *model.py* y contiene la descripción de cada una de las tablas utilizadas en el proyecto en forma de clases Python. Las clases que han sido generadas mediante el ORM a partir de los datasets son:

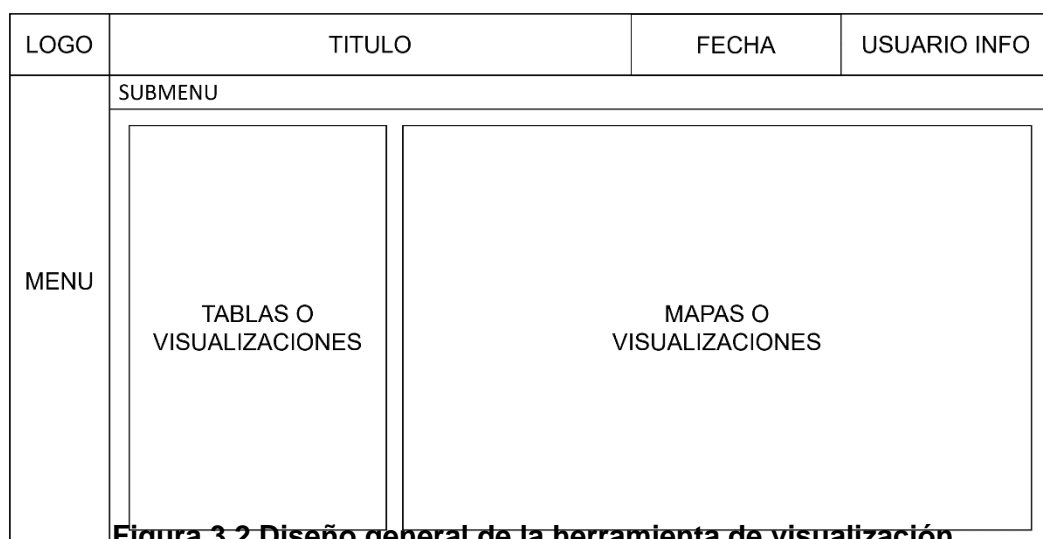
- FieldActivity
- WeatherConditions
- Production
- PodCounting

### **Vista (Templates en Django)**

Los templates en Django son los encargados de gestionar las interfaces web que el usuario visualizará, para esto Django trabaja con funciones en el archivo *views.py* que definen la conexión con la URL a la que se enviará información específica en formato de diccionarios. Los templates esencialmente contienen la estructura HTML de un sitio web y que a su vez puede contar con estilos basados en reglas CSS y funcionalidades mediante el uso de Javascript (JS). Es en esta capa en donde todo el diseño de la interfaz web debe ser realizado, así como la estructura de las funciones que van a realizar las diferentes visualizaciones utilizando las librerías de D3.js y Leaflet.js.

Debido a que Django cuenta con un sistema de tags o etiquetas `{% %}` es posible que un template o archivo html pueda *heredar* código de otro y trabajar con bloques que cambian dinámicamente, dependiendo de la sección en la que nos encontremos; en este sentido el diseño integral del sitio web debe ser elaborado inicialmente considerando cada una de las opciones ya que los cambios en las visualizaciones serán gestionados por las vistas cada vez que el usuario se desplaza por el sistema.

Para la gestión apropiada de la información en todo el flujo de trabajo de la aplicación, se ha creado un template HTML para cada uno de los componentes. La Figura 3.2 muestra un diseño general de la aplicación web a nivel de template html que es renderizado y modificado dinámicamente, de acuerdo con las opciones que han sido desarrolladas para el uso de los encargados de las actividades operativas en una hacienda de producción de cacao.



Los templates con los que cuenta la aplicación se describen a continuación:

### **Index**

Es el template original de donde se derivan todos los demás templates, y es el encargado de contener la estructura de todas las páginas web de la aplicación, así como sus estilos y funcionalidades; es aquí en donde se encuentra el menú principal, información de periodos y fechas, y los contenedores que otros templates modificarán dinámicamente.

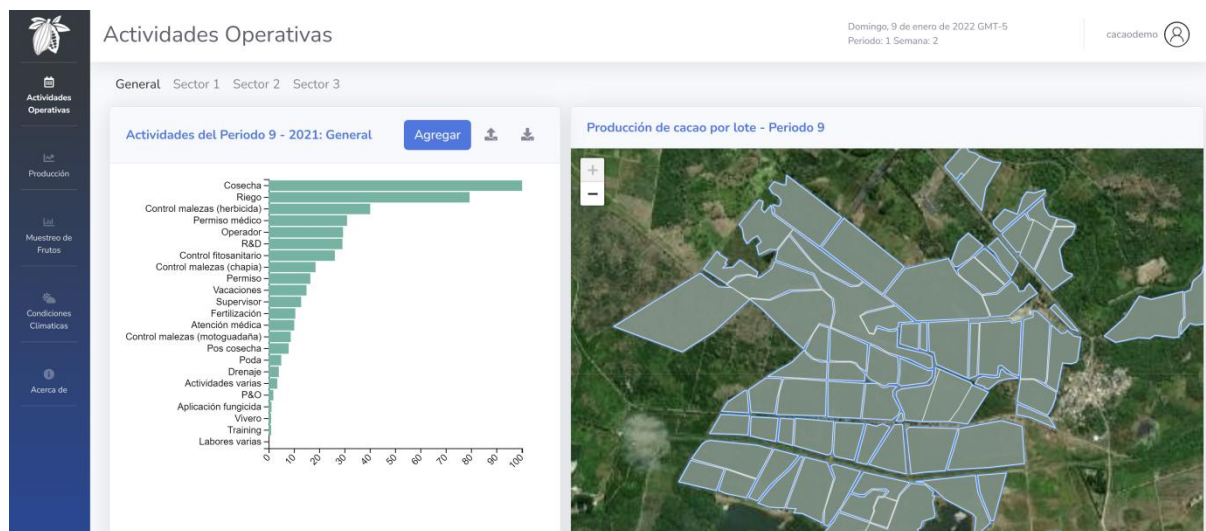
### **Actividades operativas**

En este template se encuentran las diferentes visualizaciones relacionadas a las actividades y labores agrícolas, así como un mapa que muestra que actividad se está relacionando en esa semana por cada lote productivo. Es en esta sección en donde los usuarios pueden agregar información mediante formularios que se conectan a la base de datos para registrar las características de cada una de las actividades, que posteriormente modificarán dinámicamente los gráficos de barras diseñados para este propósito. Este es el template que se muestra por defecto al ingresar las credenciales del usuario.

Las visualizaciones de las labores agrícolas son esencialmente gráficos de barras que muestran de forma ordenada las actividades más demandantes en toda la hacienda por cada sector productivo, considerando el número de jornales necesarios para ejecutar cada actividad; esto puede ser interpretado como la cantidad de personas requeridas para dicho propósito. Estas visualizaciones funcionan de manera dinámica y automatizada, y han sido implementadas utilizando la librería D3.js.

Esta información permite conocer la distribución del personal en las diferentes actividades agrícolas dentro de la empresa por cada periodo (cada 4 semanas), y conforme se registre semanalmente la planificación, el grafico cambiará hasta que

se complete el ciclo para el siguiente periodo y así durante todo el año. De la misma forma existe en este template un mapa de visualización de actividades por lotes productivos implementado utilizando la librería Leaflet.js; esto ayuda a tener en una representación gráfica de cómo se encuentra distribuido el personal por actividades en la empresa, y permite no solo al equipo operativo tomar decisiones, sino que también hace posible, conocer aquellas áreas en las que se recomienda no transitar si se está realizando alguna aplicación de pesticida o fungicida. La Figura 3.3 muestra el diseño del template desarrollado para el módulo de actividades agrícolas.



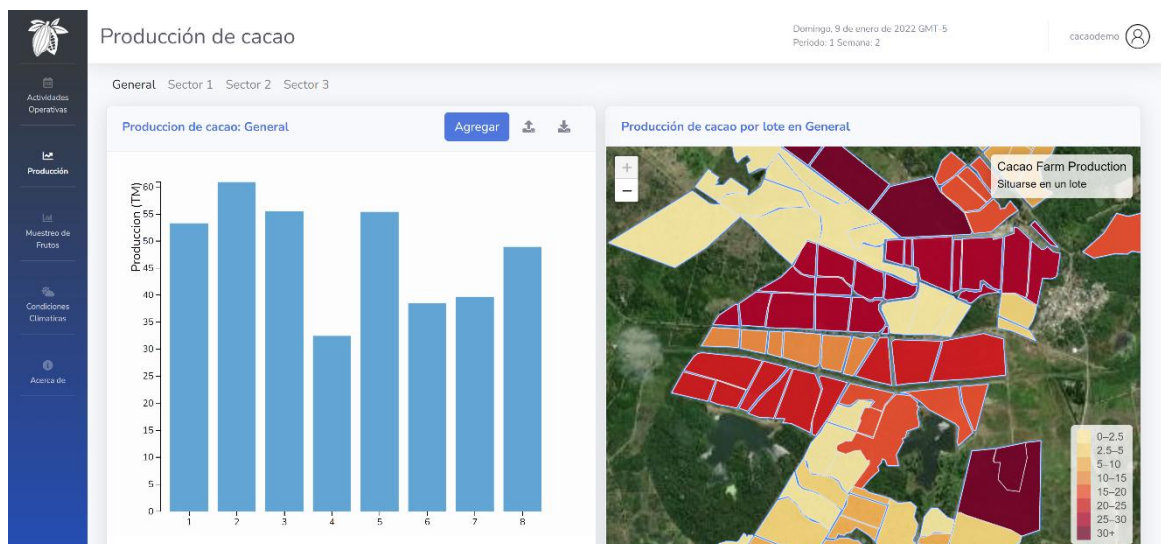
**Figura 3.3 Diseño del módulo de visualización de actividades operativas**

## Producción

La información de producción es uno de los indicadores más importantes en una hacienda dedicada al cultivo de cacao, ya que esta información es revisada de forma periódica y permite conocer el estado de la plantación en términos de cosecha. Esta información posibilita, de la misma forma, conocer o estimar indicadores económicos ya que se relacionan directamente con el producto final de una hacienda: El cacao, que será vendido para luego ser convertido en productos alimenticios. Este indicador se relaciona directamente con las métricas del negocio, ya que puede ser traducido en términos financieros y económicos.



La información de producción puede ser representada tanto como un gráfico de líneas, en serie de tiempo o como un gráfico de barras, dado que el propósito es comunicar la cantidad de cacao cosechado en unidades de peso de almendras secas, estas unidades pueden ser kilogramos, quintales, o toneladas métricas. En



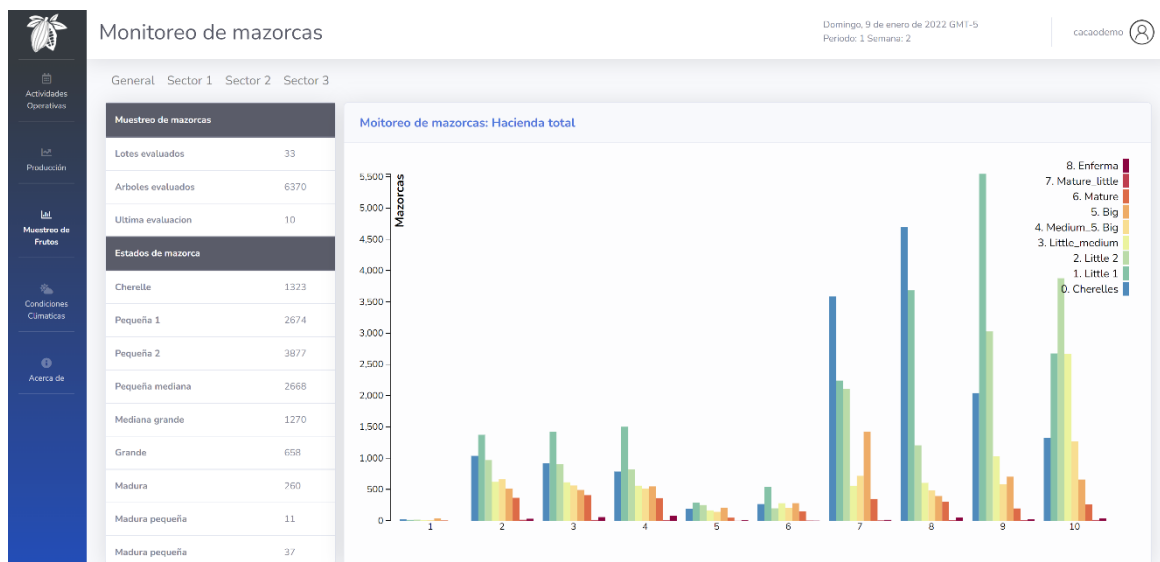
**Figura 3.4 Diseño del módulo de visualización de producción**

la aplicación web se dispone de un template único para la producción, ya que estará representado en dos visualizaciones: un gráfico de barras y un mapa interactivo que muestra la cantidad cosechada por cada uno de los lotes productivos. Esto es importante porque si bien es posible tener un gráfico general de la producción total en la hacienda, debe ser posible conocer esta información en cada uno de los sectores productivos, en cada periodo y en cada lote, para de esta forma, asistir en la toma de decisiones. La Figura 3.4 muestra una de las propuestas de mapa interactivo para conocer la productividad de cacao en la hacienda.

### Conteo de mazorcas

El conteo de mazorcas es una actividad que se realiza cada periodo (cada 4 semanas) que consiste en cuantificar las mazorcas de cacao en los diferentes lotes productivos de acuerdo con su estado de desarrollo, con la finalidad de obtener una estimación de cuanto se cosechará próximamente.

La visualización escogida para esta información es un gráfico de barras múltiples, ya que la cantidad de mazorcas muestreada está clasificada en 9 tipos: Cherelle, Pequeña 1, Pequeña 2, Pequeña mediana, Mediana grande, Grande, Madura, Madura pequeña o Enferma. Esta visualización se realizará utilizando la librería D3.js para generar un gráfico dinámico que permite conocer en detalle la cantidad por periodo de cada uno de los estados de las mazorcas en la hacienda. De la misma forma, una tabla con las diferentes cantidades se muestra junto al gráfico (Figura 3.5).



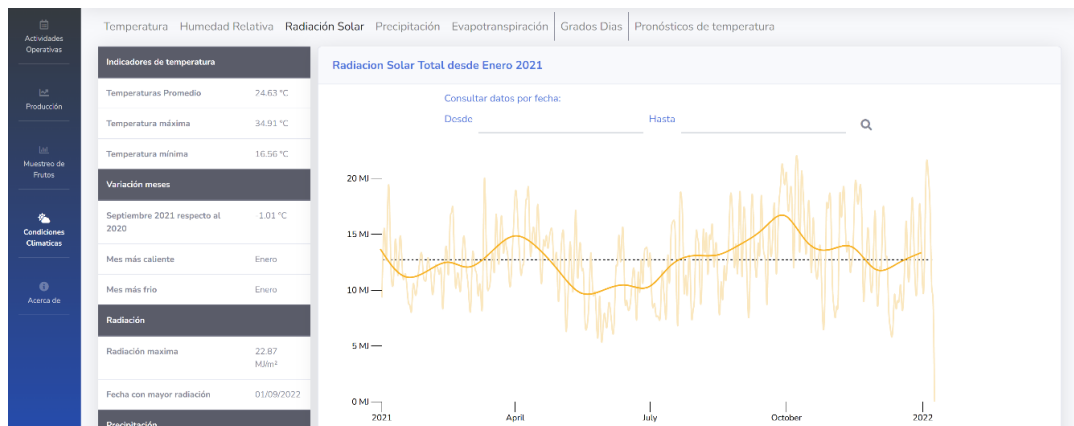
**Figura 3.5 Diseño del módulo de visualización de conteo de mazorcas**

### Condiciones climáticas

El clima es un factor importante que debe ser considerado en agricultura, es por este motivo que la información climática, obtenida directamente de sensores meteorológicos en una estación de clima automatizada, ha sido incorporada como componente en la aplicación web descrita en este documento. Las variables de interés principales que han sido escogidas para ser representadas de forma gráfica son: Temperatura, Humedad Relativa, Radiación Solar, Precipitación, Evapotranspiración y Grados días.

La representación considerada para todos los indicadores a excepción de la precipitación o lluvia, son gráficos de línea ya que los datos pueden ser

modelados como series temporales; la lluvia se representará mediante un gráfico de barras, ya que muestra la información semanal acumulada de las precipitaciones. Esta implementación se realiza utilizando la librería D3.js, ya que permite gráficos dinámicos y animaciones. De la misma forma, en cada uno de los indicadores al ser subsecciones del template de clima, se mostrará una tabla con analítica de los datos respecto a años anteriores con la finalidad de obtener de forma inmediata el comportamiento de las condiciones climáticas del año en curso. En las Figura 3.6 se muestra el diseño de los gráficos desarrollados para cada indicador climático.



**Figura 3.6 Diseño de la visualización para indicadores climáticos**

### Controlador (Vistas en Django)

Las vistas en Django se encargan del procesamiento de los datos previo a una visualización en la interfaz web, y son un intermediario entre el usuario y los datos. De forma similar a los componentes MVC descritos previamente e implementados en este proyecto, se ha considerado disponer de al menos una vista que procese los datos con los que el usuario interactuará y visualizará.

Las vistas son esencialmente funciones de Python que gestionan una petición realizada por el usuario en el navegador web y dependiendo de su propósito, devolverán información con los datos de un determinado modelo. A continuación, se detallan las vistas implementadas para cada uno de los templates del proyecto y que se encuentran en el archivo *views.py*.

**Tabla 3.1 Funciones en la vista Index**

<b>Index</b>	
<b>Función</b>	<b>Objetivo</b>
<b>loadIndex</b>	Generar la información para la vista por defecto (index) de la solución.

**Tabla 3.2 Funciones en la vista Actividades operativas**

<b>Actividades operativas</b>	
<b>Función</b>	<b>Objetivo</b>
<b>actividadesOperativas</b>	Proporcionar la información de las actividades operativas por periodo de tiempo y sector productivo a su respectivo template. Esta información viaja como registros diarios que, mediante javascript se resumen en cantidad de jornales por tipo de actividad para cada periodo (acumulados).

**Tabla 3.3 Funciones en la vista Condiciones Climáticas**

<b>Condiciones climáticas</b>	
<b>Función</b>	<b>Objetivo</b>
<b>condicionesClimaticas</b>	Proporcionar la información del clima en dos formatos: Información climática del año en curso hasta la actualidad Información histórica de los indicadores climáticos con el propósito de realizar pronósticos de series temporales
<b>getDailyData</b>	Procesar la información obtenida del modelo en forma de registros de cada 15 minutos, el resultado de esta función es devolver un dataframe con los registros en un intervalo de tiempo de 24 horas o 1 día
<b>timeSeriesPrediction</b>	Proporciona un dataframe con predicciones de un indicador climatológico, recibe como parámetros el dataset de la base de datos y los pasos o número de predicciones. El modelo de pronóstico de series de tiempo utilizado es ARIMA.

**Tabla 3.4 Funciones en la vista Producción**

<b>Producción</b>	
<b>Función</b>	<b>Objetivo</b>
<b>loadProduction</b>	Proporcionar la información de producción desde la base de datos, esta información puede ser datos históricos o un periodo

	<p>de tiempo puntual. Esta función se encarga de procesar y resumir los datos de acuerdo con el periodo de tiempo solicitado. La información se visualiza en el template de producción mediante la librería D3.js</p>
--	---

**Tabla 3.5 Funciones en la vista Conteo de Mazorcas**

Conteo de mazorcas	
Función	Propósito
<b>loadPodCounting</b>	Proveer la información del conteo de mazorcas obtenido desde la base de datos, esta función procesa los datos en un dataframe para generar resúmenes en un tiempo determinado, ya que la base de datos cuenta con registros a nivel de plantas evaluada y el lote al que pertenecen. El resultado puede ser a nivel general de la hacienda o por sector o lote productivo.

### 3.3 Prototipos de algoritmos, modelos y módulos del sistema

#### Módulos del Sistema

Debido a que la solución utiliza el patrón de diseño Modelo Vista Controlador o Modelo Template Vista en Django, las diferentes funcionalidades y responsabilidades de sus componentes están bien definidas, y es posible segmentarlas en 4 grupos o módulos, los cuales se detallan a continuación:

Módulos de ingesta de datos (Modelos):

- Datos que son ingresados por el usuario en la interfaz
- Datos obtenidos mediante fuentes externas (sensores de clima)

Módulos de procesamiento y gestión de datos (Vistas):

- Actividades agrícolas
- Condiciones Climáticas
- Producción
- Muestro de mazorcas

Módulos de visualización (Templates):

- Demanda operativa por periodos en gráfico de barras

- Indicadores climáticos en series temporales y gráficos de barras para la lluvia
- Monitoreo en la producción periódica en grafico de líneas
- Conteo de mazorcas en gráficos de barras múltiple

Módulos de predicción (Vista y Template):

- Pronóstico de series de tiempo para temperaturas promedio

El diagrama de la Figura 3.7 ilustra la interacción de los módulos durante el flujo de la información en el sistema, en donde la aplicación permite la ingesta de datos, para luego ser procesada en las funciones o vistas MVC de Django y que proporcionan los datos necesarios para las diferentes visualizaciones, o en el caso de la temperatura promedio, también hacer predicciones para días próximos. De la misma forma el diagrama especifica en que componente del patrón MVC se encuentran los diferentes módulos.

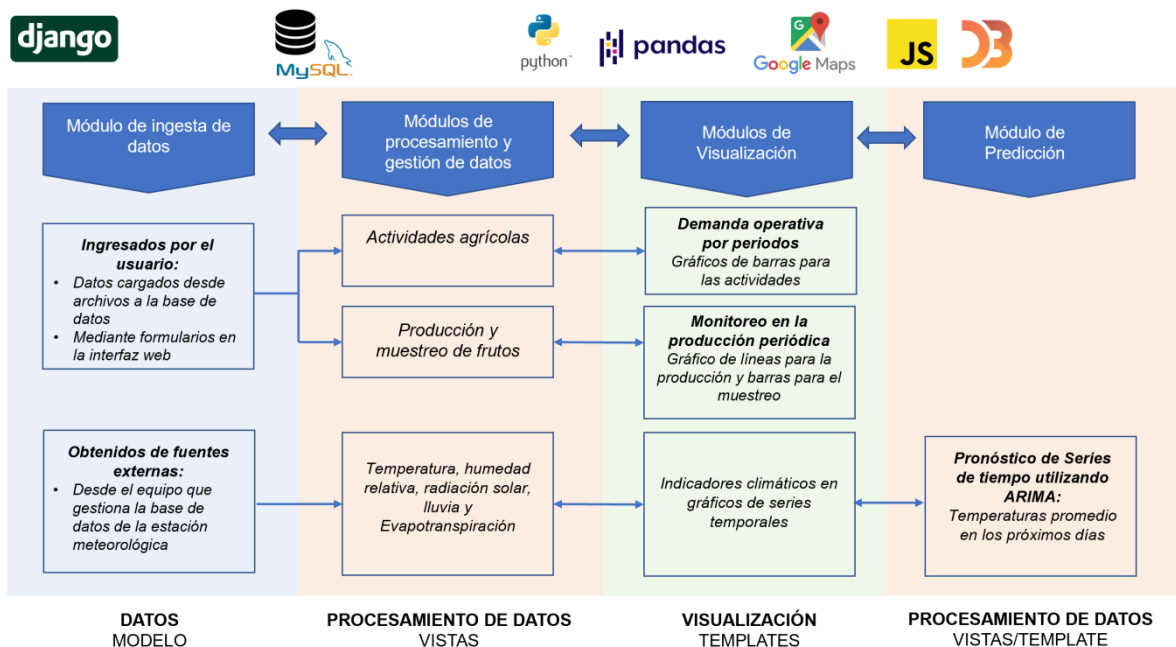


Figura 3.7 Diagrama de módulos de la herramienta de visualización

### Módulo de Predicción de temperaturas promedio

Una de las características desarrolladas para este sistema es la posibilidad de realizar predicciones de indicadores climáticos, específicamente de la temperatura

promedio, con la finalidad de proveer al equipo operativo un componente adicional en un software de visualización que les permita conocer con anticipación el comportamiento estimado del clima. Esta información es importante, ya que el clima está relacionado a muchas actividades agrícolas y procesos de desarrollo del cultivo.

Para este efecto se ha realizado un análisis preliminar en el lenguaje de programación R de los posibles modelos de series temporales que podrían ser utilizados en el sistema. Para el dataset, se utilizó la información del clima obtenida de la estación meteorológica entre enero de 2019 y agosto de 2021. Este indicador ha sido considerado inicialmente debido a las características de la serie temporal que presenta, ya que cuenta con estacionalidad y ciclos definidos a lo largo del año. Con la finalidad de obtener métricas de desempeño se ha dividido al dataset en datos de entrenamiento y datos de prueba.

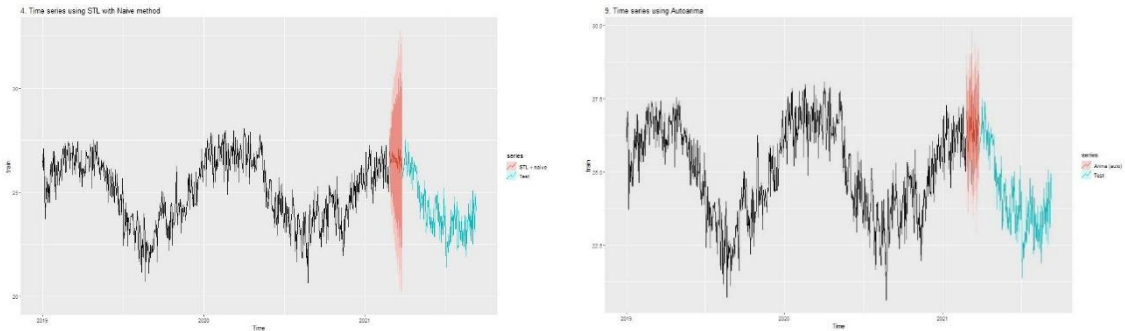
El modelo fue ajustado a los datos de entrenamiento y la validación se realizó con predicciones y datos de prueba. Los resultados obtenidos se detallan en la tabla 3.6 con los 5 modelos seleccionados.

**Tabla 3.6 Resultado de diferentes modelos de pronóstico para series temporales**

<b>METODO</b>	<b>ME</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>MASE</b>
<b>Naive</b>	-1.775	2.253	1.889	-7.602	8.023	2.105
<b>SNaive</b>	-0.572	1.323	1.070	-2.386	4.425	1.193
<b>Drift</b>	-1.781	2.260	1.894	-7.628	8.047	2.111
<b>STL</b>	-0.631	<b>0.991</b>	0.839	-2.510	3.278	0.935
<b>ARIMA</b>	-0.587	<b>1.115</b>	0.921	-2.346	3.585	1.026

De acuerdo con los resultados, el método STL y el método ARIMA producen los valores más bajos para el Error Cuadrático Medio (RMSE), utilizando datos de prueba, por lo que indican que ambos modelos tienen un buen desempeño en las predicciones realizadas. Para el análisis, se consideró máximo un periodo de tiempo de 30 pasos adelante o 30 días de pronóstico, con el propósito de realizar las respectivas validaciones. La Figura 3.8 indica gráficamente el comportamiento

de los modelos STL y ARIMA respectivamente, y las predicciones en comparación con los datos de prueba reales (en color rojo).



**Figura 3.8 Pronósticos de modelos STL y ARIMA en R**

El método ARIMA fue seleccionado para ser implementado directamente en una de las vistas o funciones de Django, ya que trabaja directamente con el lenguaje de programación Python y no requiere de un procesamiento externo en R cada vez que el usuario desee realizar predicciones desde la interfaz web. Para este propósito se utilizaron las librerías Pandas, Numpy y Statsmodel y se creó la función *timeSeriesPrediction* que recibe como parámetro un dataset con datos diarios de temperatura, y la cantidad de predicciones deseada, la salida es un dataframe con las predicciones y los días específicos en los que se estima un valor que genera el modelo.

Los datos son procesados por la función *condicionesClimaticas* en Django y la información o resultado es enviado al template de clima en un diccionario que contiene varios datasets en formato json con la siguiente información:

- Datos del indicador climático de 2021
- Datos de predicción en intervalos de 7, 15 y 30 días
- Datasets combinados del indicador climático y las predicciones en intervalos de 7, 15 y 30 días





**Figura 3.9 Visualización de temperaturas promedio (verde) y pronóstico a 30 días**

El motivo de enviar varios datasets desde el procesamiento al template o interfaz web, radica en que en una sola petición se inicializarán variables en Javascript con la información disponible. Esto hace posible que los datos se encuentren “precargados” y se disminuya el tiempo de espera, cada vez que el usuario desee cambiar el número o días a predecir.

Con los datos combinados se han desarrollado funciones en Javascript que tienen el objetivo de diferenciar mediante colores, los datos obtenidos con sensores y los datos estimados o pronosticados directamente con la librería D3.js. La Figura 3.9 indica el resultado de una visualización en la que se realizan predicciones, generando los resultados en un formato de tabla.

### 3.4 Infraestructura para procesamiento y almacenamiento

Debido al objetivo de la solución, todo el despliegue será en servidores web y de base de datos. El patrón de diseño MVC o MTV con Django, permite que el procesamiento de los datos se realice directamente en el framework. El archivo *views.py* es el encargado de contener todas las funciones de procesamiento de los datos, a partir de los modelos, como una representación abstracta de las tablas de la base de datos principal.

En este sentido toda la interacción y flujo de datos se da en el mismo entorno, del lado del servidor, facilitando la información necesaria para generar las visualizaciones del lado de cliente, debido a que en este lado de la arquitectura web, el usuario accede solo a archivos HTML, CSS y Javascript. Esta herramienta de visualización puede funcionar en un servidor único, dado que la infraestructura para el procesamiento de la información y los recursos necesarios no requiere de una arquitectura compleja.

En término de datos, si no se dispone de un servidor de base de datos, esta puede residir en el mismo servidor de aplicaciones y se requerirá tener instalado MySQL Server para poder alojar la información. La cantidad de datos que se procesarán, si bien es grande, no se consideran masivos; por lo que no se requeriría de una arquitectura de Big Data para el manejo y procesamiento. Además, en la mayoría de los componentes o módulos, no se requiere de datos en tiempo real, únicamente los indicadores climáticos tienen una frecuencia de actualización alta.

Finalmente se puede concluir que el código desarrollado en este proyecto utilizando el framework Django, puede ser desplegado en cualquier servidor o equipo que tenga todas las configuraciones necesarias para el acceso a usuarios. Si se desea habilitar esta herramienta en la web, será necesario el uso de un dominio o dirección asociada a una IP.

### **3.5 Plataformas y prototipos de visualización**

La herramienta de visualización descrita en el presente documento ha sido desarrollada y diseñada para ser utilizada como una aplicación web en su totalidad, evitando el uso de software externos para el procesamiento y/o visualización. Inicialmente el diseño de la aplicación fue realizado utilizando el software Xampp como servidor web local, en el computador utilizado para el desarrollo del proyecto; sin embargo, todas las pruebas fueron migradas en etapas tempranas a Django con la finalidad de asegurar la consistencia en la

ejecución las diferentes funciones implementadas para el procesamiento y visualización.

El análisis de series temporales, con la finalidad de definir el modelo óptimo para el pronóstico de temperaturas promedio, fue realizado con R y RStudio, lo cual permitió generar visualizaciones de las diferentes predicciones. Una vez concluido el análisis, estas características fueron implementadas en Django, mediante una función que se encarga de generar las predicciones a partir de un dataset. El desarrollo inicial del proyecto fue realizado en un servidor web local proporcionado por Django, sin embargo, el resultado será presentado en un servidor web remoto, con un dominio específico, con el propósito de permitir que todos los usuarios finales tengan acceso a la herramienta.

### **3.6 Métricas y comunicación de resultados**

La herramienta de visualización y monitoreo descrita en el presente documento tiene como finalidad proveer de una solución basada en datos, que permita la automatización de varias de las tareas que actualmente son realizadas de forma manual por el equipo operativo, encargado de dirigir y gestionar las labores agrícolas en una hacienda de producción de cacao.

Las actividades que se reducirían o suprimirían en su totalidad con la implementación de esta solución son:

- Generación de reportes con la información de jornales por periodo (actividades operativas) utilizando hojas de cálculo.
- Generación de reportes con la información climática y la baja posibilidad de acceso inmediato a los indicadores para realizar análisis.

Las posibilidades con las que actualmente la empresa no dispone y que esta solución facilitaría son:

- Visualización dinámica y automatizada de la información de producción de cacao directamente en mapas interactivos.

- Visualización dinámica de la información de producción de cacao en gráficos por periodos.
- Visualización dinámica de la distribución del personal en actividades agrícolas.
- Almacenamiento centralizado de la información y la posibilidad de acceso de las personas interesadas en los datos.
- Visualización dinámica del muestro de mazorcas realizado de forma periódica.
- Visualización dinámica y automática de los indicadores climáticos más importantes identificados en el cultivo de cacao.
- Pronóstico de temperaturas promedio en la hacienda para intervalos desde 1 hasta 30 días.

Debido a que la herramienta mejora los procesos necesarios para la obtención de resultados a partir de los datos, es posible concluir que la implementación y uso de esta generaría una optimización en el tiempo para los usuarios que actualmente elaboran reportes de forma manual. El tiempo estimado que se optimizaría sería de alrededor del 7.5% y 10%, considerando que una persona utiliza entre 3 y 4 horas de trabajo para realizar reportes relacionados al clima; sin embargo, este número aumenta un 5% más si se considera que en cada periodo los usuarios realizan diferentes reportes de acuerdo a los datos de interés con la finalidad de evaluar resultados.

# CAPÍTULO 4

## 4. ANÁLISIS DE LOS RESULTADOS

### 4.1 Recolección de datos y estrategias para validación del proyecto

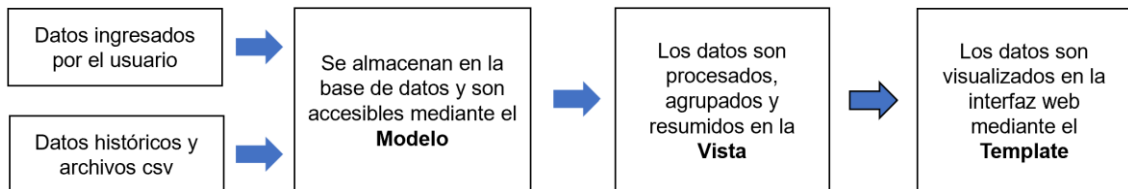
Durante la implementación de los diferentes módulos del proyecto se requirió el desarrollo de metodologías, que permitan el ingreso apropiado de los datos en las diferentes tablas diseñadas para cada uno de los módulos de la solución. Para este propósito se crearon funciones en las diferentes vistas de Django que permiten procesar los datos ingresados, convirtiéndolos principalmente en dataframes para poder utilizar todas las funciones de limpieza y preprocesamiento de la librería Pandas.

Este proceso fue fundamental, debido a que, dentro de los recursos, se disponía de información histórica que ha sido registrada manualmente durante meses, y gestionada por las diferentes personas que se encargan del manejo de la información operativa de la Hacienda. Esta información es analizada periódicamente, con la finalidad de obtener resultados que guíen en la planificación anual y asistan en la toma de decisiones que optimicen el uso de recursos en la empresa.

La metodología desarrollada no solo contribuyó al diseño de las tablas en la base de datos y su respectivo almacenamiento, sino que permitió disponer de funciones que hacen posible la carga de datos que actualmente se encuentran en hojas de cálculo, con el propósito de procesarlos y analizarlos de forma automática. De esta forma, además de los formularios de registro de datos, es posible cargar la información desde fuentes externas.

El propósito es minimizar el impacto en la transición del uso de un software de utilitarios a una herramienta de visualización de datos automatizada, con la información disponible. La Figura 4.1 muestra la metodología desarrollada para alimentar el sistema de datos y mostrarlos de forma periódica de acuerdo con las necesidades del usuario. Esta metodología permite dos formas de registro de información: mediante formularios establecidos con el tipo de datos requerido

para las visualizaciones y mediante archivos externos que serán procesados para obtener los datos necesarios.



**Figura 4.1 Proceso de ingesta de datos, procesamiento y visualización**

### **Validación de los gráficos generados automáticamente**

Con la finalidad de asegurar que los gráficos generados automáticamente en el sistema se realicen de forma correcta, se han implementado diferentes funciones que tomen los registros directamente de la base de datos y los procesen. Debido a la naturaleza de las visualizaciones es necesario un proceso de transformación, ya que muchos registros tienen una frecuencia diaria.

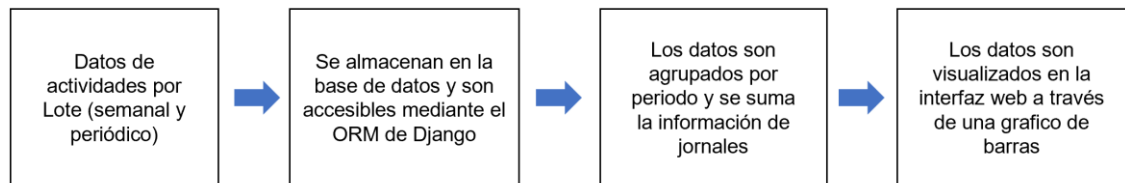
De la misma forma y con el propósito de validar que las agrupaciones de datos y las funciones que se realizan de forma interna con Pandas sean ejecutadas correctamente se utilizaron las siguientes estrategias:

1. Los procedimientos y resultados esperados fueron previamente validados mediante Jupyter Notebook; luego de esto, las instrucciones fueron implementadas en cada vista, de acuerdo con el módulo, en Django.
2. Se utilizó R con RStudio como respaldo para comparar y validar el resultado obtenido en Jupyter Notebook con Python y se verificó la agrupación de: actividades, producción, tipos de mazorca por periodo, etc.

### **Visualización y validación de la información de actividades operativas**

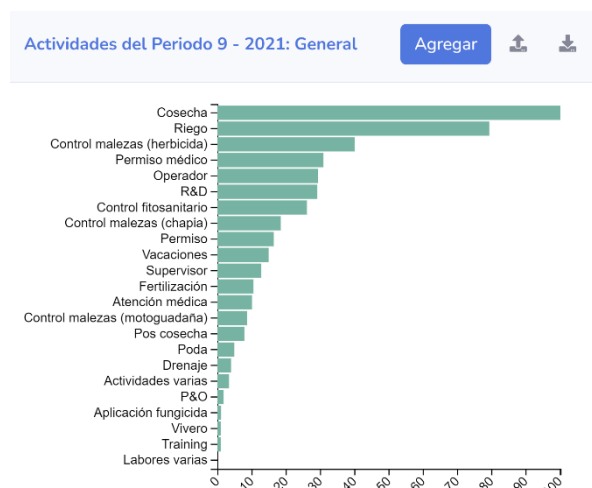
La visualización de las actividades operativas es un gráfico de barras ordenado de forma descendente de acuerdo con la cantidad de jornales (personal) requerido para cada labor. Esta información se registra con una frecuencia diaria para cada actividad y el área productiva en donde se realizará.

El resultado del procesamiento es una visualización con el personal requerido (suma de jornales) por actividad y por periodo (cada 4 semanas), mostrando en el sistema siempre el último por defecto, ya que conforme los técnicos agrícolas agreguen datos, esta visualización se actualizará. La Figura 4.2 muestra el flujo de trabajo necesario para obtener este gráfico.



**Figura 4.2 Flujo de datos para generar visualizaciones de actividades operativas**

La Figura 4.3 refleja este resultado en una visualización interactiva con D3.js en el sistema y la Tabla 4.1 muestra el ejemplo de las actividades realizadas para el Periodo 9 obtenida en un cuadro resumen ordenado (10 actividades) por demanda operativa obtenida con RStudio.



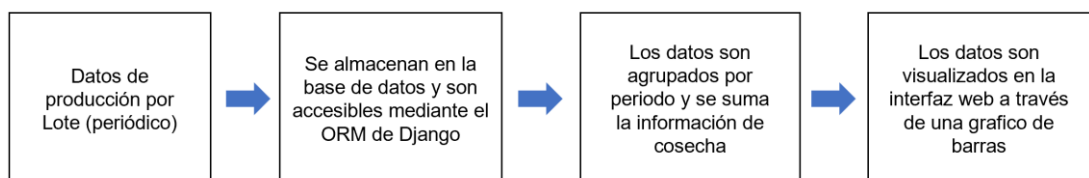
**Figura 4.3 Visualización de actividades operativas**

**Tabla 4.1 Resumen de actividades de un periodo obtenido con RStudio**

Orden	Actividad	Jornales
1	Cosecha	100.18
2	Riego	79.42
3	Control malezas (herbicida)	40.15
4	Permiso médico	31
5	Operador	29.44
6	R&D	29.2
7	Control fitosanitario	26.21
8	Control malezas (chapia)	18.57
9	Permiso	16.51
10	Vacaciones	15

### **Visualización y validación de la información de producción**

El gráfico de la producción obtenida puede ser representado como una serie de tiempo, o un gráfico de barras, ya que los datos están asociados a un periodo en el año. Estos datos se han registrado en la base de datos por periodo (cada 4 semanas) y el agrupamiento se realiza utilizando esta frecuencia, en donde se suman todos los resultados de productividad de todos los sectores y áreas de plantación; sin embargo, esta visualización también permite conocer la producción por sector productivo, en donde el agrupamiento sería primero por sector y luego por periodo, mostrando la suma de los resultados obtenidos. La plataforma por defecto siempre muestra la producción obtenida en toda la hacienda hasta el último periodo de registro. La Figura 4.4 muestra el flujo de trabajo necesario para obtener esta visualización.

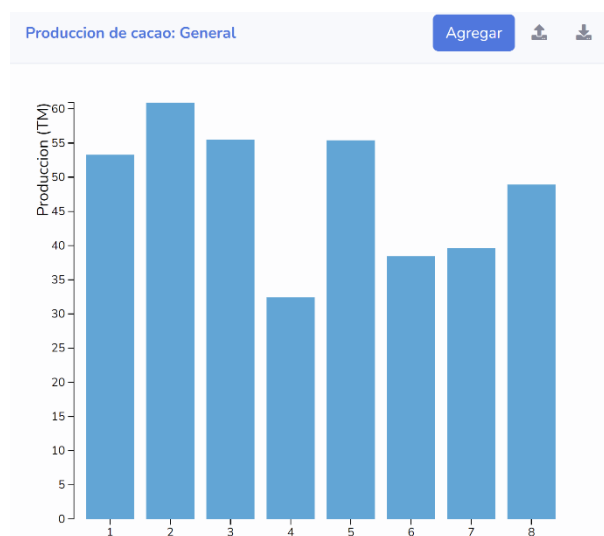


**Figura 4.4 Flujo de datos para generar visualizaciones de producción**



El gráfico con el resultado de la productividad general en lo que va del año se muestra en la Figura 4.5, y los resultados de la visualización han sido validados utilizando RStudio.

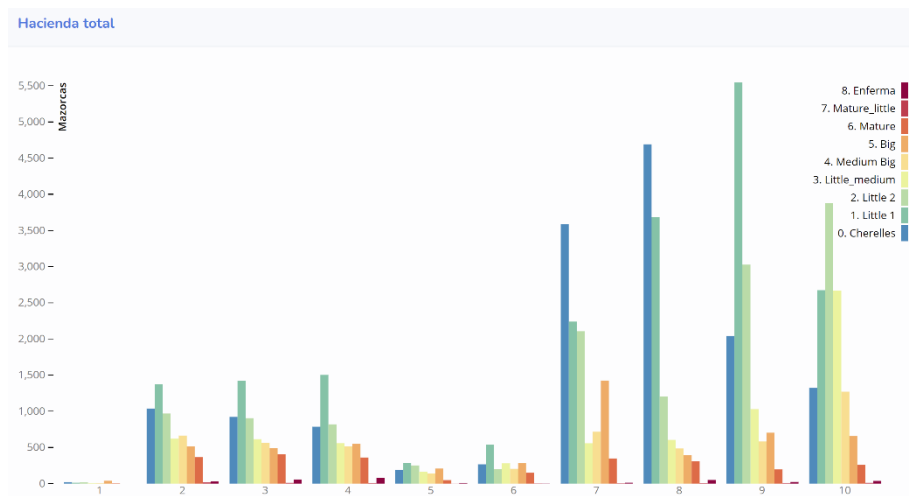
La información de productividad que se almacena en la base de datos proviene tanto de información histórica como de un formulario en el que cada encargado de un sector productivo puede registrar la información de cuanto se ha cosechado por área productiva en cada periodo.



**Figura 4.5 Visualización de producción por periodos**

### **Visualización y validación del monitoreo de mazorcas**

La visualización del monitoreo de mazorcas es un gráfico de barras agrupado por periodos para cada tipo de mazorca que se evalúa por planta en las áreas productivas. Los registros son realizados dos semanas antes de cada periodo, y el nivel de detalle incluye la cantidad de mazorcas por tipo en todas las plantas que se muestrean. El resultado es una visualización por periodo de la cantidad total de los tipos de mazorcas, el cual permite conocer cuales tienen alta probabilidad de ser cosechadas pronto. La información del monitoreo se ilustra en la Figura 4.6 y el gráfico cambia dependiendo de la información de interés, teniendo el área operativa la posibilidad de escoger la información general de la hacienda o conocer esta información por sectores productivos.

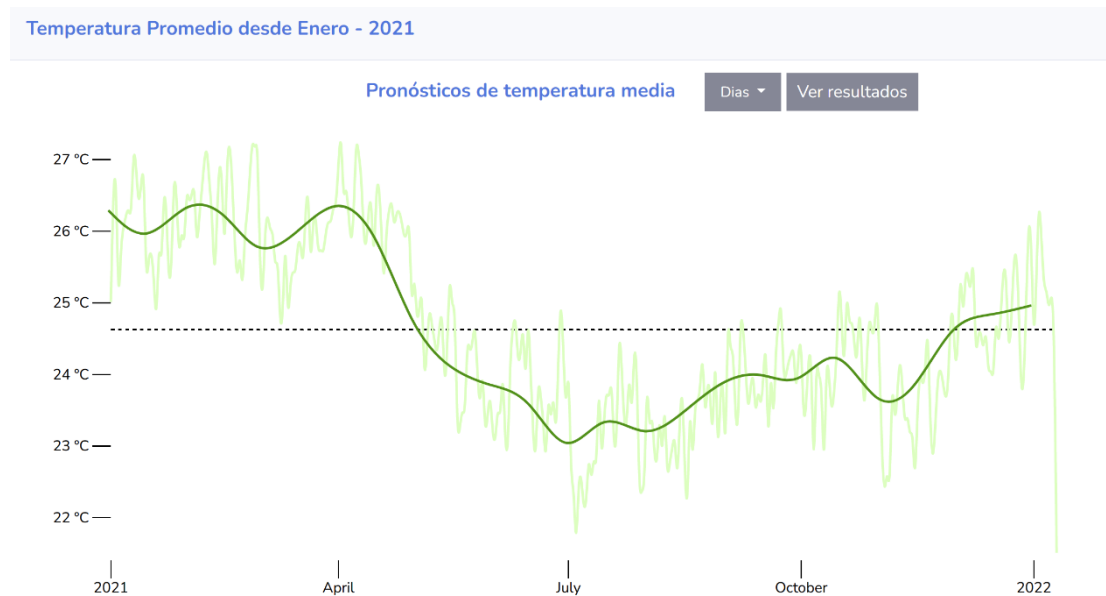


**Figura 4.6 Visualización para el monitoreo de mazorcas periódico**

Esta información ha sido obtenida directamente de la base de datos y ha sido procesada en una vista de Django, por lo que, a medida que se agregue más información será posible visualizar el monitoreo de forma actualizada. De forma similar a las visualizaciones anteriores, se ha validado el resultado utilizando RStudio para generar una tabla con el resumen de las mazorcas de cacao totales por estado y por periodo, con la finalidad de comprobar que la información se muestra correctamente. Los datos almacenados en el sistema, provienen tanto de información histórica, como de datos que son registrados utilizando una tablet y que pueden ser cargados directamente.

### **Información de indicadores climáticos**

La información del clima se muestra en un gráfico de líneas como serie de tiempo de los datos obtenidos directamente de sensores externos. Un menú con



**Figura 4.7 Visualización de temperaturas promedio**

las variables climáticas más utilizadas en las labores operativas de la hacienda ha sido implementado en la interfaz web. Con la finalidad de procesar los datos, se ha utilizado la librería Pandas para resumir la información en formato diario y generar visualizaciones mediante la librería D3.js. Por defecto se consideran únicamente los datos del año en curso desde enero hasta la fecha en la que se realiza la consulta, pero es posible cambiar el rango de visualización dependiendo de las necesidades del usuario. La Figura 4.7 ilustra el tipo de gráfico generado en el sistema, en el que se muestran las temperaturas promedio, tanto como valores diarios, como en tendencia calculada mensualmente.

De la misma forma en cada una de las visualizaciones para indicadores climáticos se genera como resultado, una tabla con información descriptiva (Figura 4.8) en donde se reflejan métricas para cada indicador, tales como:

Indicador	Valor
Temperaturas Promedio	24.8 °C
Temperatura máxima	34.91 °C
Temperatura mínima	16.56 °C
Agosto 2021 res climáticos	+ 0.33 °C

**Figura 4.8 Tabla de análisis de indicadores**

promedio, máximo, mínimo y que la diferencia climática entre el mes actual y el mismo mes del año anterior. Este resultado proporciona una idea directa del comportamiento actual del clima y crea la posibilidad de disponer de información adicional que puede ser utilizada para un análisis más profundo del comportamiento del cultivo y como se relaciona con el clima.

### Pronóstico de temperaturas promedio

El sistema cuenta con una opción que permite pronosticar el comportamiento de la temperatura promedio basado en los datos que son almacenados en el sistema; para esto y como se describió en el Capítulo 3 se utiliza el Modelo Autorregresivo Integrado de Promedio Móvil o ARIMA (por sus siglas en inglés), para series de tiempo, el cual permite realizar un cálculo estimado de las temperaturas promedio en los siguientes 7, 15 y 30 días. El resultado se muestra en la gráfica de la temperatura promedio como se ilustra en la Figura 4.9, pero en color rojo para identificar en que rangos se estima la temperatura en este



**Figura 4.9 Visualización de temperaturas promedio y pronósticos tiempo.**

De la misma forma la temperatura se muestra en una tabla de valores que indica

Fecha	Temperatura estimada
2022-01-11	24.374
2022-01-12	24.761
2022-01-13	25.001
2022-01-14	25.324
2022-01-15	23.557
2022-01-16	24.814
2022-01-17	24.855

**Figura 4.10 Resultado del pronóstico en una tabla en la interfaz web**

el pronóstico para cada uno de los días a partir de la consulta. La Figura 4.10 muestra el resultado que es visualizado en la plataforma.

Este modelo de pronóstico fue desarrollado con Python utilizando las librerías Pandas y Statsmodel con los parámetros 7,0 y 1 basado en el análisis realizado en R y que se describe en el capítulo anterior. Con los datos históricos desde el 1 de enero de 2019 se realizó la evaluación del modelo obteniendo un RMSE de 0.69 al comparar las predicciones con los datos reales obtenidos por el sensor en la estación meteorológica.

## **4.2 Puesta en marcha y funcionamiento**

El desarrollo de la herramienta de visualización y monitoreo ha sido realizado utilizando a Django como framework principal, debido a que trabaja con Python y permite la implementación de funciones para manejo y manipulación de datos con librerías como Pandas y Numpy.

Debido a que Django crea directamente un servidor web local en todo el proceso de desarrollo, todas las funciones implementadas tanto en el lado del procesamiento mediante las vistas de Django, así como los templates o interfaces de usuario han sido verificadas para garantizar su correcto funcionamiento. De forma similar la base de datos ha sido creada utilizando MySQL Server y la interfaz gráfica MySQL Workbench, para la creación de las diferentes tablas y la validación de las diferentes consultas de datos que posteriormente fueron implementados en los modelos y vistas de Django para el manejo de la información. Debido a que las visualizaciones han sido creadas utilizando D3.js y los mapas con Leaflet.js, fue necesario disponer de librerías Javascript, debido a que fue necesario el uso de recursos externos para la generación de gráficos y mapas interactivos.

Desde esta perspectiva, el desarrollo web para el proyecto ha sido realizado de forma local en primera instancia, con la finalidad de garantizar el correcto funcionamiento e interacción entre MySQL y Django para el registro de los datos,

procesamiento y posterior visualización de la información. Esta metodología de trabajo tiene como ventaja evaluar cualquier inconveniente que se presente en el proceso de desarrollo, debido a que se disponen de opciones de depuración para verificar la ejecución del código y analizar los errores existentes.

Una vez concluido el desarrollo de forma local es posible cargar el proyecto a un servicio de alojamiento externo, que permita el acceso a todos los módulos y componentes de la herramienta desde cualquier computador con conexión a internet. Para este propósito es necesario disponer de:

- Un proveedor de servicios de hosting web y almacenamiento para los archivos y la base de datos.
- Un dominio que facilite el acceso a la aplicación web a través de una dirección URL.

Con estas herramientas es posible migrar toda la información y recursos locales al servicio remoto y acceder a los mismos mediante una URL o dirección web del sistema que ha sido contratada. Luego del proceso de migración, todos los usuarios que dispongan de conectividad a internet tendrán acceso al sistema, como herramienta de visualización y monitoreo, lo cual les permitirá registrar las actividades planificadas y visualizar la demanda operativa por tipo de actividad realizada por periodo. De forma similar, será posible ingresar la producción obtenida periodo a periodo y monitorear los avances en la cosecha por área de cultivo, así como el estado de la plantación, mediante un conteo de mazorcas, realizado una semana de cada mes. La información del clima también estará disponible en la plataforma y es obtenida directamente de la estación meteorológica.

### **4.3 Pruebas de funcionalidad**

Con la finalidad de garantizar el funcionamiento apropiado del sistema, se han realizado diferentes pruebas para evaluar el código desarrollado en Django en base a los resultados obtenidos. La interacción de la herramienta con la base de datos se realiza mediante el Mapeo Objeto Relacional (ORM) del Modelo y las

diferentes funciones implementadas en la vista para el respectivo procesamiento de los datos; esto, a su vez, permite disponer en la interfaz de usuario de los datos procesados y representados mediante visualizaciones. Para este propósito fue necesaria la creación de las funciones detalladas en la Tabla 4.2 con el objetivo de procesar la información y generar predicciones de las temperaturas promedio.

**Tabla 4.2 Funciones utilizadas en el procesamiento de los datos**

<b>Función</b>	<b>Objetivo</b>	<b>Entrada</b>	<b>Salida</b>
<b>getDailyData</b>	Proporcionar datos del clima en formato diario	Datos en una frecuencia de 15 minutos y función a utilizar (suma, promedio, máximo, mínimo, etc.)	Datos en formato diario
<b>timeSeriesPrediction</b>	Generar predicciones en base a la serie de tiempo del clima ingresada	Dataset del indicador, cantidad de predicciones o pasos	Dataframe con las predicciones y fechas
<b>formatJsonPod</b>	Convertir el formato de datos del monitoreo o conteo de mazorcas de dataframe a json	Dataframe con los datos obtenidos de la base de datos del monitoreo de mazorcas	Datos en formato JSON requeridos para la visualización
<b>processPeriodicPodCounting</b>	Tomar los datos del monitoreo de mazorcas y resumirlos por periodo	Dataframe con valores de pod counting general o por sector productivo	Datos resumidos por periodo con el total de mazorcas

Estas funciones han sido evaluadas, tanto en las entradas o parámetros, así como en las salidas o valores de retorno; con el propósito de asegurar que el resultado esperado sea el correcto en cada una de las vistas que las requieren.

### **Pruebas de funcionalidad en la generación de visualizaciones**

Uno de los componentes de este proyecto es el módulo de visualizaciones para cada una de las secciones disponibles en la interfaz, por lo cual, en la implementación de cada una de las vistas, se ha evaluado que los datos entregados a los templates para la visualización sean transferidos

correctamente. Este proceso ha sido crucial en la generación de gráficos con Javacript ya que los datos son transferidos en formato json, desde la vista hacia el template para que la librería D3.js pueda utilizarlos apropiadamente.

En cada una de las visualizaciones del sistema se han realizado pruebas de funcionalidad utilizando como recurso inicial el servidor web local de XAMPP y un archivo HTML con código Javascript. De esta forma, y una vez implementado el gráfico deseado, fue transferido a código de Django, en el que la vista procesa los datos y los transfiere al template para la generación apropiada de los gráficos.

### Ingreso de información mediante formularios

Finalmente, el registro de la información, además de ser ingresada como datos históricos en el sistema o archivos provenientes de fuentes externas como csv, es posible ingresarlos directamente en la aplicación web, mediante el uso de formularios, con dos propósitos principales:

- Centralizar la información que es manejada por los encargados de las áreas productivas
- Verificar que los datos ingresados sean válidos, ya que posteriormente serán utilizados para las diferentes visualizaciones.

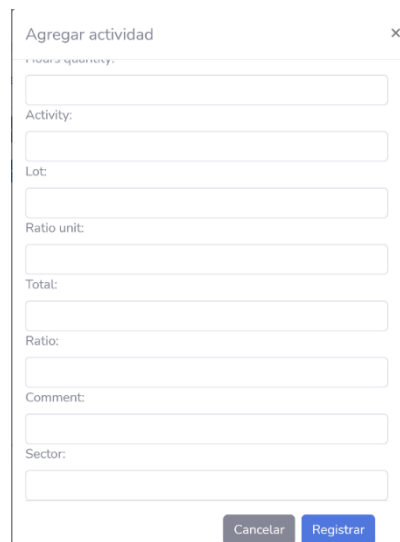
Los formularios han sido implementados mediante el uso de la clase ModelForm del paquete forms de Django, en las Vistas, y con el uso de la librería Bootstrap en las interfaces web del lado del usuario, a través de componentes del tipo Modal, que son mostrarlos al presionar un botón (Figura 4.11).



**Figura 4.11 Gráfico de actividades con botón para agregar datos**



Los formularios han sido implementados en las secciones: Actividades operativas, Producción y Monitoreo de mazorcas. Estos componentes han sido evaluados rigurosamente, con el propósito de verificar que la información que es enviada desde la interfaz web (template) llega correctamente a la base de datos, mediante el Modelo de Django y utilizando el método POST.

The image shows a web form titled "Agregar actividad" with a close button (x) in the top right corner. The form contains several input fields: a large text area at the top, followed by "Actividad:", "Lot:", "Ratio unit:", "Total:", "Ratio:", "Comment:", and "Sector:". At the bottom of the form, there are two buttons: "Cancelar" and "Registrar".

**Figura 4.12** Formulario de registro de actividades operativas

#### **4.4 Análisis costo/beneficio**

Uno de los beneficios principales de la implementación de un sistema que permita monitorear procesos agrícolas en cacao, es la disponibilidad de una herramienta desarrollada específicamente para este propósito, en base al análisis de los requerimientos de la empresa. Esto se relaciona directamente con la transformación digital en sectores productivos en donde existen tecnologías emergentes, y potenciales oportunidades de implementación de soluciones de software para el manejo de cultivos. El sistema desarrollado combina aspectos de diferentes sistemas descritos a continuación:

**Sistemas de manejo agrícola:** Debido a que permite ingresar y revisar la demanda operativa del personal para las diferentes actividades agrícolas y como se distribuye a lo largo del año. También es posible revisar el avance en la producción obtenida al momento, así como generar una representación gráfica del monitoreo de mazorcas para estimar la próxima cosecha.

**Sistemas de información geográfica:** Ya que la información de producción además de ser representada en un gráfico, se visualiza en un mapa interactivo, que indica las áreas de producción y como esta se distribuye en toda la hacienda.

**Sistemas de visualización de datos:** Debido a que las diferentes visualizaciones escogidas para ser implementadas en esta solución tienen relación a conceptos de visualización de datos, representando de forma efectiva la información disponible.

**Sistemas agrometeorológicos:** Uno de los componentes esenciales en esta herramienta es el clima como factor determinante en la agricultura, para esto se han desarrollado visualizaciones de los diferentes indicadores climáticos principales, y sobre todo de metodologías para trabajar directamente con datos obtenidos en la zona de producción, además de disponer de un módulo de pronóstico de temperaturas promedio.

Por este motivo, la implementación de una solución de software para haciendas de producción de cacao aportaría a la modernización y automatización de procesos que actualmente son realizados manualmente, y demandan recursos operativos, tal y como se describe en el Capítulo 1. Al momento no se dispone en la empresa de una solución de este tipo, y trabajar con archivos independientes u hojas de cálculo para manejar y analizar los datos, supone un riesgo relacionado directamente con la calidad de la información.

### **Optimización del tiempo**

La automatización de procesos implementados en esta herramienta mediante las diferentes visualizaciones e información que se genera periódicamente contribuye a la optimización del tiempo que actualmente es requerido para estas tareas. Las principales actividades que buscan ser suprimidas al utilizar esta solución y que actualmente son realizadas manualmente están relacionadas a la

generación de reportes con las actividades operativas y a la generación de visualizaciones relacionadas al conteo de mazorcas, producción y clima.

Debido a esto se realizó un análisis del tiempo requerido para realizar estas tareas con una frecuencia semanal, utilizando los programas MS. Excel, RStudio y QGIS; dependiendo de la funcionalidad. El resultado fue comparado con el tiempo de ejecución de la aplicación en una consulta web, para cada módulo.

**Tabla 4.3 Análisis de tiempo requerido para cada funcionalidad del sistema**

Módulo	Funcionalidad	Software utilizado	Tiempo (min)	Proyecto	Tiempo (min)
<b>Actividades operativas</b>	Análisis de demanda por tipo	Excel	20	Automático	0.043
	Visualización por periodo	Excel	15	Automático	0.043
	Visualización en mapas	QGIS	45	Automático	5
<b>Productividad</b>	Visualización por periodo	Excel	15	Automático	0.057
	Visualización en mapas	QGIS	45	Automático	5
<b>Conteo de frutos</b>	Análisis por periodo	Excel	3	Automático	0.113
	Visualización del conteo	Excel	3.5	Automático	0.113
	Análisis de estadio de mazorca	Excel	3	Automático	0.113
<b>Condiciones climáticas</b>	Visualización de temperaturas	RStudio	7.5	Automático	0.233
	Análisis de temperatura	RStudio	9.5	Automático	0.233
	Visualización de humedad relativa	RStudio	7.5	Automático	0.267
	Análisis de humedad relativa	RStudio	9.5	Automático	0.267
	Visualización de radiación solar	RStudio	7.5	Automático	0.267
	Análisis de radiación solar	RStudio	9.5	Automático	0.267
	Visualización de precipitación	RStudio	7.5	Automático	0.267
	Análisis de precipitación	RStudio	3	Automático	0.267
	Visualización de evapotranspiración	RStudio	7.5	Automático	0.267
	Análisis de evapotranspiración	RStudio	9.5	Automático	0.267
	Visualización de grados días	RStudio	8	Automático	0.267
	Análisis de grados días	RStudio	9.5	Automático	0.267
	Pronóstico de temperaturas promedio	RStudio	-	Automático	2
	Visualización de pronósticos de temperaturas promedio	-	-	Automático	2
	Descarga de datos del clima	-	-	Automático	3.5
<b>Total</b>			<b>245.5</b>		<b>17.7</b>

De acuerdo con la Tabla 4.3 se evidencia que, con la implementación de un dashboard de visualizaciones automatizado y personalizado con los datos de la

empresa, es posible realizar todas las funcionalidades desarrolladas en **17.7** minutos, comparado con los **245.5** minutos que actualmente se requieren. o **3.8** horas semanales. Esto significa que aproximadamente se ahorraría **9.5%** del tiempo semanal para desempeñar este trabajo o **3.8 horas** (227.8 min), en base a las 40 horas laborales.

En vista de que el proceso en la obtención de resultados involucra más de una persona, este número podría aumentar, además de que no se considera el tiempo requerido para transcribir datos de una fuente como papel al computador, que podría ser optimizado utilizando directamente la herramienta. De la misma forma y como resultado, se ha analizado las personas que estarían interesadas en revisar la información periódicamente en el sistema y que actualmente no disponen de un medio para este propósito. A continuación, se presenta la cantidad de personas estimadas que monitorearán los datos en la herramienta:

**Tabla 4.4 Personal por áreas que utilizarían la herramienta**

<b>Área o Departamento</b>	<b>Personas</b>
<b>Operaciones</b>	5
<b>Investigación</b>	13
<b>Calidad</b>	2
<b>Equipo externo que monitorea clima en LATAM</b>	2

Siendo un total estimado de **22 personas** en la organización que tendrían acceso a esta herramienta de forma instantánea.

La herramienta no solo posibilitará la generación de gráficos automáticos e información relevante, sino que permitirá descargar archivos csv; esta característica es importante ya que en muchas ocasiones se requieren los resultados para ser detallados en presentaciones.

### **Costos**

El costo de disponer de la herramienta siempre operativa en internet está relacionado al costo del servicio de hosting web contratado para alojar los recursos que han sido desarrollados. En el mercado existen diferentes

alternativas para servicios de hosting web y cuyo costo variará dependiendo de diferentes características disponibles, entre ellas, la cantidad de almacenamiento y los sitios web que se alojen; los valores también oscilarán dependiendo de la cantidad de meses del servicio contratado, siendo más económico contratarlo a largo plazo.

Tomando como referencia los costos de dos de los proveedores de hosting web más conocidos y un plan básico con almacenamiento de hasta 100 GB, con el sistema operativo Linux para instalar Django y 1 sitio web se tienen los siguientes valores:

**Tabla 4.5 Valores de 2 proveedores de servicios de hosting web**

<b>Meses</b>	<b>Proveedor 1</b>	<b>Proveedor 2</b>
<b>12</b>	60	83.99
<b>24</b>	102	155.76
<b>36</b>	126	215.64

El costo por el dominio o URL también debe ser considerado, en el presente proyecto se eligió la dirección [www.cocoafarmapp.com](http://www.cocoafarmapp.com) y su valor fue de \$12 anual. Asumiendo una operatividad del sistema de 2 años los costos están entre \$102 y \$155.76 (tabla 4.5) por 24 meses, con un costo de \$24 por el dominio.

# CONCLUSIONES Y RECOMENDACIONES

## Conclusiones

1. De acuerdo con los resultados obtenidos y el análisis costo-beneficio realizado, es posible concluir que una solución basada en datos permite optimizar procesos que, al momento son realizados manualmente o mediante el uso de sistemas utilitarios. En este sentido, es evidente que un sistema personalizado, reducirá la necesidad de contar con diferentes soluciones para realizar diferentes tareas o análisis de datos de forma aislada. Además, de esta forma, se genera la oportunidad de integrar diferentes tipos de sistemas en uno solo, que fueron diseñados con un propósito común.
2. El desarrollo de una aplicación web para el manejo y visualización de datos, en haciendas de producción de cacao, puede contar con varios módulos, que pueden funcionar de forma integral, con la finalidad de registrar datos, procesarlos y obtener información relevante, y que asista al departamento encargado del manejo y administración de recursos y personal. En este proyecto se consideraron únicamente cuatro módulos, sin embargo, módulos adicionales enfocados en aspectos agronómicos del cultivo pueden ser agregados progresivamente.
3. Una herramienta o software de visualización puede mejorar la forma en la que se analizan los datos en una empresa, esto será posible siempre y cuando los gráficos sean los más apropiados dependiendo del tipo de datos y se disponga de una apropiada interacción en los mismos, ya que, con esto, el usuario tendrá un mayor control en cómo se presentan los resultados obtenidos.
4. Es posible aplicar ciencia de datos en casi cualquier sector o disciplina, con la finalidad de contribuir con nuevas tecnologías que procesen los datos y generen

resultados, en este sentido, el sector agrícola es un área que presenta oportunidades para implementar nuevas metodologías. En el desarrollo de este proyecto se han considerado diferentes aspectos tales como la visualización de datos y el análisis y pronósticos de series temporales.

## **Recomendaciones**

1. Es recomendable realizar un análisis previo del entorno de trabajo y los usuarios finales, con el propósito de seleccionar las tecnologías más apropiadas para el desarrollo de una solución basada en datos, debido a que, por ejemplo, en muchos lugares aún no se dispone de conexión a internet, en este caso, será mejor implementar una aplicación móvil o una aplicación de escritorio que no requiera conectividad.
2. Un levantamiento de requerimientos con todos los usuarios finales es esencial, ya sea si se desarrolla una tecnología nueva, o si se está implementando una solución que reemplazará a otra, ya que de esta forma será posible conocer todos los aspectos que harán posible una buena experiencia de usuario.
3. En proyectos desarrollados específicamente para el sector agrícola, es necesario considerar los aspectos más importantes en el desarrollo biológico de un cultivo, con la finalidad de priorizar las etapas de implementación y alinear las funcionalidades del software con las diferentes prácticas agronómicas. De forma similar, se recomienda consultar y trabajar en conjunto con un especialista en agronomía durante todas las etapas de implementación.
4. Para el desarrollo de una herramienta orientada a automatizar procesos, en un determinado sector productivo o industria, se recomienda disponer de los datos necesarios para la implementación de los diferentes módulos. De no disponer de algún dataset, es importante realizar una planificación previa con la finalidad de coleccionar la información requerida antes iniciar el desarrollo.

5. Es importante, durante el proceso de implementación de este tipo de soluciones, contar con equipos de trabajo vinculados a la empresa y que estén familiarizados con todas las actividades que se busca desarrollar. De forma similar, es recomendable realizar la transición progresiva para su correcto uso. De ser necesaria una capacitación previa, se sugiere hacerlo con todo el equipo que utilizara el software.

## BIBLIOGRAFÍA

- Akpanta, A., Okorie, I., & Okoye, N. (2015). SARIMA modelling of the frequency of monthly rainfall in Umuahia, Abia State of Nigeria. *American Journal of Mathematics and Statistics*, 5(2), 82-87.
- ANECACAO. (Junio de 2021). *Asociación Nacional de Exportadores de Cacao*. Obtenido de Estadísticas de Exportación: <http://www.anecacao.com/index.php/es/estadisticas/estadisticas-actuales.html>
- Balyani, Y., Fazelniya, G., & Bayat, A. (2012). A study and prediction of annual temperature in shiraz using ARIMA model. *GEOGRAPHIC SPACE*, 12(38), 127-144. Obtenido de <https://www.sid.ir/en/journal/ViewPaper.aspx?id=269692>
- Brath, A., Montanari, A., & Toth, E. (1999). Short-term rainfall prediction with time series analysis techniques for real-time flash flood forecasting. *Mediterranean storms, Proceedings EGS Plinius Conference, Maratea (PZ)*.
- Carrer, M., de Souza Filho, H., & Batalha, M. (2017). Factors influencing the adoption of Farm Management Information Systems (FMIS) by Brazilian citrus farmers. *Computers and Electronics in Agriculture*, 138, 11-19.
- De Arce, R., & Mahía, R. (2003). Modelos Arima. *Programa CITUS: Técnicas de Variables Financieras*.
- Dzemyda, G., Kurasova, O., Medvedev, V., & Dzemydaitė, G. (2019). Visualization of data: methods, software, and applications. En *Advances in Mathematical Methods and High Performance Computing* (Vol. 41, págs. 295-307). Springer, Cham.
- Johnson, C., Schafer, R., & Young, S. (1983). Controlling agricultural machinery intelligently. In: *Agricultural Electronics-1983 and Beyond. National Conference*



- on Agricultural Electronics Applications* (págs. 114 - 119). St Joseph, MI, USA: American Society of Agricultural Engineers.
- Katz, R., & Skaggs, R. (1981). On the use of autoregressive-moving average processes to model meteorological time series. *Monthly Weather Review*, 109(3), 479-484.
- Khedhiri, S. (2016). Forecasting temperature records in PEI, Canada. *Letters in Spatial and Resource Sciences*, 9(1), 43-55.
- León Villamar, F., Calderón Salazar, J., & Mayorga Quinteros, E. (2016). Estrategias para el cultivo, comercialización y exportación del cacao fino de aroma en Ecuador. *Ciencia Unemi*, 9(18), 45-49.
- Liu, X., Wang, X., Wright, G., Cheng, J., Li, X., & Liu, R. (2017). A state-of-the-art review on the integration of Building Information Modeling (BIM) and Geographic Information System (GIS). *ISPRS International Journal of Geo-Information*, 6(2), 53.
- Macías Barberán, R., Cuenca Nevárez, G., Intriago Flor, F., María Caetano, C., Menjivar Flores, J. C., & Pacheco Gil, H. A. (2019). Vulnerability to climate change of smallholder cocoa producers in the province of Manabí, Ecuador. *Revista Facultad Nacional de Agronomía Medellín*, 72(1), 8707-8716.
- Mehrmolaei, S., & Keyvanpour, M. (2016). Time series forecasting using improved ARIMA. *2016 Artificial Intelligence and Robotics (IRANOPEN)* (págs. 92-97). IEEE.
- Murat, M., Malinowska, I., Gos, M., & Krzyszczak, J. (2018). Forecasting daily meteorological time series using ARIMA and regression models. *International agrophysics*, 32(2), 253-264.
- Nikkilä, R., Seilonen, I., & Koskinen, K. (2010). Software architecture for farm management information systems in precision agriculture. *Computers and electronics in agriculture*, 70(2), 328-336.
- Paras, S., Kumar, A., & Chandra, M. (2009). A feature based neural network model for weather forecasting. *International Journal of Computational Intelligence*, 4(3), 209-216.
- Pechter, V. (2021). *What Is Farm Management Software and Who Is It For?* Recuperado el 28 de Agosto de 2021, de eAgronom: [https://eagronom.com/en\\_au/blog/what-is-farm-management-software/](https://eagronom.com/en_au/blog/what-is-farm-management-software/)

- Pettit, C., Bishop, I., Cartwright, W., Park, G., & Kemp, O. (2007). Enhancing web based farm management software through the use of visualisation technologies. *MODSIM 2007 International Congress on Modeling and Simulation. Modeling and Simulation Society of Australia and New Zealand*, (págs. 1280-1286).
- Power BI. (29 de Agosto de 2021). Obtenido de Microsoft Power BI: <https://powerbi.microsoft.com/es-es/what-is-power-bi/>
- Sadiku, M., Shadare, A., Musa, S., Akujuobi, C., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12), 11-16.
- Schmidt, J. (2020). Usage of Visualization Techniques in Data Science Workflows. *15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP '20) (3: IVAPP)* , (págs. 309-316).
- Sigrimis, N., Hashimoto, Y., Munack, A., & De Baerdemaeker, J. (1999). Prospects in Agricultural Engineering in the Information Age: Technological Developments for the Producer and the Consumer. *CIGR e-journal*. Obtenido de <http://www.agen.tamu.edu/cigr>
- Stafford, J. (2000). Implementing precision agriculture in the 21st century. *Journal of agricultural engineering research*, 76(3), 267-275.
- Tableau. (2021). Recuperado el 29 de Agosto de 2021, de Tableau: <https://www.tableau.com/why-tableau/what-is-tableau>
- Top 9 Farm Management Software in 2021. (2021). Recuperado el 29 de Agosto de 2021, de PAT RESEARCH: <https://www.predictiveanalyticstoday.com/top-farm-management-software/>
- Xiao, J., Liu, H., Luan, X., & Zhou, Z. (2007). The design of a map database. *Geoinformatics 2007: Cartographic Theory and Models*, 6751, 675114.