



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Implementación de herramientas Python en el proceso de
producción de cultivos agrícolas del fundo “San Juan de Buenavista”

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas

AUTOR:

Mendoza Vásquez, Eduardo Arturo (ORCID: 0000-0001-9300-2838)

ASESOR:

Vargas Vargas, Gautama (ORCID: 0000-0001-7181-7497)

LÍNEA DE INVESTIGACIÓN:

Sistemas de información

LIMA – PERÚ

2021

Dedicatoria

Sin el apoyo de mi familia no hubiera recibido la formación que me ha permitido presentar la investigación con el fin de obtener mi grado de Ingeniero. Al docente por su apoyo brindado en la investigación, por haberme transmitido los conocimientos obtenidos y haberme llevado pasó a paso en el aprendizaje.

Agradecimiento

Agradezco al dueño del fundo San Juan de Buenavista, quien es mi abuelo, por la confianza de poder innovar con soluciones tecnológicas dentro del negocio. También agradezco a mis familiares quienes me apoyaron facilitándome el conocimiento de cada proceso.

Índice de contenidos

Dedicatoria	ii
Agradecimiento.....	ii
Índice de contenidos.....	ii
Índice de tablas.....	iv
Índice de figuras	v
Resumen	vi
Abstract	vii
I. INTRODUCCIÓN	1
1.1. Realidad Problemática	2
1.2. Justificación de la investigación	2
1.3. Objetivo general.....	3
1.4. Hipótesis	3
II. MARCO TEÓRICO.....	4
2.1. Antecedentes	4
2.2. Marco Teórico	7
2.3. Marco Conceptual	8
III. METODOLOGÍA.....	13
3.1. Diseño de investigación	13
3.2. Variables y operacionalización	13
3.3. Población, muestra y muestreo	14
3.4. Técnicas e instrumentos de recolección de datos	14
3.5. Procedimientos	15
3.6. Método de análisis de datos.....	16
3.7. Aspectos éticos.....	17
IV. RESULTADOS	18
4.1. Análisis descriptivo.....	18
4.2. Análisis de Inferencial	25
V. DISCUSIÓN	27
VI. CONCLUSIONES.....	30
VII. RECOMENDACIONES.....	31
REFERENCIAS	32
ANEXOS.....	35
Anexo 01: Flujo de procesos.....	35
Anexo 02 – Matriz de operacionalización de variables.....	36
Anexo 03 – Matriz de operacionalización de indicadores.....	37

Anexo 04 – Diario de trabajo del proceso Logístico	38
Anexo 05 – Diario de trabajo del proceso de producción	39
Anexo 06 – Diagrama de constelación.....	40
Anexo 07 – Declaración de originalidad del Autor/Autores.....	41

Índice de tablas

Tabla 01: Técnicas e instrumentos de recolección de datos	15
Tabla 02: Técnicas, instrumentos e indicadores	16
Tabla 03: Fechas de recolección de datos.....	18
Tabla 04: Modelos de normalidad	18
Tabla 05: Tabla estadística de la muestra.....	21
Tabla 06: Tabla estadística de la muestra.....	22
Tabla 07: Tabla estadística de la muestra.....	23
Tabla 08: Tabla estadística de la muestra.....	24
Tabla 09: test normalidad Modelo 1	26
Tabla 10: test normalidad Modelo 2.....	26
Tabla 11: test normalidad Modelo 3.....	26
Tabla 12: test normalidad Modelo 4.....	26

Índice de figuras

Figura 01: Proceso de elaboración de la proyección.....	7
Figura 02: Proceso de elaboración de la proyección.....	8
Figura 03: Fases de la metodología CRISP-DM	8
Figura 04: Cuadrante Gartner Análisis predictivo.....	9
Figura 05: Fases de la metodología CRISP-DM	11
Figura 06: Esquema constelación	12
Figura 07: formula hipótesis.....	13
Figura 08: Diagnostico de residuos – Modelo 1	19
Figura 09: Diagnostico de residuos – Modelo 2	19
Figura 10: Diagnostico de residuos – Modelo 3	20
Figura 11: Diagnostico de residuos – Modelo 4	21
Figura 12: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “precisión de materiales” (pretest y postest).....	22
Figura 13: Porcentaje de precisión de materiales	22
Figura 14: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “perdida de materiales” (pretest y postest).....	23
Figura 15: Porcentaje de perdida de materiales.....	23
Figura 16: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “precisión de producción” (pretest y postest).....	24
Figura 17: Porcentaje de precisión de producción	24
Figura 18: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “precisión de descarte” (pretest y postest)	25
Figura 19: Porcentaje de precisión de descarte	25

Resumen

En el presente desarrollo de tesis detalla el estudio del proceso de producción agrícola en el fundo San Juan de Buenavista, así mismo se presenta el desarrollo y la implementación de la herramienta Python para la realización de modelos predictivos. El proceso de producción se controla por fichas de monitoreo de forma manual que se ingresa por el jefe de operaciones a un archivo Excel. Este proceso ocasiona lentitud en la toma de decisiones debido a que el jefe de operaciones puede ocupar mas de 2 días en preparar y realizar proyecciones, las cuales tienen menos del 80% de precisión. El objetivo del estudio es diseñar e implementar la herramienta Python, para generar alertas y brindar soporte en la toma de decisiones del área de planta, el cual se toma en cuenta los objetivos específicos para construcción de la herramienta.

La herramienta es de software libre y se realizó con la metodología CRISP-DM, la cual ayudara a diseñar e implementar los modelos predictivos. La investigación es de tipo aplicada, y como diseño de estudio es experimental. Los indicadores a tomar en cuenta son la precisión en predecir la producción, el uso de materiales, la producción de descarte y la pérdida de materiales. Se utilizan dos segmentos de datos, datos de entrenamiento (train) que utilizan datos hasta el 2019 y para el segundo segmento denominado datos de prueba (test) se utilizara datos del 2020. Se utilizaron las pruebas Shapiro-Wilk test y D'Agostino's K-squared.

El resultado de los modelos predictivos fue superior al 80% con el cual se considera un modelo óptimo. Gracias a la implantación de la herramienta se logró aumentar todos los indicadores de producción y gestión de materiales. Con esto se demostró que la herramienta mejoró la forma de hacer proyecciones en la producción agrícola de la campaña 2020 en el fundo San Juan de Buenavista.

Palabras clave: modelo predictivo, proceso de producción agrícola, Python

Abstract

In this thesis development, the study of the agricultural production process in the San Juan de Buenavista farm is detailed, as well as the development and implementation of the Python tool for the realization of predictive models. The production process is controlled by manual monitoring files that are entered by the operations manager into an Excel file. This process causes slow decision-making because the operations manager can spend more than 2 days preparing and making projections, which are less than 80% accurate. The objective of the study is to design and implement the Python tool, to generate alerts and provide support in the decision-making of the plant area, which takes into account the specific objectives for the construction of the tool.

The tool is free software and was made with the CRISP-DM methodology, which will help to design and implement predictive models. The research is of an applied type, and as a study design it is experimental. The indicators to take into account are the precision in predicting the production, the use of materials, the production of waste and the loss of materials. Two data segments are used, training data (train) that use data up to 2019 and for the second segment called test data (test), data from 2020 will be used. The Shapiro-Wilk test and D'Agostino's K-squared tests were used.

The result of the predictive models was higher than 80% with which it is considered an optimal model. Thanks to the implementation of the tool, it was possible to increase all the production and material management indicators. With this, it was shown that the tool improved the way of making projections in the agricultural production of the 2020 campaign in the San Juan de Buenavista farm.

Keywords: predictive model, agricultural production process, Python

I. INTRODUCCIÓN

En la actualidad una de las principales razones para que las empresas puedan aumentar su productividad, es el análisis predictivo. La mayoría de empresas del País solo llegan a realizar análisis descriptivos, lo que comúnmente se presentan en grafico donde se muestran la situación actual de la empresa (Indicadores de gestión, alcance de metas, bonificaciones, ranking, etcétera). En el mercado actual que cada día se vuelve más competitivo, las empresas no solo desean saber cómo van, sino como les ira a futuro según su avance a la fecha y las estrategias a aplicar. Es donde nace el análisis predictivo, y para poder llegar a realizarlo, se tienen que conocer los procesos internos, tener centralizados los datos y una base de datos histórica de mínimo 1 año. En el fundo “San Juan de Buenavista”, empresa que cuenta con RUC de persona natural con negocio, cada inicio de campaña se realiza una reunión para administrar los recursos (insumos, planilla, ventas), la cual tiene como finalidad saber la proyección mes a mes, información que toma 3 semanas en realizarse con menos del 80% de precisión. Gran parte de este problema se debe a que cuenta con información descentralizada, déficit en el tratamiento de los datos y ausencia de análisis de información. A pesar de tener al jefe de operaciones realizando seguimiento en los cultivos, una vez que los productos llegan a planta para ser procesados, los datos se dejan gestionar. Como se ve el escenario, fue una clara oportunidad para la implementación de un modelo predictivo.

1.1. Realidad Problemática

En el fundo “San Juan de Buenavista” se detectó déficit en la gestión de datos. Los datos se encontraban descentralizados debido a que todas las fichas de monitoreo eran registradas de forma manual en hojas físicas, para luego ser ingresadas en un archivo Excel. Mientras que los datos del tareo diario son registrados por dispositivos móviles y almacenados en una Base de datos independiente.

Debido a que los datos no están centralizados, cuando el jefe de operaciones requiere generar un reporte existente o crear uno nuevo, el tiempo de elaboración no es óptimo y ocasiona que se cruce con tareas diarias de superior prioridad, como resultado de ello el tiempo se extiende. Para las proyecciones realizadas por campaña, se realizan por cálculos manuales mediante fórmulas en archivos Excel, debido a ello se obtuvieron resultados con menos de un 80% de precisión.

En este contexto se planteó la siguiente problemática: ¿Cómo puede mejorar la implementación de la herramienta Python en la gestión de datos y la precisión en la proyección del proceso de producción agrícola para poder brindar soporte en la toma de decisiones?

1.2. Justificación de la investigación

La justificación general de la investigación es brindar mejorar el análisis predictivo con el fin de ayudar a predecir las utilidades en base al nivel de producción agrícola.

La herramienta Python a implementar, tiene como finalidad brindar soporte a la toma de decisiones. La implementación requiere 5 etapas, cada una de ellas soluciona las deficiencias con las que cuenta la empresa en la actualidad. Como es la descentralización de datos, la cual ocasiona demora en creación de los informes y reportes que se presentan en las reuniones de inicio y cierre de campaña. El tratamiento de datos y el análisis de información para la elaboración de las proyecciones por cada rango de tiempo seleccionado, al contar con menos del 80% de precisión, ocasiona que no se designe de forma óptima los recursos a las áreas correspondientes.

En el área de Planta se genera la información más relevante del negocio, debido a que se procesan la materia prima y se utiliza la mayor cantidad de recursos. Cuando se plantea las proyecciones, se destina una cantidad de insumos aproximados, pero al tener una precisión muy baja pueden ocurrir dos situaciones. La primera, se destina una cantidad excesiva de insumos y personal de trabajo para la producción que termina siendo mínima y los recursos se pierden. Caso contrario cuando se destina pocos recursos se generan retrasos debido a que se tiene que solicitar más recursos para

culminar la producción. Los gastos en los últimos 3 años crecieron 10% el primer año y 13% el segundo año.

La parte tecnológica del negocio también tiene a aprovechar, debido a que la empresa en el año anterior adquirió un servidor para salvaguardar su información en una Base de datos. De esta manera se aprovecharán los recursos y evitara gastos adicionales.

1.3. Objetivo general

Diseñar e implementar la herramienta Python, para brindar soporte en la toma de decisiones del área de planta.

1.3.1. Objetivos específicos

Dentro del objetivo general se ramifican los siguientes objetivos específicos:

- Diseñar Datamart para el área de Planta
- Diseñar procesos de ETL para alimentar la información histórica
- Implementar la herramienta Python en los servidores del are TI
- Diseñar e implementar modelo predictivo en la herramienta Python 8

1.4. Hipótesis

- Hipótesis alterna: La implementación de la herramienta Python podrá brindar soporte en la toma de decisiones del área de planta, con un nivel de precisión superior al 80%.
- Hipótesis Nula: La implementación de la herramienta Python podrá brindar soporte en la toma de decisiones del área de planta, con un nivel de precisión menor al 80%.

II. MARCO TEORICO

2.1. Antecedentes

2.1.1. Antecedentes nacionales

Edwin Williams Heredia Salinas (2019, p. 11), la “Inteligencia de negocios aplicada a la gestión de ventas de una empresa agroindustrial” Tesis (Título de Ingeniero en sistemas e informática) Lima, Universidad Cesar Vallejo. El objetivo principal que tuvo el proyecto en mención fue de poder aplicar inteligencia de negocios en la gestión de ventas para una empresa agroindustrial. Su investigación fue de tipo explicativo y según los resultados obtenidos en el análisis inferencial al contraste de Hipótesis, se logró determinar que aplicar Inteligencia de negocios influye positivamente en la gestión de ventas con un nivel de confianza del 97%.

Montero Silva Boris Javier (2018, p. 23), la “Implementación de una solución de inteligencia de negocios para una empresa agroindustrial” Tesis (Título de Ingeniero en sistemas e informática) Piura, Universidad nacional de Piura. El proyecto en mención tiene como su objetivo principal Implementar una solución de inteligencia de negocios en el acopio de productos agroindustriales. Su investigación es de tipo Descriptiva y el diseño de investigación es no Experimental – transaccional. Para lograr su objetivo se tomó en cuenta determinar los requerimientos en el acopio de productos, evaluar la información de la base de datos obtenida del sistema transaccional e implementar el Data Warehouse, esta debe proporcionar y describir el análisis de la información estratégica obtenida a partir del modelo dimensional realizado. Para realizar el proyecto se eligió el Metodología Kimball tomando como modelo a aplicar el Copo de Nieve.

Barboza Morales Italo Shamir, Huamaní Portocarrero Javier Eduardo (2016), presentó la tesis “Implementación de un modelo de business intelligence orientado a tecnología mobile basado en sap businessobjects para pymes del sector retail”, Universidad Peruana de Ciencias aplicadas, Perú. El objetivo de esta investigación fue Implementar un modelo de Business Intelligence orientado a tecnología Mobile utilizando SAP BusinessObjects que permita la mejora de toma de decisiones empresariales en pymes del sector retail. Su investigación fue de tipo explicativo.

Nixon Omar Fernández Carrión (2018) presentó la tesis “La influencia de la inteligencia de negocios en el análisis de información de ventas de la importadora y distribuidora Jiménez E.I.R.L, en la ciudad de Nueva Cajamarca, 2017”, Universidad nacional de san martín-Tarapoto, Perú. El objetivo de esta investigación fue Aplicar inteligencia de negocios para mejorar el análisis de

información de ventas en la Importadora y Distribuidora Jiménez E.I.R.L, en la ciudad de Nueva Cajamarca, 2017. Su investigación fue de tipo aplicativo.

Marlene Elisa Carhuaricra Inocente, Jenny Isabel Gonzales Caporal (2017) presentó en la tesis “Implementación de business intelligence para mejorar la eficiencia de la toma de decisiones en la gestión de proyectos”, Universidad San Ignacio de Loyola, Perú. El objetivo de esta investigación fue Demostrar la influencia del uso de BI en la eficiencia de la toma de decisiones en la Gestión de los Proyectos. Su investigación fue de tipo experimental.

Lorena Beatriz Bellido Camacho (2019) presentó la tesis “Propuesta de mejora en la toma de decisiones aplicando business intelligence caso: agencia de promoción de inversiones”, Universidad San Ignacio de Loyola, Perú. El objetivo de esta investigación fue BI es la herramienta que mejorará la toma de decisiones en campañas de promoción en aporte a la gestión de clientes. Su investigación fue de tipo experimental.

2.1.2. Antecedentes Internacionales

Candanoza Rey Carlos Andrés, Pacheco Casiani Henuar Javier, Puerta Botero Paula Andrea, Murillo Sandra Marcela, Villamil Valencia Fredy Alexander (2019, p. 17), la “Propuesta de implementación de herramienta de Business Intelligence (BI) orientada en la reducción de costos de adquisición de materia prima en una empresa procesadora y comercializadora de alimentos en la región del eje cafetero”, Tesis (Título de Ingeniero en sistemas e informática) Colombia, Institución Universitaria Politécnico grancolombiano. El proyecto en mención tiene como objetivo tiene proponer el diseño e implementación de un sistema de información (BI) a través de una herramienta tecnológica para obtener ventaja competitiva en el proceso de adquisición de materias primas, permitiendo integrar y aumentar la eficiencia en los procesos; y apoyar la toma de decisiones. La metodología para lograr el objetivo interrelaciona el esquema PMI involucrando las áreas de conocimiento tales como integración, alcance, tiempo, costes, calidad, recursos humanos, comunicación, riesgo, adquisiciones y partes interesadas en los procesos de inicio, planeación, ejecución, monitorización y control, y cierre. Para la construcción de la solución se tomó en cuenta la metodología Kimball con el esquema estrella.

Espinoza Vega Angel Andrés, Jiménez Ruiz Fernanda Elizabeth (2016, p. 21), la “Implementación de un módulo de soporte de decisiones para agricultura utilizando Big Data e inteligencia de negocios integrado al sistema iotmach”. Tesis (Título

Ingeniero informático) Colombia, Universidad técnica de Machala. El proyecto en mención tiene como objetivo principal Implementar un módulo de soporte de decisiones para agricultura utilizando Big Data e Inteligencia de Negocios BI aplicando la metodología Hefesto, para la obtención de análisis descriptivos de los cultivos y cosechas, integrado al sistema IoT Mach. El problema que se pretende resolver, es la generación de información estadística descriptiva que ayude en la toma de decisiones a través de un sistema de soporte de decisiones (DSS) y un Sistema de Información Ejecutiva.

Arrobo Lapo Estalin Vladimir (2017), presentó en la tesis “BUSINESS Intelligence para la generación de indicadores y control de desempeño, en el área de compras agrícolas de la empresa Pronaca”, Universidad regional autónoma de los Andes, Colombia. El objetivo de esta investigación fue Implementar una solución Business Intelligence para la generación de indicadores y control de desempeño, en el área de compras agrícolas de la empresa Pronaca. Su investigación fue de tipo aplicativo.

Santiago Leonardo Morales Cardoso (2019), presentó en la tesis “Metodología para procesos de Inteligencia de negocios con mejores en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones”, Universidad de Alicante, España. El objetivo de esta investigación fue Conseguir que se disminuyan los tiempos de desarrollo de los proyectos de IN, por medio de una metodología orientada a la toma de decisiones, mejorando principalmente los procesos de extracción y transformación de fuentes de datos heterogéneos con la ayuda de árboles de decisión que aplican aprendizaje automático en las decisiones. Su investigación fue de tipo aplicativo.

Pablo Esteban Villota Neira (2019), presentó en la tesis “Implementación de una estación prototipo con visión artificial, aplicado a la agricultura de precisión”, Universidad de Cuenca, Ecuador. El objetivo de esta investigación fue Diseñar y desarrollar una estación prototipo que use visión artificial enfocada a la agricultura de precisión. Su investigación fue de tipo aplicativo.

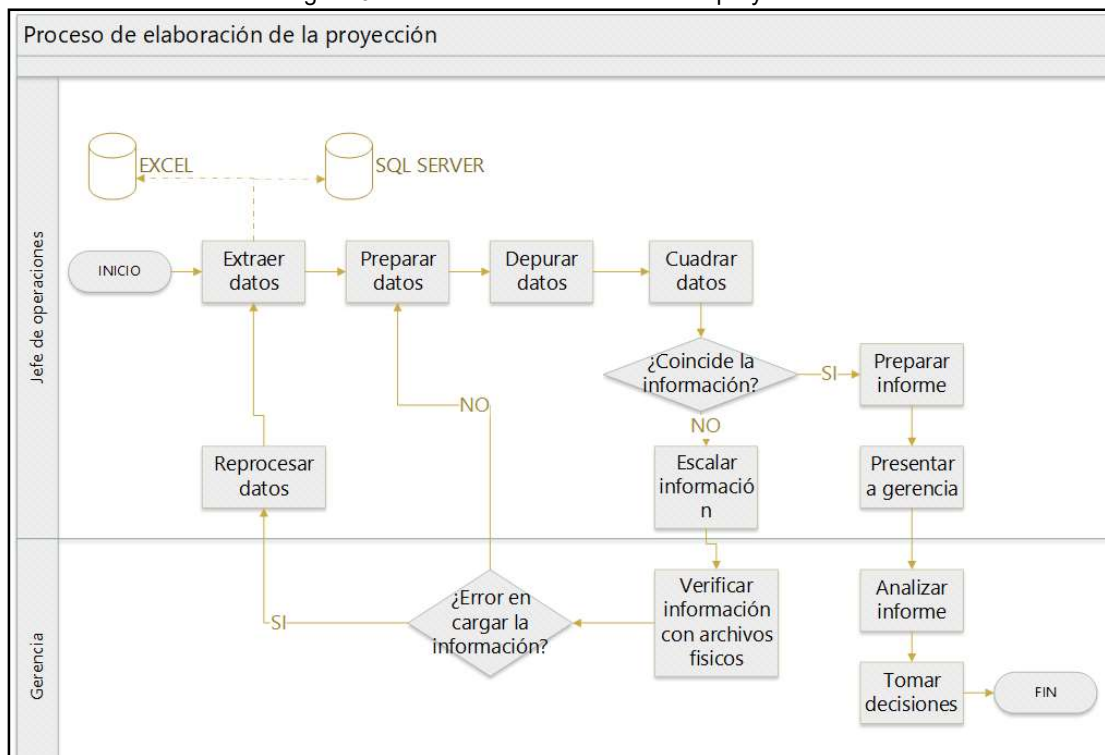
Andrés Fernando Cárdenas Pepinós (2018), presentó en la tesis “Análisis de datos para mejorar la toma de decisiones en la distribución de agua potable del cantón Ibarra, utilizando Business Intelligence”, Universidad técnica del norte, Ecuador. El objetivo de esta investigación fue desarrollar una herramienta Business Intelligence para análisis del consumo de agua del cantón en Ibarra, analizando los datos existentes para ayudar en la toma de decisiones. Su investigación fue de tipo explicativo.

2.2. Marco Teórico

En los últimos años las grandes empresas de agroindustria y fondos de producción agrícola han ido implementando nuevos procesos en el área de Planta, área donde se realiza la conversión de materia prima en productos comerciales, debido a ello, la cantidad de información a aumentado y el control que se le debe realizar también. La empresa facilitó información para poder elaborar un diagrama de flujos, ya que no contaban con uno establecido (Anexo 1).

En la gestión de recursos es donde se realizan las proyecciones, donde se establecen la cantidad de materiales por cada tipo de producto, la cantidad de personal que se requiere en cada área de producción. En este proceso fue identificado el problema. Para realizar las proyecciones se empleaba un personal dedicado a extraer los datos de forma manual desde archivos Excel y base de datos SQL Server, luego preparaba los datos, depuraba por diversas reglas de negocio, cruzaba datos para poder cuadrarlos. El grafico siguiente se describe el flujo.

Figura 01: Proceso de elaboración de la proyección

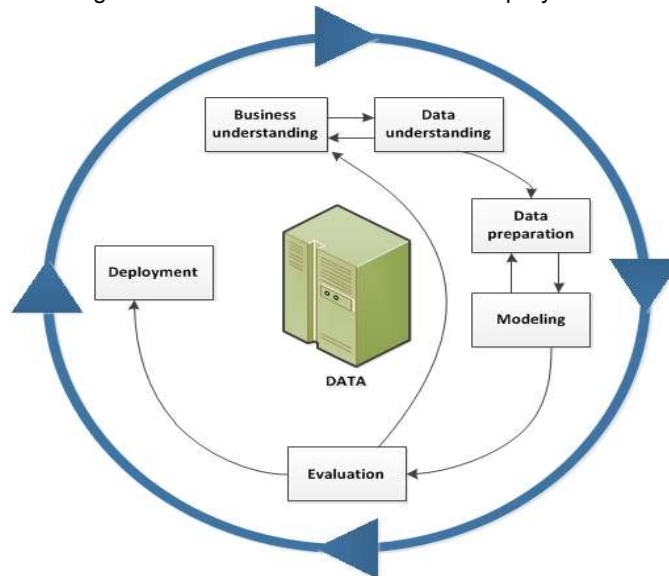


Fuente: Elaboración propia

Como dice William Thomson Kelvin (1890) “Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre”. La propuesta implementada fue una integración de toda la información, para poder realizar análisis predictivos con la herramienta Python. Para poder implementar la

herramienta se usará la metodología Cross Industry Standard Process for Data Mining (CRISP-DM).

Figura 02: Proceso de elaboración de la proyección



Fuente: IBM

Figura 03: Fases de la metodología CRISP-DM

Comprensión del Negocio	Comprensión de Datos	Preparación de Datos	Modelamiento	Evaluación	Despliegue
<ul style="list-style-type: none"> • Determinar los Objetivos del Negocio <ul style="list-style-type: none"> ✓ Antecedentes ✓ Objetivos del Negocio • Evaluar la situación <ul style="list-style-type: none"> ✓ Inventario de requerimientos de Recursos, Hipótesis y Limitaciones ✓ Riesgos y Contingencias ✓ Terminología ✓ Costos y Beneficios • Determinar el objetivo de Minería de Datos <ul style="list-style-type: none"> ✓ Objetivos de Minería de Datos ✓ Criterio de Éxito de Minería de Datos • Desarrollar el Plan de Proyecto <ul style="list-style-type: none"> ✓ Plan de proyecto 	<ul style="list-style-type: none"> • Obtener los datos iniciales <ul style="list-style-type: none"> ✓ Reporte de la obtención de los datos iniciales • Describir los Datos <ul style="list-style-type: none"> ✓ Reporte con la descripción de los datos • Explorar de Datos <ul style="list-style-type: none"> ✓ Reporte de la Exploración de Datos • Verificar de la calidad de los Datos <ul style="list-style-type: none"> ✓ Reporte de la calidad de los datos 	<ul style="list-style-type: none"> • Conjunto de Datos <ul style="list-style-type: none"> • Descripción de los Datos • Seleccionar los Datos <ul style="list-style-type: none"> ✓ Justificación de la inclusión / Exclusión • Limpiar Datos <ul style="list-style-type: none"> ✓ Reporte de Limpieza de Datos • Construir Datos <ul style="list-style-type: none"> ✓ Atributos Derivados ✓ Registros Generados • Integrar Datos <ul style="list-style-type: none"> ✓ Datos Combinados • Dar formato a los Datos <ul style="list-style-type: none"> ✓ Datos Formateados 	<ul style="list-style-type: none"> • Seleccionar Técnica de Modelamiento <ul style="list-style-type: none"> ✓ Técnica de Modelamiento ✓ Modelamiento • Generar el Diseño de Prueba <ul style="list-style-type: none"> ✓ Diseño de Prueba • Construir el Modelo <ul style="list-style-type: none"> ✓ Configuración de los parámetros del Modelo ✓ Descripción del Modelo • Evaluar el Modelo <ul style="list-style-type: none"> ✓ Evaluación del Modelo ✓ Revisión de la configuración de los parámetros del modelo 	<ul style="list-style-type: none"> • Evaluar Resultados <ul style="list-style-type: none"> ✓ Hipótesis de Minería de Datos ✓ Resultados ✓ Criterio de éxito del negocio ✓ Modelos aprobados • Revisar el Proceso <ul style="list-style-type: none"> ✓ Revisión del Proceso • Determinar los siguientes pasos <ul style="list-style-type: none"> ✓ Lista de Posibles Acciones ✓ Decisión 	<ul style="list-style-type: none"> • Desplegar el Plan <ul style="list-style-type: none"> ✓ Plan de Despliegue • Monitorear y Mantener <ul style="list-style-type: none"> ✓ Plan de monitoreo y Mantenimiento • Desarrollar el reporte final <ul style="list-style-type: none"> ✓ Reporte Final ✓ Presentación Final • Revisión del Proyecto <ul style="list-style-type: none"> ✓ Documentación de las experiencias

Fuente: Universidad ESAN

2.3. Marco Conceptual

2.3.1. Definición de proceso de producción agrícola

El proceso de producción se refiere a la serie de pasos mecánicos o químicos utilizados para crear un objeto, que generalmente se repiten para así crear múltiples unidades del mismo elemento. Fabricar un producto implica el uso de materias primas, maquinarias y mano de obra.

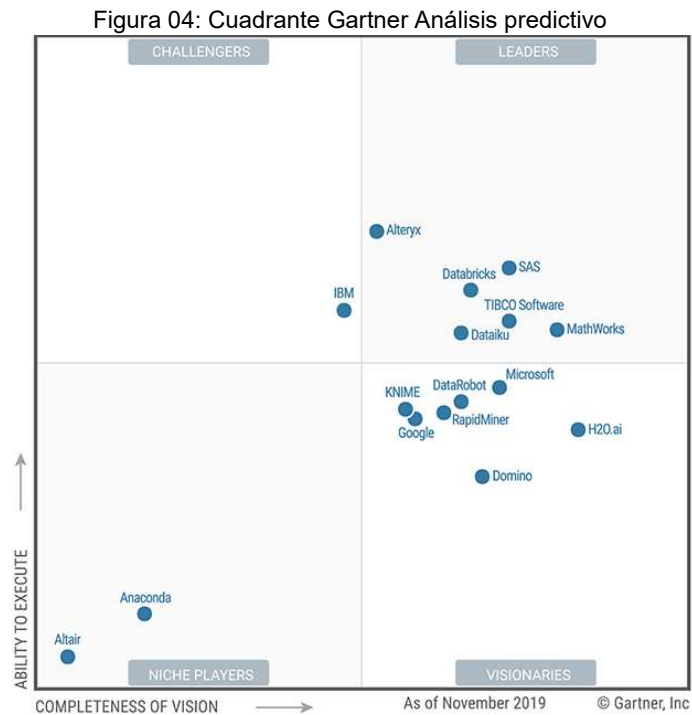
2.3.2. Definición de herramientas Python

Regina Pernaz (2021, p. 2) “Python es un lenguaje de programación interpretado, multiparadigma y multiplataforma usado, principalmente, en Big Data, Inteligencia Artificial, Data Science, frameworks de pruebas y desarrollo web”.

Regina Pernaz (2021, p. 3) “Está desarrollado bajo una licencia de código abierto, por lo que es de libre uso y distribución”.

Dentro de la librería de clases que contiene Python existen herramientas con las cuales se puede realizar análisis predictivos.

- Ventajas de la utilización de herramientas Python: Existen muchas herramientas para realizar análisis predictivos, pero según el cuadrante de Gartner realizado en el 2020 con datos al cierre del año 2019, la mayoría de herramientas que lideran el mercado como Alteryx, Dataliku, Rapidminer y SAS, también Microsoft e IBM que a pesar que en el 2020 no tienen un buen posicionamiento son usados en la mayoría de empresa. Todas las herramientas en mención tienen en común que trabajan con programación Python en sus módulos de trabajo. Trabajar con Python como herramienta independiente brinda una mayor flexibilidad al momento de diseñar los modelos predictivos. Se utiliza como interfaz Spyder, un programa que también es libre.



Fuente: Cuadrante Gartner

2.3.3. Definición de análisis predictivo

El análisis predictivo consiste en tomar de un origen de datos históricos (DataWarehouse) y aplicar un modelo analítico que permita predecir el comportamiento de las variables en rangos de tiempo. Dentro del análisis emplea técnicas estadísticas, aprendizaje automático y minería de datos. Las principales funcionalidades son:

- ✓ Predecir probabilidades y tendencias futuras.
- ✓ Encontrar relaciones entre datos que pueden no ser evidentes en el análisis descriptivo.

Clay Halton (2019) considera que “los modelos predictivos examinan los datos pasados para determinar la probabilidad de ciertos resultados futuros, mientras que los modelos descriptivos examinan los datos pasados para determinar cómo un grupo puede responder a un conjunto de variables”.

UNIR (2021) lo define como: “El análisis predictivo se basa en estimar eventos futuros en función de datos históricos, a los que se les aplican diversas técnicas analíticas, estadísticas y de aprendizaje automático.”

2.3.4. Base de datos

Oracle (2021) define como, “base de datos una recopilación organizada de información estructurada, o datos, normalmente almacenados electrónicamente en un sistema informático. Normalmente, una base de datos se controla mediante un sistema de gestión de bases de datos (DBMS). Juntos, los datos y el DBMS, junto con las aplicaciones asociadas a ellos, se conocen como un sistema de base de datos, a menudo acortado a solo base de datos”.

Microsoft (2021) define como. “base de datos una herramienta para recopilar y organizar información. Las bases de datos pueden almacenar información sobre personas, productos, pedidos u otras cosas.

2.3.5. Definición de Datawarehouse

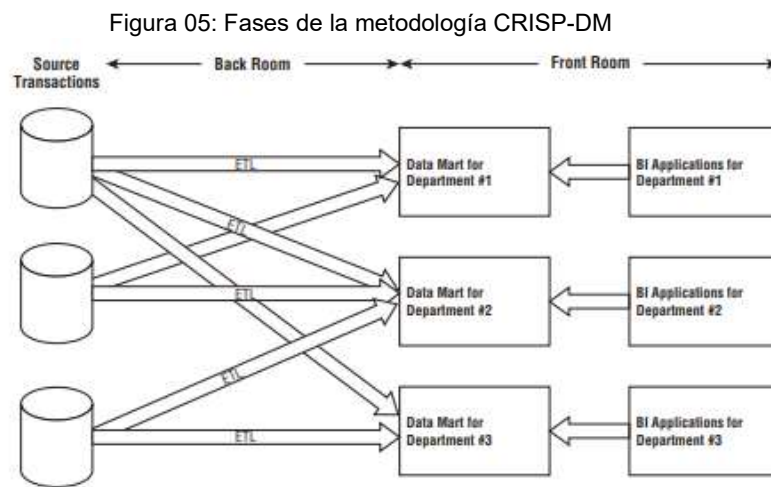
Ralph Kimball y Margy Ross (2013, p.19) El concepto clave de almacenamiento de datos es permitir a los usuarios acceder a una versión unificada de la verdad para la toma de decisiones empresariales, los informes y la previsión oportunos. DWH funciona como un sistema de información que tiene todos los datos pasados y conmutativos almacenados de una o más fuentes.

Irfan Ahmed (2021) define lo siguiente: “El almacenamiento de datos e (DWH) se define como un repositorio de los datos almacenados electrónicamente de una

organización extraídos de sistemas operativos y puestos a disposición para consultas ad hoc e informes programados”.

2.3.6. Metodología Kimball (Arquitectura independiente de Data Mart)

Ralph Kimball y Margy Ross (2013, p.26) Con este enfoque, los datos analíticos se despliegan de forma departamental sin preocuparse de compartir e integrar información en toda la empresa, como se ilustra en la figura. Por lo general, un solo departamento identifica los requisitos para los datos de un sistema de fuentes operativas. El departamento trabaja con personal de TI o consultores externos para construir una base de datos que satisfaga sus necesidades departamentales, reglas de negocio y etiquetado preferido.

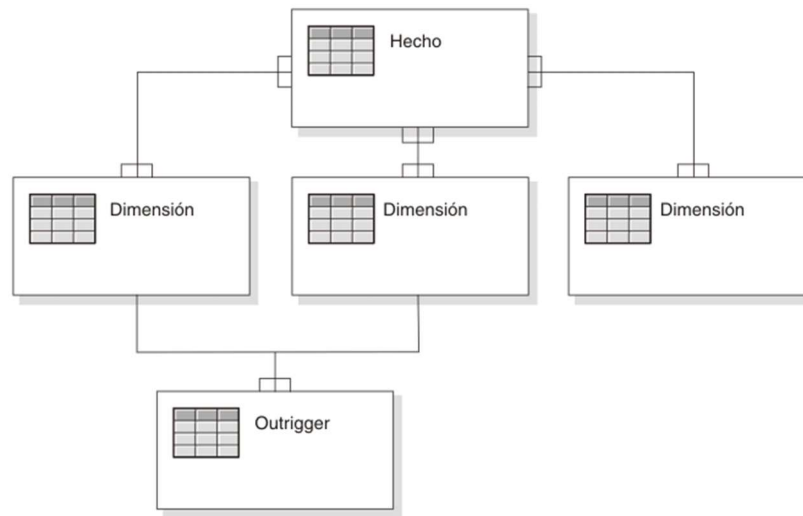


Fuente: The Data Warehouse Toolkit (2013)

2.3.7. Esquema Constelación

IBM define un esquema constelación de la siguiente manera: “Un esquema de constelación es una combinación de un esquema de estrella y un esquema de copo de nieve. Los esquemas de constelación son esquemas de copo de nieve en los que sólo algunas de las tablas de dimensiones se han desnormalizado. El objetivo de los esquemas de constelación es aprovechar las ventajas de los esquemas de estrella y de copo de nieve. Las jerarquías de los esquemas de estrella están desnormalizadas, mientras que las jerarquías de los esquemas de copo de nieve están normalizadas. Los esquemas de constelación están normalizados para eliminar las redundancias de las dimensiones. Para normalizar el esquema, las jerarquías dimensionales compartidas se colocan en outriggers.”

Figura 06: Esquema constelación



Fuente: IBM – Esquema Constelación

2.3.8. Spyder

MTA y otros desarrolladores definen la herramienta de la siguiente manera: “Spyder es un entorno científico gratuito y de código abierto escrito en Python, para Python, y diseñado por y para científicos, ingenieros y analistas de datos. Cuenta con una combinación única de la funcionalidad avanzada de edición, análisis, depuración y creación de perfiles de una herramienta de desarrollo integral con la exploración de datos, ejecución interactiva, inspección profunda y hermosas capacidades de visualización de un paquete científico”.

III. METODOLOGÍA

3.1. Diseño de investigación

3.1.1. Tipo de investigación

El desarrollo de esta investigación es de tipo aplicada, debido a que tiene como objetivo resolver un problema específico, el cual es brindar soporte en la toma de decisiones a través de las proyecciones en el área de planta (análisis predictivo) con la herramienta Python, la cual será utilizada por el analista de la empresa y mostrado a los directivos y personal responsable.

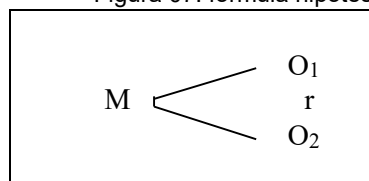
3.1.2. Nivel de investigación explicativa

La investigación busca conocer las variables que tienen mayor significancia y correlación en la producción de los productos agrícolas, así poder tener un análisis más preciso al momento de realizar las proyecciones por periodos de tiempo.

3.1.3. Diseño de investigación

El diseño de la investigación no experimental transeccional correlacional-causal, ya que dentro de la investigación se busca encontrar correlaciones entre las variables a estudiar. Como finalidad se contrastará con la hipótesis para encontrar similitud.

Figura 07: formula hipótesis



Fuente: Elaboración propia

3.2. Variables y operacionalización

Variable Independiente: Herramientas Python

Concepto Operacional: La variable definida para la investigación son de uso del negocio, donde se opera mediante los informes estadísticos realizados por la herramienta Python.

Variable Dependiente:

Variable Proceso de Producción, dentro del proceso se derivan 3 sub procesos:

- Proceso logístico, este proceso se realiza internamente para el área de Planta, consiste en proveer una cantidad de materiales por tipo de proceso a realizar. Gestiona con diferentes proveedores el envío de materiales el cual se almacena y/o desecha.

- Proceso de producción, en el proceso en mención se convierte cierta cantidad de materia prima en productos comerciales, que, dependiendo la materia prima puede tener uno o más productos finales. En este proceso intervienen la mayoría de variables a analizar en la presente investigación.
- Proceso de reclutamiento, este proceso a pesar de pertenecer al área de recursos humanos, interviene muchos los jefes de planta debido a que son ellos quienes deciden la cantidad de personas que se necesitan por rango de tiempo.

Matriz de Operacionalización de variables (Anexo 2)

Matriz de Operacionalización de indicadores (Anexo 3)

3.3. Población, muestra y muestreo

3.3.1. Población

La población para esta investigación son los registros de producción del área de planta de los últimos 4 años desde la campaña iniciada el 15/01/2017 hasta el 15//01/2021. Se llego a determinar una población de 12508 registros en los cuales se incluyen la producción de cada personal por tipo de producto.

3.3.2. Muestra

Por tratarse de un modelo predictivo usando la herramienta Python se tomará dos tipos de muestra.

- Muestra de aprendizaje: Se toma los primeros periodos antes de aplicar el modelo.
- Muestra de test: Se toma el último periodo para poder estimas la desviación en el tiempo. Se usará como estrategia para saber qué hubiese pasado si se actuaba de una determinada manera en un tiempo pasado, así poder simular condiciones pasadas con datos históricos.

3.3.3. Muestreo

Dentro de la metodología a implementar en la investigación, se utilizará todos los registros. Solo para efectos del modelo predictivo se le llama muestra a la segmentación mencionada en el punto anterior.

3.4. Técnicas e instrumentos de recolección de datos

Las técnicas e instrumentos de recolección de datos fueron los siguientes:

Tabla 01: Técnicas e instrumentos de recolección de datos

Técnica	Instrumento	Fuente	Personal necesario
Análisis documental	Herramienta Python	Base de datos Usuario Interno	Analista BI
Observación de campo no experimental	Diario de trabajo	Usuario interno	Usuario del área de Planta

Fuente: Creación propia

3.4.1. Análisis documental

Para la investigación es una operación fundamental para poder contrastar y analizar la información obtenida de las demás técnicas de recolección de datos. Se logró examinar la base de datos con la información histórica de las campañas por cada tipo de producto.

A pesar que dentro de los propios conceptos de la técnica una de las desventajas es que puede recolectar información incompleta o inexacta, la metodología CRISP-MD pudo complementar y corregir este déficit en el instrumento, debido a que en la fase de preparación de datos se aseguró la integridad de la información.

El instrumento usado es la herramienta Python, con la cual se pudo realizar el recorrido y análisis de los datos, también generó gráficos estadísticos que fueron fundamentales para la medición de los indicadores.

3.4.2. Observación de campo no experimental

Para la investigación los datos no son controladas y se desconoce el resultado final, no se manipularon las variables. El instrumento aplicado fue el diario de trabajo, debido a que se necesitaba que personal interno del área de planta pueda monitorear los resultados en todo el proceso de producción, logística y de reclutamiento de personal.

Cada aplicación de los instrumentos se apoyaba del análisis documental ya que era almacenado en la base de datos.

3.5. Procedimientos

Para iniciar la investigación se pudo aplicar una entrevista al responsable del negocio para identificar las necesidades del mismo y así poder estructurar su situación problemática. Una vez detallado la problemática y propuesto los objetivos, se realizaron las observaciones de campo no experimental para las cuales se utilizaron diarios de trabajo independientes por cada variable de estudio

En el área de almacén se realizó la aplicación del diario de trabajo a la jefa de almacén (Ver Anexo 4), donde se detalla el proceso logístico. Para el encargado de producción quien es el jefe de operaciones, se creó un diario de trabajo con su validación como

experto del proceso (Ver Anexo 6). Por último, para el área de recursos humanos se solicitó brindar un diario de trabajo para evaluar el proceso de reclutamiento (Ver Anexo 7).

Para el mejor entendimiento de cada proceso nos facilitaron la información para poder crear un diagrama de flujo del proceso de producción (Ver Anexo 1). Según la metodología CRISP – MD, una vez aplicados todos los instrumentos se procedieron a ingresar a un DataWarehouse (creado en SQL Server con la metodología Kimball), todo ello dentro de la Fase de preparación de datos. Luego se dio inicio a la fase de modelamiento donde se tomó la información ya procesada de los instrumentos y de la base de datos.

En la fase de evaluación se corrieron los modelos predictivos y se hizo la simulación del llenado de los instrumentos en base a los datos arrojados por el modelo, a fin de poder hacer un comparativo de los indicadores del pretest y del post-test.

3.6. Método de análisis de datos

3.6.1. Análisis descriptivo

A fin de poder tener un orden al aplicar los instrumentos de investigación, se organizaron los diarios de trabajo de la siguiente manera:

Tabla 02: Técnicas, instrumentos e indicadores

Técnica	Instrumento	Indicador
Análisis documental	Herramienta Python	Base de datos Usuario Interno
Observación de campo no experimental	Diario de trabajo del proceso logístico	Porcentaje de precisión de materiales. Porcentaje de pérdida de materiales
Observación de campo no experimental	Diario de trabajo del proceso de producción	Porcentaje de precisión de producción Porcentaje de descarte

Fuente: Creación propia

Para la elaboración del post-test se realizó la fase de evaluación donde se utilizó la muestra de test, se registraron cambios en los indicadores. Gracias a la realización de los gráficos de barras, se puede visualizar la información para su correcto análisis y medición.

3.6.2. Análisis inferencial

Se realizó la prueba de normalidad se dio a conocer si siguen o no una distribución normal gracias a los datos que fueron recolectados en el pretest y post-test.

Se aplicó Shapiro-Wilk test y D'Agostino's K-squared test para validar el modelo predictivo. Usando la misma herramienta Python se pudo efectuar el cálculo de las pruebas por cada indicador.

3.7. Aspectos éticos

El investigador es consciente que para la empresa es importante la confidencialidad de los datos, por ello asume el compromiso de resguardar datos sensibles para la empresa y solo mostrar los resultados obtenidos para efectos de la presente investigación.

IV. RESULTADOS

4.1. Análisis descriptivo

En la presente investigación se evaluaron los indicadores que ayudaron a mejorar las proyecciones por campaña de los productos agrícolas, como resultado la toma de decisiones fue más ágil y óptima. Esto se realizó en 2 momentos, antes de implementar la herramienta Python y luego de implementarla. El resultado del post-test fue llenado de forma automática con la herramienta, mientras que el pre-test se había realizado el llenado de forma manual en cada proceso.

Tabla 03. Fechas de recolección de datos

Tipo de prueba	Fecha de inicio	Fecha de término
Pretest	01/04/2017	15/01/2021
Posttest	01/02/2021	20/02/2021

Fuente: Elaboración propia

- **INDICADOR: Nivel de confianza**
Los modelos para obtener la proyección de la producción agrícola se segmentaron en 4. Segmentados por Área y tipo de materia prima:

Tabla 04: Modelos de normalidad

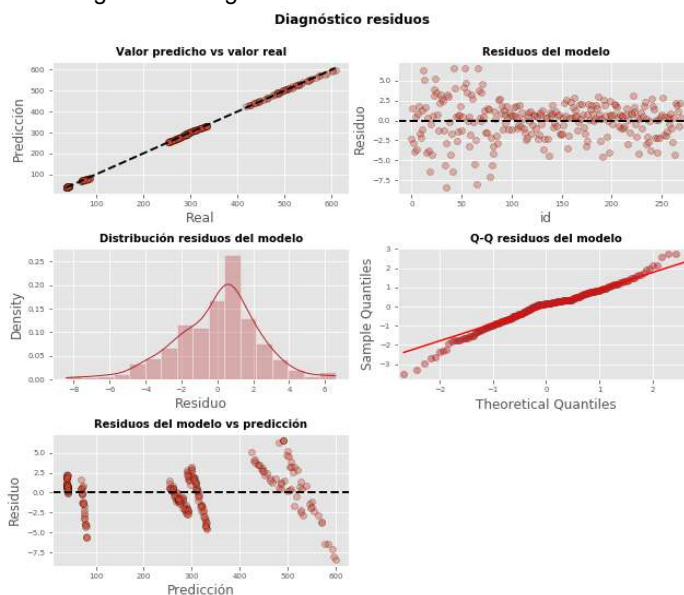
Modelo	Area	Producto
Modelo 1	selección, procesamiento, descarte	Pecana
Modelo 2	Empaquetado	Pecana
Modelo 3	selección, procesamiento, descarte	Esparrago
Modelo 4	Empaquetado	Esparrago

Fuente: elaboración Propia

✓ Modelo 1

El modelo puede predecir al 83.1% de precisión, siendo un nivel aceptable para realizar las proyecciones. Se observa que en los gráficos de existe simetría y que el grafico Normal Q-Q se muestra un comportamiento normal, es decir cumple el supuesto de normalidad. Al cumplirse los supuestos y tener un R2 alto, se acepta el modelo como óptimo.

Figura 08: Diagnostico de residuos – Modelo 1

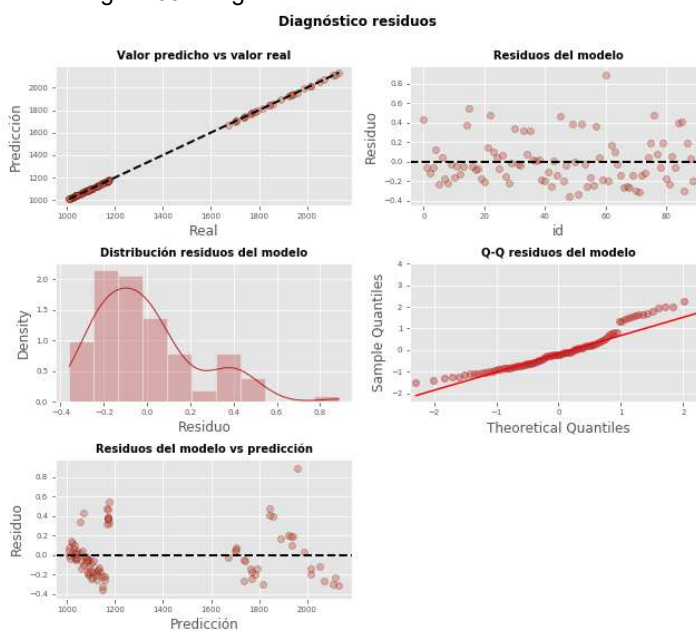


Fuente: elaboración Propia

✓ Modelo 2

El modelo puede predecir al 83.1% de precisión, siendo un nivel aceptable para realizar las proyecciones. Se observa que en los gráficos de existe simetría y que el grafico Normal Q-Q se muestra un comportamiento normal, es decir cumple el supuesto de normalidad. Al cumplirse los supuestos y tener un R2 alto, se acepta el modelo como óptimo.

Figura 09: Diagnostico de residuos – Modelo 2

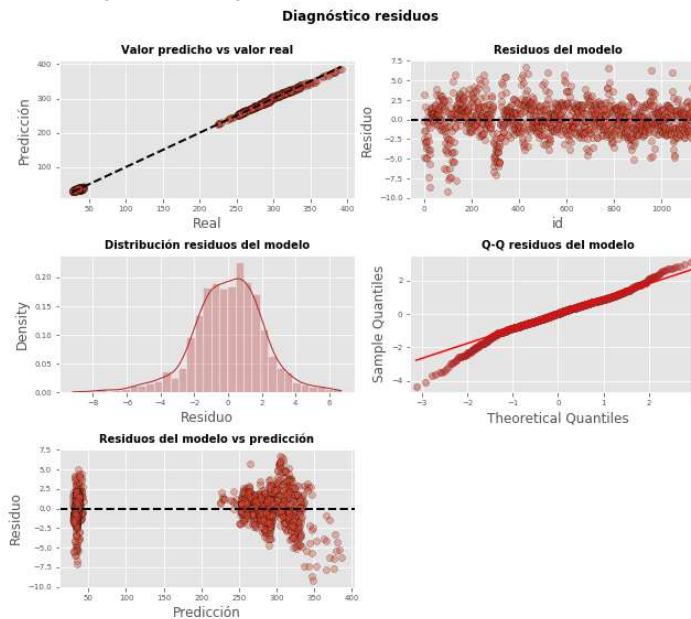


Fuente: elaboración Propia

✓ Modelo 3

El modelo puede predecir al 98% de precisión, siendo un nivel óptimo. Se observa que en los gráficos de existe simetría y que el grafico Normal Q-Q se muestra un comportamiento normal, es decir cumple el supuesto de normalidad. Al cumplirse los supuestos y tener un R2 alto, se acepta el modelo como óptimo.

Figura 10: Diagnostico de residuos – Modelo 3

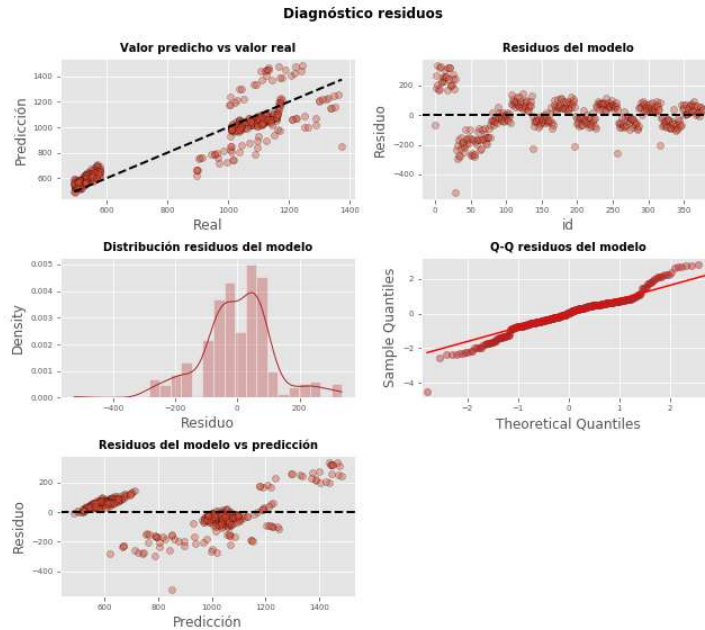


Fuente: elaboración Propia

✓ Modelo 4

El modelo puede predecir al 83.35% de precisión, siendo un nivel aceptable para realizar las proyecciones. Se observa que en los gráficos de existe simetría y que el grafico Normal Q-Q se muestra un comportamiento normal, es decir cumple el supuesto de normalidad. Al cumplirse los supuestos y tener un R2 alto, se acepta el modelo como óptimo.

Figura 11: Diagnóstico de residuos – Modelo 4



Fuente: elaboración Propia

- **INDICADOR:** Porcentaje de precisión de materiales
Los resultados descriptivos en la producción agrícola se observan en la siguiente tabla:

Medias descriptivas del uso de materiales dentro de la producción agrícola antes y después de implementar la herramienta Python.

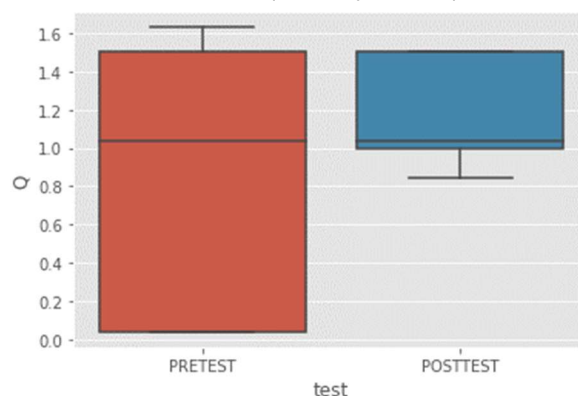
Tabla 05: Tabla estadística de la muestra

Tipo de evaluación	N	Media	viación Estar	Mínimo	Máximo
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico
Pre-test	450	0.91	0.55	0.04	1.63
Post-test	450	1.14	0.22	0.84	1.5

Fuente: elaboración Propia

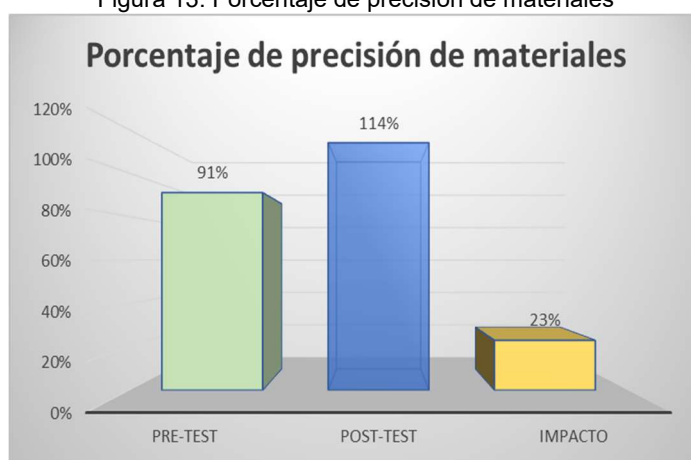
En el caso de la producción agrícola en el pre-test de la muestra se obtuvo un valor de media de 0.91 por registro anual, por lo tanto, en el post-test fue de media 1.14 por año, a primera lectura se podría interpretar una deficiencia en el Post-test. Por ello, como análisis adicional se detectaron los extremos del post-test como datos más próximos a la media. Adicionalmente, la desviación estándar del pre-test es muy alta en comparación a la del Post-test, lo que indican que la variabilidad del post-test es menor y se concluye que es más óptimo.

Figura 12: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “precisión de materiales” (pretest y postest)



Fuente: elaboración Propia (Python)

Figura 13: Porcentaje de precisión de materiales



Fuente: elaboración Propia (Python)

- **INDICADOR: Porcentaje de pérdida de materiales**
Los resultados descriptivos en la producción agrícola se observan en la siguiente tabla:

Medias descriptivas del uso de materiales dentro de la producción agrícola antes y después de implementar la herramienta Python.

Tabla 06: Tabla estadística de la muestra

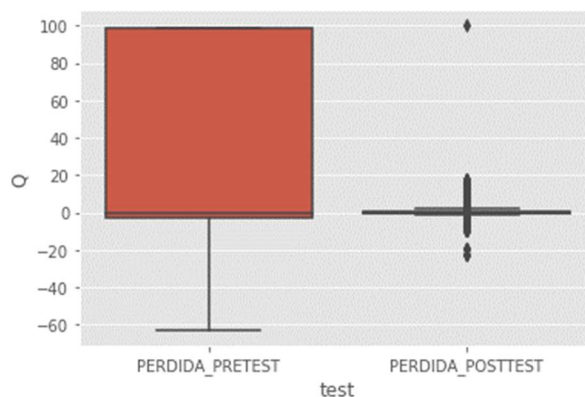
Tipo de evaluacion	N	Media Estadístico	viación Estar Estadístico	Mínimo Estadístico	Máximo Estadístico
Pre-test	450	0.76	0.44	0.01	1.63
Post-test	450	0.99	0.07	0	1.23

Fuente: elaboración Propia

En el caso de la producción agrícola en el pre-test de la muestra se obtuvo un valor de media de 0.76 por registro anual, por lo tanto, en el post-test fue de media 0.99 por año, siendo así el más óptimo. Como análisis adicional se detectaron los extremos del post-test como datos más próximos a la media. Adicionalmente, la desviación estándar del pre-test es muy alta en comparación a la del Post-test, lo

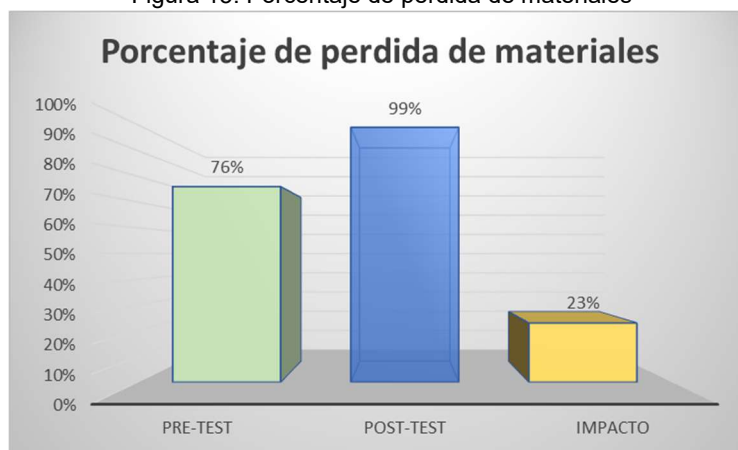
que indican que la variabilidad del post-test es menor y se concluye que es más óptimo.

Figura 14: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “perdida de materiales” (pretest y posttest)



Fuente: elaboración Propia (Python)

Figura 15: Porcentaje de perdida de materiales



Fuente: elaboración Propia (Python)

- INDICADOR: Porcentaje de precisión de producción

Los resultados descriptivos en la producción agrícola se observan en la siguiente tabla:

Medias descriptivas de los ingresos en la producción agrícola antes y después de implementar la herramienta Python.

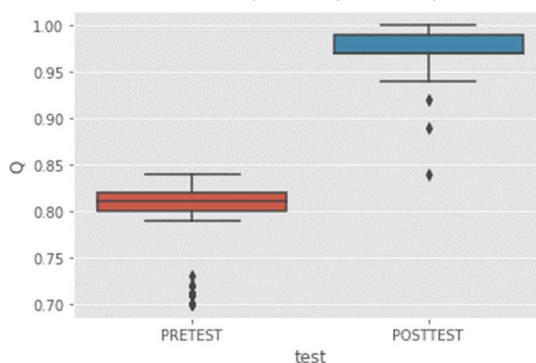
Tabla 07: Tabla estadística de la muestra

Tipo de evaluacion	N	Media	viación Estar	Mínimo	Máximo
	Estadistico	Estadistico	Estadistico	Estadistico	Estadistico
Pre-test	600	0.82	0.26	0.04	0.98
Post-test	600	0.97	0.03	0.76	1

Fuente: elaboración Propia

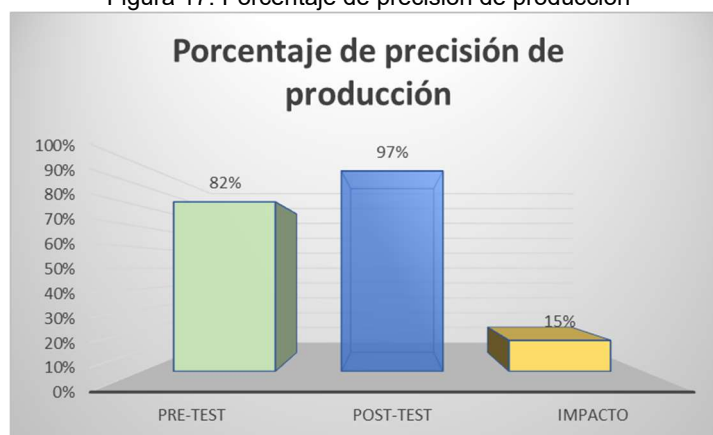
En el caso de la producción agrícola en el pre-test de la muestra se obtuvo un valor de media de 0.82 por registro anual, por lo tanto, en el post-test fue de media 0.97 por año, siendo así el más óptimo. Adicionalmente, la desviación estándar del pre-test es muy alta en comparación a la del Post-test, lo que indican que la variabilidad del post-test es menor y se concluye que es más óptimo.

Figura 16: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de “precisión de producción” (pretest y postest)



Fuente: elaboración Propia (Python)

Figura 17: Porcentaje de precisión de producción



Fuente: elaboración Propia (Python)

- **INDICADOR: Porcentaje de precisión de descarte**
Los resultados descriptivos en la producción agrícola se observan en la siguiente tabla:

Medias descriptivas de los ingresos para el área de descarte en la producción agrícola antes y después de implementar la herramienta Python.

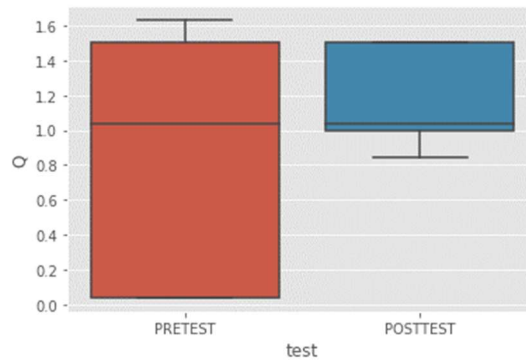
Tabla 08: Tabla estadística de la muestra

Tipo de evaluación	N Estadístico	Media Estadístico	viación Estar Estadístico	Mínimo Estadístico	Máximo Estadístico
Pre-test	150	0.79	0.04	0.7	0.84
Post-test	150	0.97	0.02	0.84	1

Fuente: elaboración Propia

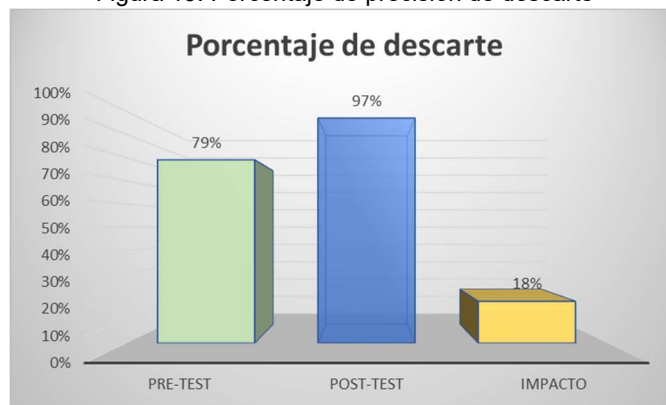
En el caso de la producción agrícola en el pre-test de la muestra se obtuvo un valor de media de 0.79 por registro anual, por lo tanto, en el post-test fue de media 0.97 por año, siendo así el más óptimo. Adicionalmente, la desviación estándar del pre-test es ligeramente mayor en comparación a la del Post-test, lo que indica que la variabilidad del post-test es menor y se concluye que es más óptimo.

Figura 18: Diagrama de Caja y Bigote del Indicador del indicador Porcentaje de "precisión de descarte" (pretest y posttest)



Fuente: elaboración Propia (Python)

Figura 19: Porcentaje de precisión de descarte



Fuente: elaboración Propia

4.2. Análisis de Inferencial

Se comprueba si los residuos siguen una distribución normal empleando dos test estadísticos: Shapiro-Wilk test y D'Agostino's K-squared test. Para la investigación se realizaron los test de normalidad por cada modelo de regresión múltiple.

- **Modelo 1**
Según ambos test consideran que los datos siguen una distribución normal debido a que los P-value son inferiores al nivel de referencia Alpha (0.01). Se puede aceptar la hipótesis nula.

Tabla 09: test normalidad Modelo 1

Tipo de test	Estadístico	P-value
Shapiro-Wilk	0.97	0.0001
D'Agostino's K-squared	13.85	0.0009

Fuente: elaboración Propia

- **Modelo 2**
Según ambos test consideran que los datos siguen una distribución normal debido a que uno de los P-value es inferior al nivel de referencia Alpha (0.01). Se puede aceptar la hipótesis nula.

Tabla 10: test normalidad Modelo 2

Tipo de test	Estadístico	P-value
Shapiro-Wilk	0.914	1.84
D'Agostino's K-squared	18.35	0.0001

Fuente: elaboración Propia

- **Modelo 3**
Según ambos test consideran que los datos siguen una distribución normal debido a que los P-value son superiores al nivel de referencia Alpha (0.01). Se rechaza la hipótesis nula.

Tabla 11: test normalidad Modelo 3

Tipo de test	Estadístico	P-value
Shapiro-Wilk	0.98	7.25
D'Agostino's K-squared	60.68	6.63

Fuente: elaboración Propia

- **Modelo 4**
Según ambos test consideran que los datos siguen una distribución normal debido a que uno de los P-value es inferior al nivel de referencia Alpha (0.01). Se puede aceptar la hipótesis nula.

Tabla 12: test normalidad Modelo 4

Tipo de test	Estadístico	P-value
Shapiro-Wilk	0.96	1.62
D'Agostino's K-squared	15.88	0.0003

Fuente: elaboración Propia

V. DISCUSIÓN

En base a los resultados de la presente investigación, se analizó el nivel de precisión que el modelo de regresión tiene para predecir los resultados de la producción agrícola para el año 2020 en el fundo “San Juan de Buenavista”. Para desarrollar la implementación se utilizó la metodología CRISP-DM, metodología que a diferencia de las demás no está ligada a ningún producto, es abierta y gratuita. Esta orientada a los objetivos empresariales y una metodología de gestión de proyectos, también es adaptable a una metodología ágil.

Juan Francisco Vallalta Rueda (2021) afirma que, en el ámbito de la minería de datos, la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) se ha convertido en un estándar de facto del mercado. Surgida en dos empresas que han sido pioneras en la aplicación de minería de datos a los procesos de negocio: DaimlerChrysler y SPSS.

La herramienta Python es ligera a comparación de las herramientas de pago que necesitan especificaciones técnicas más avanzadas, lo que ayuda a pequeñas empresas que disponen de pequeños servidores en los cuales almacenan su información. Al contar con un analista ya sea como posición fija o como consultor, se pueden monitorear cambios en el modelo productivo, que tiende a mejorar un 5% por cada periodo de tiempo. Para la implementación se usó como motor de base de datos SQL Server Express 2017, con la finalidad de poder centralizar la información y crear el datamart, al utilizar Python no se necesitan las funcionalidades de alguna suscripción de SQL Server. Álex Rayón (2015) menciona lo siguiente: “no es fácil quedarse con una herramienta de analítica y las personas que trabajamos en grandes compañías estamos habituados a trabajar con varias. SAS ofrece soluciones integradoras a un costo elevado. R tiene muchas capacidades de innovación debido a su origen y Python tiene la ventaja de ser un lenguaje de programación generalista que además puede servir para hacer Data Mining o Machine Learning. La elección dependerá de lo que estés dispuesto a pagar y tus necesidades específicas”. Por los motivos mencionados por el autor la elección de la herramienta Python se hizo más fácil.

Se utilizó el esquema constelación para optimizar la comunicación entre diferentes áreas, como Almacén y Planta. Según lo investigado, todos los Datawarehouse terminan convirtiendo sus esquemas copo de nieve o estrella en constelaciones al momento de comunicarse entre sus áreas del negocio. Bernabeu Dario (2019) refiere lo siguiente: “Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos”.

PowerData (2020) la empresa menciona que “La preparación y optimización de datos nunca es una tarea lineal, sino que es un proceso de refinamiento iterativo. De allí el valor clave del trabajo de gestión de datos. Adquirir los datos correctos para entrenar al algoritmo es una

tarea central”. De esta manera se realizó el tratamiento de los datos y el mapeo de procesos que llevaron a la implementación de la herramienta. Mientras que Stephen Goldsmith (2017) menciona “el análisis predictivo se debería utilizar para mejorar la eficacia y la discreción de los empleados del gobierno, no para reemplazar su intuición, conocimiento local y experiencia. El uso óptimo de esta herramienta corresponde al complemento de las prácticas existentes”. La herramienta ayudó a realizar los análisis predictivos usados para predecir la el proceso de producción agrícola, evaluados por los indicadores propuestos en la investigación. Guilherme Barra (2018) define el análisis predictivo como: “la unión del avance tecnológico que tenemos, que incluye minería de datos, machine learning, inteligencia artificial y estadística, con el alto volumen de informaciones que creamos diariamente”. Se especifica los siguientes indicadores:

1. Para el indicador Nivel de confianza de la variable dependiente, se obtuvieron para todos los modelos índices mayores al 80% de precisión. El cuál nos indica que toda proyección que se realizará optima. Según los diferentes modelos existentes, el modelo con el que se trabajó logró mejorar el nivel de precisión donde se evalúa el R cuadrado y los supuestos de normalidad son óptimos.
2. Para el indicador Porcentaje de precisión de materiales, disminuyo su desviación estándar en 23 puntos porcentuales, lo cual demostró que los datos son menos variantes y más confiables a comparación de antes de implementar la herramienta. La media resultante es de 82% antes de la implementación y 97% después. Por tanto, se corrobora que la herramienta es 15% más precisa que las proyecciones realizadas en la actualidad en el fundo “San Juan de Buenavista”.
3. Para el indicador Porcentaje de precisión de la perdida de materiales, antes de la implementación de la herramienta contaba con una media de 76%, posterior a ella aumento hasta un 99%. Con este nivel de precisión se corrobora que los datos calculados para las siguientes campañas serán 13% más exactos que las campañas anteriores del fundo “San Juan de Buenavista”.
4. Para el indicador Porcentaje de precisión de producción, se pudo mejorar el nivel de precisión antes de implementar la herramienta, la cual era de 82% y después de la implementación llego a 97%. Con este nivel de precisión se corrobora que los datos calculados para las siguientes campañas serán 15% más exactos que las campañas anteriores del fundo “San Juan de Buenavista”.
5. Para el indicador Porcentaje de precisión de producción, se pudo mejorar el nivel de precisión antes de implementar la herramienta, la cual era de 79% y después de la implementación llego a 97%. Con este nivel de precisión se corrobora que los datos

calculados para las siguientes campañas serán 18% más exactos que las campañas anteriores del fondo "San Juan de Buenavista".

VI. CONCLUSIONES

La implementación de la herramienta Python ha logrado aumentar el nivel de precisión de las proyecciones del proceso de producción agrícola y logró centralizar los datos, se tienen las siguientes conclusiones:

- Se logró la centralización de los datos del proceso de producción agrícola gracias a la creación de un datamart para el área de Planta, los datos ingresados fueron sincronizados a un 100% gracias a la implementación del ETL.
- Se llevó a cabo la implementación de la herramienta Python en los servidores del área TI, integrando las herramientas y colocando el desarrollo en producción.
- En cuanto al diseño del modelo predictivo con la herramienta Python, se logró optimizar el nivel de precisión superando el 80%. Con ello se pudo concluir que el modelo creado puede predecir la producción agrícola de las siguientes campañas con un error máximo del 10%, teniendo en cuenta que cada modelo predictivo tiende a mejorar un 5% cada periodo de tiempo ingresado.

VII. RECOMENDACIONES

- Realizar un mapeo optimo de los procesos con el fin de diseñar un datamart optimo con la granularidad correspondiente a la necesidad del negocio.
- Comprobar la significancia de las variables por medio de un cuadro de correlación para identificar las variables que se usaran en el modelo predictivo.
- Validar cada inicio de campaña que el modelo predictivo aumente en un 5% comparado con el ultimo nivel de precisión, de no ser el caso, reconsiderar las variables y los outliers del modelo.

REFERENCIAS

Edwin Williams Heredia Salinas. Inteligencia de negocios aplicada a la gestión de ventas de una empresa agroindustrial (Grado de Ingeniero de sistemas). Perú: Universidad Cesar Vallejo, 2019. Disponible en: [Heredia_SEW.pdf \(ucv.edu.pe\)](#)

Montero Silva Boris Javier. Implementación de una solución de inteligencia de negocios para una empresa agroindustrial (Grado de Ingeniero informático). Perú: Universidad nacional de Piura, 2018. Disponible en: https://drive.google.com/file/d/1_05wTfJp2e5kh1iCheZIM8E5jIIPUPGr/view?usp=sharing

Candanoza Rey Carlos Andrés, Pacheco Casiani Henuar Javier, Puerta Botero Paula Andrea, Murillo Sandra Marcela, Villamil Valencia Fredy Alexander. Propuesta de implementación de herramienta de business intelligence orientada en la reducción de costos de adquisición de materia prima en una empresa procesadora y comercializadora de alimentos en la región del eje cafetero (Post grado Business Intelligence). Colombia: Institución universitaria politécnico grancolombiano, 2019. Disponible en: <http://hdl.handle.net/10823/1845>

Espinoza Vega Ángel Andrés, Jiménez Ruiz Fernanda Elizabeth. Implementación de un módulo de soporte de decisiones para agricultura utilizando Bigdata e inteligencia de negocios integrado al sistema iotmach (Grado de Ingeniero de sistemas). Ecuador: Universidad técnica de Machala, 2016. Disponible en: <https://drive.google.com/file/d/1a412p82HD-cLjuD65oXYEU1rUFdv8Qt4/view?usp=sharing>

Nixon Omar Fernández Carrión. La influencia de la inteligencia de negocios en el análisis de información de ventas de la importadora y distribuidora Jiménez E.I.R.L, en la ciudad de Nueva Cajamarca, 2017 (Grado de Ingeniero de sistemas). Perú: Universidad nacional de san martín-tarapoto, 2018. Disponible en: [FISI - Nixon Omar Fernández Carrión.pdf \(unsm.edu.pe\)](#)

Arrobo Lapo, Estalin Vladimir. Business intelligence para la generación de indicadores y control de desempeño, en el área de compras agrícolas de la empresa "Pronaca" (Grado de Ingeniero de sistemas). Ecuador: Universidad regional autónoma de los andes, santo domingo, 2017. Disponible en: <http://dspace.uniandes.edu.ec/handle/123456789/7524>

Marlene Elisa Carhuaricra Inocente, Jenny Isabel Gonzales Caporal. Implementación de business intelligence para mejorar la eficiencia de la toma de decisiones en la gestión de proyectos (Grado de Ingeniería Empresarial y de Sistemas). Perú: Universidad San Ignacio de Loyola, 2017. Disponible en: [2017_Gonzales-Caporal.pdf \(usil.edu.pe\)](#)

Barboza Morales Italo Shamir, Huamaní Portocarrero Javier Eduardo. Implementación de un modelo de business intelligence orientado a tecnología mobile basado en sap businessobjects para pymes del sector Retail (Grado de Ingeniero de sistemas). Perú: Universidad peruana de ciencias aplicadas, 2016. Disponible en: https://drive.google.com/file/d/1KVJXqTviWI_5ldEDMRVnDH4HO9OwG4IR/view?usp=sharing

Santiago Leonardo Morales Cardoso. Metodología para procesos de inteligencia de negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones. (Doctorado en Informatica). España: Universidad de alicante, 2019.

Disponible en: [Metodología para procesos de Inteligencia de Negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones. \(ua.es\)](#)

Lorena Beatriz Bellido Camacho. Propuesta de mejora en la toma de decisiones aplicando business intelligence caso: agencia de promoción de inversiones (Grado de Ingeniería Empresarial) Perú: Universidad San Ignacio de Loyola, 2019. Disponible en: [2019_Bellido-Camacho.pdf \(usil.edu.pe\)](#)

Andrés Fernando Cárdenas Pepinós. Análisis de datos para mejorar la toma de decisiones en la distribución de agua potable del cantón Ibarra, utilizando Business Intelligence (Magíster en Ingeniería de sistemas), Ecuador: Universidad técnica del norte, 2018. Disponible en: [PG 655 TESIS.pdf \(utn.edu.ec\)](#)

Pablo Esteban Villota Neira. Implementación de una estación prototipo con visión artificial, aplicado a la agricultura de precisión (Grado Ingeniero en Electrónica y Telecomunicaciones), Ecuador: Universidad de Cuenca, 2019. Disponible en: [Trabajo de Titulación.pdf \(ucuenca.edu.ec\)](#)

William Thomson Kelvin (1890): Conceptos básicos de ayuda de CRISP-DM - IBM Documentation

Regina Pernaz (2021): [¿Qué es Python? El lenguaje de programación del 2021 | Crehana ES](#)

Ralph Kimball y Margy Ross (2013): Se encuentra en el libro “The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling”

William Thomson Kelvin (1890):

Cuadrante de Gartner: Gartner Magic Quadrant 2020 - Data Science and Machine Learning Tools ([analyticsvidhya.com](#))

IBM – Esquema constelación: Esquemas de constelación - IBM Documentation

MTA y otros desarrolladores: Home — Spyder IDE ([spyder-ide.org](#))

PowerData (2020): Modelos predictivos: sus datos, ¿están listos? ([powerdata.es](#))

Stephen Goldsmith (2017): <https://blogs.iadb.org/administracion-publica/es/analisis-predictivo-impulsar-mejoras-mediante-uso-datos/>

Clay Halton (2019): <https://www.investopedia.com/terms/p/predictive-analytics.asp>

Oracle (2021): What Is a Database | Oracle

Álex Rayón (2015): Eligiendo una herramienta de Analítica: SAS, R o Python - Deusto Data

Irfan Ahmed (2021): What is Data Warehousing? Concepts, Tools, Examples ([astera.com](#))

Juan Francisco Vallalta Rueda (2021): <https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/>

Bernabeu Dario(2019): 3.4 Datawarehouse manager | Dataprix TI

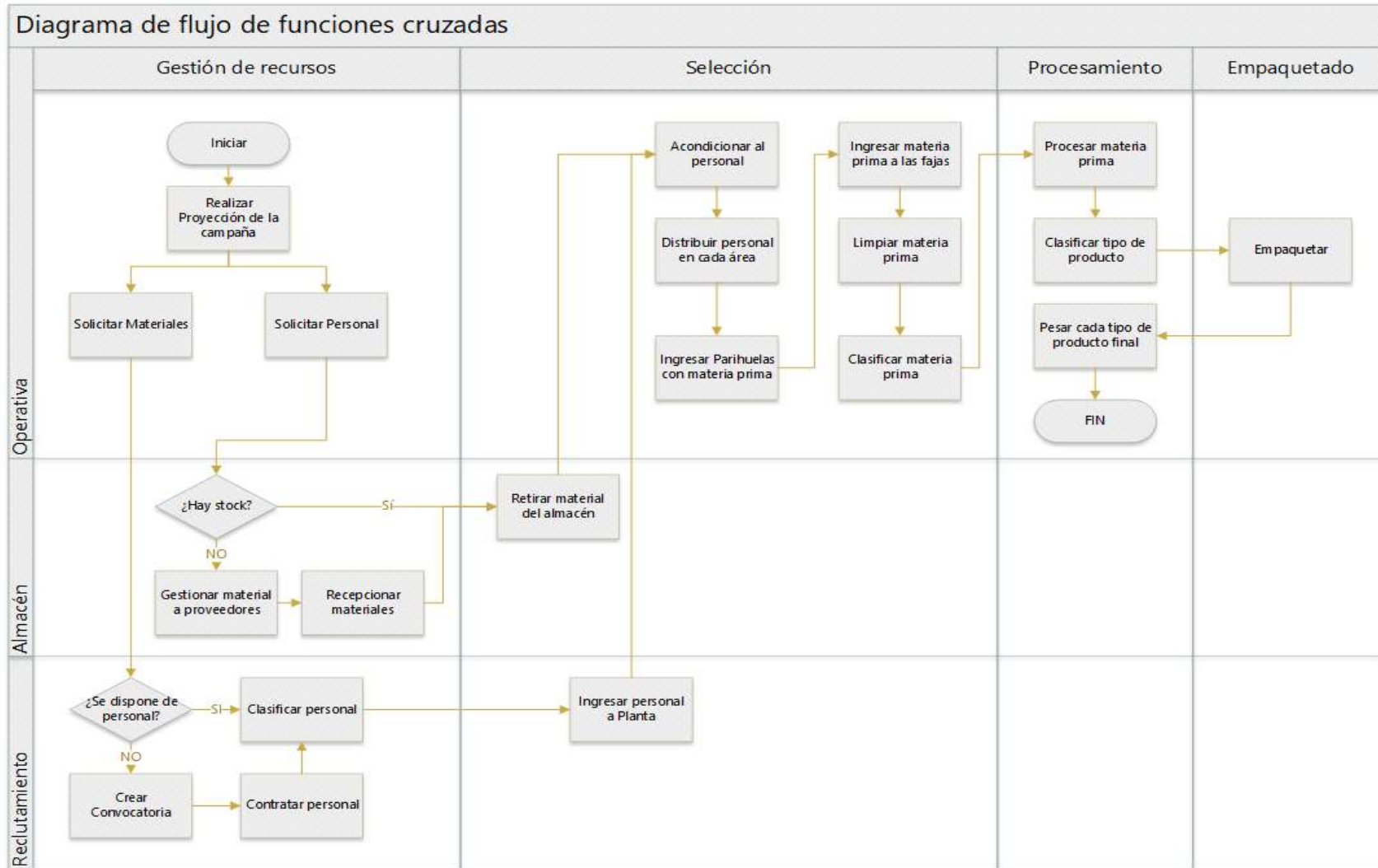
Microsoft (2021): Conceptos básicos sobre bases de datos - Access ([microsoft.com](#))

UNIR (2021): Análisis predictivo, ¿sabes en qué consiste? ([unir.net](#))

Guilherme Barra (2018): Análisis predictivo: cómo aplicar los datos para preveer el futuro
(rockcontent.com)

ANEXOS

Anexo 01: Flujo de procesos



Anexo 02 – Matriz de operacionalización de variables

Variables de estudio	Definición conceptual	Definición operacional	Dimensión	Indicadores	Escala de medición
Herramientas Python	Regina Pernaz (2021, p. 2) Python es un lenguaje de programación interpretado, multiparadigma y multiplataforma usado, principalmente, en Big Data, Inteligencia Artificial, Data Science, frameworks de pruebas y desarrollo web.	Se realizan las ejecuciones de los modelos creados, los cuales comparan la información proyectada con el último año.		Nivel de confianza	Razón
Proceso de producción	El proceso Logístico consiste en gestionar el ingreso de materiales a almacén, conservarlo y distribuirlo para que el proceso de producción se desarrolle sin retrasos.	Se distribuye la cantidad de materiales según sea estimada en la proyección de la producción en el proceso.	Proceso Logístico	Porcentaje de precisión de materiales Porcentaje de precisión de la pérdida de materiales	Razón
	El proceso de producción se encarga de convertir la materia prima en productos comerciales, dentro del cual se encuentran todas las variables que pueden impactar en el desempeño y productividad.	Se estima la cantidad de productos finales que se obtendrá por proceso	Proceso de producción	Porcentaje de precisión de producción Porcentaje de precisión de descarte	Razón

Anexo 03 – Matriz de operacionalización de indicadores

Indicador		Descripción	Técnica	Instrumento	Unidad de medida	Fórmula
Intervalo de confianza	de	Nivel de exactitud del modelo predictivo en predecir los resultados esperados.	Análisis documental	Herramienta Python	porcentual	$IC = \bar{x} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{(n)}}$ <p>Donde: \bar{x} : Media de la muestra $z_{\frac{\alpha}{2}}$: Coeficiente de confianza σ : Desviación estándar</p>
Porcentaje precisión materiales	de de	Se designa la cantidad de materiales acorde a la proyección de la producción y se contrasta con la cantidad de materiales usados.	Observación de campo no experimental	Diario de trabajo	Numérico Porcentual	$PRM = \left(1 - \left \frac{Pr-C}{C}\right \right) * 100\%$ <p>Donde Pr: Cantidad materiales en la proyección C: Cantidad de materiales usados</p>
Porcentaje perdida materiales	de de	Cantidad de materiales perdidos en almacén. (se pierden por desperfecto, caducidad, etcétera)	Observación de campo no experimental	Diario de trabajo	Numérico Porcentual	$PM = \left(1 - \left \frac{P-C}{C}\right \right) * 100\%$ <p>Donde P: Cantidad materiales perdidos C: Cantidad de materiales usados</p>
Porcentaje precisión producción	de de	Cada inicio de campaña se realizan reuniones donde se estiman las proyecciones de producción, esto se contrasta con la producción final	Observación de campo no experimental	Diario de trabajo	Numérico Porcentual	$PP = \left(1 - \left \frac{Pr-C}{C}\right \right) 100\%$ <p>Donde Pr: Cantidad productos en la proyección C: Cantidad productos terminados</p>
Porcentaje descarte	de	Dentro del proceso de producción se realiza dos descartes, uno por selección de producto y otro por tipo de producto.	Observación de campo no experimental	Diario de trabajo	Numérico Porcentual	$PD = \left(1 - \left \frac{Pr-D}{D}\right \right) * 100\%$ <p>Donde Pr: Cantidad descarte en la proyección D: Cantidad de descarte</p>

Anexo 04 – Diario de trabajo del proceso Logístico

Diario de Trabajo

Encargado:
Codigo de proceso:

Fecha:
Campaña:

Distribución de materiales					
Esparragos			Pecana		
Preguntas	Proyección	Real	Preguntas	Proyección	Real
¿Cuántas cajas de 24 unidades se solicitaron?			¿Cuántas bolsas hermeticas se utilizaron?		
¿Cuántas ligas para atados se utilizaron?			¿Cuántas cintas de sellos se utilizaron?		
¿Cuántas bolsas se utilizaron?			¿Cuántas Javas se utilizaron?		
¿Cuántas tiras de sellos se utilizaron?					
¿Cuántas Javas se utilizaron?					

Informe de almacen					
Esparragos			Pecana		
Preguntas	Proyección	Real	Preguntas	Proyección	Real
¿Cuántas cajas de 24 unidades se dañaron en el proceso?			¿Cuántas bolsas hermeticas se dañaron en el proceso?		
¿Cuántas cajas de 24 unidades se dañaron en almacen?			¿Cuántas bolsas hermeticas se dañaron en almacen?		
¿Cuántas tiras de sellos se dañaron en el proceso?			¿Cuántas tiras de sellos se dañaron en el proceso?		
¿Cuántas tiras de sellos se dañaron en el almacen?			¿Cuántas tiras de sellos se dañaron en almacen?		
¿Cuántas bolsas se dañaron?			¿Cuántas Javas fueron prestadas de otra area?		
¿Cuántas Javas fueron prestadas de otra area?					

Observaciones					
Esparragos			Pecana		
Preguntas	SI	NO	Preguntas	SI	NO
¿Se acabó algun material duran el proceso?			¿Se acabó algun material duran el proceso?		
¿El proceso se detuvo?			¿El proceso se detuvo?		
¿Se realizo una compra de materiales al proveedor de emergencia?			¿Se realizo una compra de materiales al proveedor de emergencia?		

Responsable del documento

Cecilia Mendoza Arnao
Jefe de almacen

Anexo 05 – Diario de trabajo del proceso de producción

Diario de Trabajo de Produccion en Planta

Encargado: _____

Fecha: _____

Codigo de proceso: _____

Campaña: _____

Proceso de produccion							
Esparragos					Pecana		
Preguntas	Proyección	Real			Preguntas	Proyección	Real
¿Cuántos Kg ingresaron?					¿Cuántos Kg ingresaron?		
¿Cuántos productos terminados se obtuvieron?					¿Cuántos productos terminados se obtuvieron?		
Etapa de selección					Etapa de selección		
¿Cuántos Kg continuaron el proceso?					¿Cuántos Kg continuaron el proceso?		
¿Cuánto tiempo demoró la etapa?					¿Cuánto tiempo demoró la etapa?		
Etapa de procesamiento					Etapa de procesamiento		
¿Cuántos Kg continuaron el proceso?					¿Cuántos Kg continuaron el proceso?		
¿Cuánto tiempo demoró la etapa?					¿Cuánto tiempo demoró la etapa?		
Etapa de empaquetado					Etapa de empaquetado		
¿Cuántos Kg continuaron el proceso?					¿Cuántos Kg continuaron el proceso?		
¿Cuánto tiempo demoró la etapa?					¿Cuánto tiempo demoró la etapa?		
Peso del descarte					Peso del descarte		
¿Cuántos Kg ingresaron al descarte?					¿Cuántos Kg ingresaron al descarte?		
¿Cuántos kg se dieron como perdida?					¿Cuántos kg se dieron como perdida?		

Observaciones							
Esparragos					Pecana		
Preguntas	Proyección	Real			Preguntas	Proyección	Real
¿Cuántos Kg se quedaron sin procesar?					¿Cuántos Kg se quedaron sin procesar?		
¿Cuántos adiconales del almacen ingresaron?					¿Cuántos adiconales del almacen ingresaron?		

Responsable del documento

Yuri Mendoza Arnao
Jefe de operaciones

Anexo 06 – Diagrama de constelación

