



Universidad CENFOTEC

Maestría en Ingeniería del Software Énfasis: Inteligencia Artificial

Documento final de Proyecto de Investigación Aplicada 1

Sistema predictivo basado en inteligencia artificial para anticipar la calidad del jugo natural a partir de características fisicoquímicas de la fruta antes del procesamiento

Elaborado por: Chavarría Cubero Luis David

Diciembre, 2025

## Tabla de contenido

<i>Tabla de Figuras</i>	5
<i>Índice de Tablas</i>	5
<i>Abstract</i>	6
<b>1 Capítulo 1. Introducción</b>	<b>7</b>
<b>1.1 Generalidades</b>	<b>7</b>
<b>1.2 Definición y Descripción del Problema</b>	<b>8</b>
<b>1.3 Definición y Descripción del Problema</b>	<b>8</b>
<b>1.4 Justificación</b>	<b>9</b>
<b>1.5 Viabilidad</b>	<b>10</b>
1.5.1 Punto de Vista Técnico.	10
1.5.2 Punto de Vista Operativo	11
1.5.3 Punto de Vista Económico	12
<b>1.6 Objetivos</b>	<b>13</b>
1.6.1 Objetivo General.	13
1.6.2 Objetivos Específicos.	14
<b>1.7 Alcances y Limitaciones</b>	<b>14</b>
1.7.1 Alcances.	15
1.7.2 Limitaciones	15
<b>1.8 Estado de la cuestión</b>	<b>16</b>
1.8.1 Formulación de la pregunta.	18
<b>2 Capítulo 2. Marco Conceptual</b>	<b>30</b>
<b>2.1 Concepto de calidad del jugo natural</b>	<b>31</b>
2.1.1 Factores que influyen en la calidad del jugo	32
2.1.2 Parámetros sensoriales vs. parámetros fisicoquímicos	33
2.1.3 Estándares de calidad en la industria agroindustrial	33
<b>2.2 Características fisicoquímicas de las frutas</b>	<b>34</b>
2.2.1 Acidez titulable	35
2.2.2 Sólidos solubles (°Brix)	35
2.2.3 Relación	35
2.2.4 Firmeza y color	35
2.2.5 Madurez	36
<b>2.3 Introducción a la inteligencia artificial</b>	<b>36</b>
2.3.1 Historia y evolución de la inteligencia artificial	38
2.3.2 La inteligencia artificial (IA)	39
2.3.3 Diferencias entre IA, ML y Deep Learning	40
<b>2.4 Modelos predictivos basados en IA</b>	<b>41</b>
2.4.1 La regresión lineal	42

	2.4.2	Redes neuronales artificiales	43
	2.4.3	Selección y entrenamiento de modelos	44
	<b>2.5</b>	<b>Aplicación de inteligencia artificial en la agroindustria</b>	<b>45</b>
	2.5.1	Casos de uso en predicción de calidad de alimentos	46
	2.5.2	Casos en frutas, jugos y control de calidad	47
	2.5.3	Desafíos y oportunidades del uso de IA en el sector	48
	<b>2.6</b>	<b>Evaluación del rendimiento de modelos</b>	<b>49</b>
F1)	2.6.1	Métricas de evaluación Métricas de evaluación (MSE, R <sup>2</sup> , precisión, recall, F1)	50
	2.6.2	La validación cruzada	51
	2.6.3	La interpretabilidad y la confiabilidad	52
	<b>2.7</b>	<b>Los sistemas predictivos integrados</b>	<b>53</b>
	2.7.1	Interfaz de usuario para sistemas predictivos	54
	2.7.2	Consideraciones técnicas para su implementación en planta	55
	2.7.3	Limitaciones tecnológicas en plantas agroindustriales	56
	<b>3</b>	<b>Capítulo 3. Marco Metodológico</b>	<b>58</b>
	<b>3.1</b>	<b>Tipo de investigación</b>	<b>58</b>
	<b>3.2</b>	<b>Alcance investigativo</b>	<b>59</b>
	<b>3.3</b>	<b>Enfoque</b>	<b>60</b>
	<b>3.4</b>	<b>Diseño</b>	<b>61</b>
	<b>3.5</b>	<b>Población y muestreo</b>	<b>62</b>
	<b>3.6</b>	<b>Instrumentos de recolección de datos</b>	<b>62</b>
	3.6.1	Documentación	63
	<b>3.7</b>	<b>Técnicas de análisis de información</b>	<b>63</b>
	<b>3.8</b>	<b>Diagrama de flujo</b>	<b>64</b>
	<b>4</b>	<b>Capítulo 4. Análisis del diagnóstico</b>	<b>65</b>
	<b>4.1</b>	<b>Revisión del proceso actual de control de calidad en frutas para jugo</b>	<b>65</b>
	<b>4.2</b>	<b>Análisis de los registros históricos de calidad</b>	<b>67</b>
	<b>4.3</b>	<b>Evaluación de la trazabilidad de los datos</b>	<b>69</b>
	<b>4.4</b>	<b>Limitaciones del sistema actual de evaluación</b>	<b>69</b>
	<b>4.5</b>	<b>Conclusiones del diagnóstico</b>	<b>70</b>
	<b>5</b>	<b>Capítulo 5 Desarrollo Del Modelo Predictivo</b>	<b>71</b>
	<b>5.1</b>	<b>Marco Teórico</b>	<b>73</b>
	5.1.1	Inteligencia Artificial en la Agroindustria	73
	5.1.2	Control de Calidad del Jugo Natural	73
	5.1.3	Importancia de las Características Fisicoquímicas	74
	5.1.4	Modelos Predictivos en Procesos Industriales	75
	5.1.5	Limitaciones Clave Para Considerar	75
	5.1.6	Fundamentación del Proyecto dentro de la Industria	76
	<b>5.2</b>	<b>Metodología</b>	<b>76</b>
	5.2.1	Enfoque y Tipo de Investigación	76

5.2.2	Diseño Metodológico	77
5.2.3	Población y Muestra	77
5.2.4	Recolección de Datos	77
5.2.5	Preprocesamiento de Datos	78
5.2.6	Modelos Predictivos Evaluados	78
5.2.7	Validación del Modelo	79
5.2.8	Métricas Utilizadas	79
5.2.9	Herramientas Tecnológicas	80
5.2.10	Cumplimiento del Objetivo Específico	80
<b>5.3</b>	<b>Resultados y Discusión</b>	<b>80</b>
5.3.1	Comparación visual: gráficas de barras y línea	81
5.3.2	Análisis de las matrices de confusión	83
5.3.3	Análisis PCA: Visualización en dos dimensiones	86
5.3.4	Heatmap de correlación	87
5.3.5	Discusión General	88
5.3.6	Modelo ganador: MLP (Red Neuronal Multicapa)	89
5.3.7	Relación entre clusterización y mejora del rendimiento	90
5.3.8	Implicaciones industriales de los resultados	90
<b>5.4</b>	<b>Conclusiones</b>	<b>91</b>
<b>5.5</b>	<b>Recomendaciones</b>	<b>92</b>
5.5.1	Arquitectura General del Sistema	94
5.5.2	Diseño de la Interfaz para Usuarios Técnicos	95
5.5.3	Flujo de Operación en Planta	96
5.5.4	Beneficios Industriales (Integrando archivo .txt)	96
5.5.5	Requisitos Técnicos para Implementación	97
5.5.6	Proyección a Futuro	97
<b>6</b>	<b>Capítulo VI: Conclusiones y Recomendaciones</b>	<b>98</b>
<b>6.1</b>	<b>Conclusiones</b>	<b>98</b>
6.1.1	Conclusiones Objetivo específico 1	99
6.1.2	Conclusiones Objetivo específico 2	99
6.1.3	Conclusiones Objetivo específico 3	100
6.1.4	Conclusiones Objetivo específico 4	100
6.1.5	Conclusiones Objetivo específico 5	101
<b>6.2</b>	<b>Recomendaciones</b>	<b>101</b>
<b>7</b>	<b>Referencias</b>	<b>104</b>
<b>8</b>	<b>Anexo</b>	<b>107</b>
	<b>Anexo 1</b>	<b>107</b>
	<b>Entrevista estructurada</b>	<b>107</b>

### Tabla de Figuras

Figura 1 F1-Score – Diagrama de Dispersión .....	81
Figura 2 F1-Score – Barras Verticales .....	81
Figura 3 F1-Score – Línea Comparativa .....	83
Figura 4 Matriz de Confusión – RandomForest .....	84
Figura 5 Matriz de Confusión – MLP .....	84
Figura 6 Matriz de Confusión – ExtraTrees .....	85
Figura 7 Matriz de Confusión – SVM-RBF .....	85
Figura 8 Matriz de Confusión – GradientBoosting .....	86
Figura 9 PCA 2D – Calidad Natural .....	86
Figura 10 Heatmap de correlación química .....	87

### Índice de Tablas

Tabla 1: Desglose de salarios .....	15
Tabla 2: Costo de horas consultor .....	15

### **Abstract**

This research presents a predictive system based on artificial intelligence (AI) to estimate the quality of natural juice prior to its final processing. Leveraging physicochemical features of the fruit, the system aims to improve decision-making in agro-industrial processes by providing early quality assessments. The project integrates advanced machine learning algorithms capable of analyzing multiple variables and uncovering complex relationships, thus enabling standardization of product quality, resource optimization, and reduction of waste. The development of such predictive models aligns with international and regional regulations, ensuring product safety and compliance. Overall, this AI-driven approach offers a competitive advantage by enhancing process efficiency, ensuring consistent product quality, and increasing market value in the dynamic agro-industrial sector.

## **1 Capítulo 1. Introducción**

### **1.1 Generalidades**

En las últimas décadas, el desarrollo de la tecnología ha experimentado cambios radicales, y aún más con la llegada de la Inteligencia Artificial, siendo esto un factor que marcó un antes y un después en muchas áreas de trabajo, como la agricultura y la industria. Espacios que antes parecían ser exclusivos para aspectos más académicos o sitios más científicos, hoy resultan más accesibles gracias a programas que poseen una mayor eficiencia, datos de fácil acceso y un poder de computación cada día mayor. Especialmente en sistemas de aprendizaje automático, que han permitido optimizar procesos, logrando predicciones y, aún más importante, tomar decisiones inteligentes basadas en gran cantidad de datos, algo que anteriormente parecía inalcanzable para la tecnología.

Dentro de un contexto más agroindustrial, este tipo de tecnologías, tan innovadoras, han logrado ser implementadas en una gran variedad de actividades, como la selección de cultivos, el monitoreo de calidad, la estimación del rendimiento de cultivos o jugos, e inclusive la detección de posibles enfermedades en las plantas. Sin embargo, uno de los mayores retos con que se enfrenta el mercado, día con día, es garantizar la calidad del producto final, dado que existen múltiples factores que podrían afectarla, especialmente en la producción de jugos naturales, donde aspectos como la madurez de la fruta, la coloración, la acidez, la exposición al sol, el exceso de lluvia, entre muchos otros factores, pueden influir negativamente en su calidad.

Es por ello que la integración de herramientas que trabajen con inteligencia artificial pueden permitir predecir la calidad del jugo desde antes de su procesamiento, siendo esto un factor crucial en la toma de decisiones. Este tipo de avances crea un nuevo marco de posibilidades para la industria, permitiendo anticipar posibles problemas futuros, reducir

múltiples desperdicios en el proceso o bien, lograr un rápido accionar de manera informada ante decisiones críticas, gracias a estas herramientas predictivas.

Este documento aborda el desarrollo de un sistema predictivo para el procesamiento de jugos y el uso de registros históricos de características fisicoquímicas de la fruta. Esto permitirá obtener un producto de mayor calidad, con menos variaciones, siendo esto un aporte de gran relevancia para el sector.

## **1.2 Definición y Descripción del Problema**

Históricamente, el control de calidad del jugo natural ha sido un proceso bastante laborioso, en el cual la evaluación del producto suele realizarse hasta después de haber sido procesado. Con el tiempo, y conforme la práctica y los diferentes desarrollos tecnológicos han avanzado, se han ido implementando diversos procesos evaluativos durante la producción, aspectos que poco a poco han contribuido al mejoramiento de la calidad del jugo final. Sin embargo, este enfoque puede llegar a ser mucho más eficiente y ofrecer mejores resultados mediante la implementación de un modelo predictivo.

Ahora bien, dentro del amplio marco de posibilidades que ofrecen los modelos predictivos para evaluar la calidad del jugo final, se encuentra la capacidad de analizar grandes cantidades de datos históricos, lo que mejora significativamente la toma de decisiones en tiempo real y, aún más importante, la eficiencia operativa, características que, en un entorno de negocio, permiten una mayor rentabilidad, lo que convierte a este tipo de herramientas en un recurso poderoso para el sector agroindustrial.

## **1.3 Definición y Descripción del Problema**

En la actualidad, la industria agroalimentaria enfrenta retos crecientes relacionados con el control de calidad, la estandarización de sus procesos y la eficiencia en el manejo de la materia prima, especialmente en la producción de jugo. Si bien se han recibido múltiples avances gracias a la tecnología, en muchos casos existe resistencia al cambio y dependencia de

métodos manuales o empíricos para la evaluación de la calidad de la fruta. Todo esto genera variaciones en el producto final, desperdicios, o incluso situaciones tan críticas como el descarte de un lote completo. Por lo tanto, este tipo de afectaciones son percibidas por muchos de los consumidores finales.

En este contexto, con el desarrollo de nuevas tecnologías innovadoras como la inteligencia artificial, se logra demostrar con hechos su gran capacidad para predecir de manera precisa distintos resultados. Gracias a herramientas como redes neuronales, árboles de decisión y modelos de aprendizaje supervisado, se convierte en un marco de desarrollo muy atractivo y plantea el reto de anticipar la calidad del jugo a partir de características fisicoquímicas, evitando así tantas pérdidas y logrando mejores estándares de calidad, con menos variaciones, lo que genera un consumidor final más satisfechos.

#### **1.4 Justificación**

La industria alimentaria, enfrenta desafíos constantes día con día debido a la variabilidad de la materia prima, influenciada por factores climáticos y biológicos que afectan directamente la calidad del producto final. En el caso de los jugos naturales, se consideran parámetros como la relación brix<sup>1</sup> y la acidez, los cuales determinan aspectos esenciales como el sabor, el color o el aroma. Sin embargo, estos factores suelen evaluarse de forma empírica o tardía, generando pérdidas o reprocesos. Por ello, un modelo predictivo basado en inteligencia artificial que, mediante el análisis de datos fisicoquímicos históricos, permita anticipar la calidad del producto antes del procesamiento, representa una propuesta viable, innovadora y útil, tanto a nivel industrial como académico, ya que mejora la toma de decisiones, reduce el desperdicio y aporta una nueva visión al control de calidad en alimentos.

---

<sup>1</sup> Brix es una unidad de medida para determinar el contenido de sólidos disueltos, principalmente azúcares, en un líquido, expresado como gramos de sacarosa por cada 100 gramos de solución

En definitiva, esta investigación resulta adecuada para el propósito de ofrecer soluciones concretas a un problema actual de la industria del jugo natural. Desde un enfoque social, permite avanzar hacia una producción más estandarizada, con mejores características nutricionales, lo cual representa un beneficio directo para el consumidor. Además, aporta un valor teórico al integrar conocimientos tecnológicos, como la inteligencia artificial, y sus aplicaciones en el análisis predictivo. Finalmente, desde una perspectiva metodológica, este tipo de enfoque permite ser aplicado o adaptado a otros productos del mercado agroindustrial, modificando ciertas variables según el contexto, pero manteniendo su impacto y utilidad en cada implementación

## **1.5 Viabilidad**

La viabilidad del proyecto se basa en la disponibilidad de los datos fisicoquímicos de las frutas y en los avances que actualmente ofrece la inteligencia artificial, además del interés creciente del sector agroindustrial por mejorar la calidad de sus productos y maximizar su rentabilidad. Asimismo, la implementación de sistemas que integran algoritmos de aprendizaje automático permite anticipar la calidad del jugo natural antes de su procesamiento, lo cual representa una ventaja competitiva en el mercado. Desde un punto de vista más técnico, el proyecto resulta factible gracias a la existencia de herramientas de código abierto, el conocimiento disponible en internet y el apoyo de profesores que facilitan el diseño, entrenamiento y procesamiento de datos, garantizando así una implementación eficiente. Finalmente, la viabilidad económica se sustenta en la reducción del desperdicio y en la mejora de la toma de decisiones en la cadena productiva, contribuyendo a factores como la rentabilidad y la sostenibilidad del proceso.

### ***1.5.1 Punto de Vista Técnico.***

Desde un enfoque técnico, el proyecto posee una muy buena perspectiva, ya que las herramientas que actualmente ofrece el mercado facilitan la implementación de algoritmos de inteligencia artificial, como por ejemplo la regresión lineal, las redes neuronales o el algoritmo Random Forest; todos han demostrado ser herramientas eficaces para realizar predicciones basadas en datos. Por otro lado, este tipo de modelos se puede entrenar con base en el tipo de fruto y sus características fisicoquímicas, lo cual permite generar modelos predictivos con la capacidad de anticipar, con un alto grado de precisión, la calidad esperada del producto final antes de su procesamiento, o incluso antes de ingresar a una fábrica donde se implementan métodos de mezclado para lograr un producto estándar.

Este tipo de estrategias ya han sido implementadas en algunos ámbitos agroindustriales, por ejemplo, para predecir la calidad del café, evaluar la calidad enológica del vino, o determinar el punto óptimo de madurez y dulzura en frutas como el tomate o el mango. Todo esto refuerza la factibilidad técnica de esta propuesta aplicada al jugo de naranja. En definitiva, su correcta implementación requiere de un diseño metodológico que contemple la recolección de datos fisicoquímicos confiables, la definición de las variables más representativas para determinar la calidad del jugo y la selección del modelo predictivo más adecuado. Asimismo, se deben considerar los factores que influyen en todo este proceso y que, al final, repercuten directamente en la calidad del producto.

### ***1.5.2 Punto de Vista Operativo***

La realización del proyecto se diseña de manera flexible para adaptarse a los conocimientos y aprendizajes que se obtienen. Se prioriza el uso de recursos accesibles y se evita depender en exceso de infraestructuras complejas o costosas, evitando incurrir en procesos que realmente no generen valor.

Parte del planeamiento operativo recae en la gestión del tiempo a través de un cronograma realista que permite dedicar el tiempo necesario para el desarrollo del proyecto, además de la utilización de herramientas digitales de la actualidad y recursos de bibliotecas universitarias que proporcionan material didáctico especializado en el área de modelos de predicción e inteligencia artificial.

### **1.5.3 Punto de Vista Económico**

El desarrollo del proyecto de un sistema predictivo en inteligencia artificial para anticipar la calidad de jugo natural, a partir de características fisicoquímicas, implica ciertos costos asociados, de los cuales principalmente son aspectos relacionados con horas de investigación, licencia de software, hardware y otros gastos relacionados, los cuales son asumidos íntegramente por los autores de esta investigación.

Para ello, se ha tomado como referencia la información salarial proporcionada por Glassdoor para un investigador en Costa Rica. Con esta fuente, el salario mensual promedio para un investigador ronda los ₡634000, lo que corresponde a un salario anual de ₡7608000.

Por ende, el desarrollo del proyecto requiere aproximadamente 24 horas de trabajo tomando como referencia el salario del percentil 25 al 75 de los salarios en Costa Rica.

**Tabla 1: Desglose de salarios**

Salario Anual	Salario Mensual	Salario Diario	Salario por hora
7608000	634000	20843	2605

Fuente: Elaboración propia basada en la información de Glassdoor para un investigador en Costa Rica.

**Tabla 2: Costo de horas consultor**

Horas Investigación	Horas Semanales	Valoración TGF Meses	Valoración TGF Horas	Costo estimado por hora
Luis David Chavarría	2	3	24	6164

Fuente: Elaboración propia basada en la información de Glassdoor para un consultor en Costa Rica.

Con esta valoración, se pueden calcular los fondos requeridos para llevar a cabo el proyecto, mostrando que el gasto que implica es asequible y práctico en nuestro entorno, lo cual consolida la viabilidad económica de lo que se propone.

## **1.6 Objetivos**

La taxonomía de Bloom se implementó en la definición de objetivos de investigación debido a su facilidad de uso, así como a su amplia adopción en el ámbito investigativo, lo que permite estructurar los objetivos de manera clara, jerárquica y alineada con los niveles de pensamiento cognitivo que se buscan desarrollar.

### **1.6.1 Objetivo General.**

Diseñar un sistema predictivo basado en inteligencia artificial que permita anticipar la calidad del jugo natural antes del procesamiento, mediante el análisis de características fisicoquímicas, durante el periodo comprendido entre marzo y diciembre de 2025.

### **1.6.2 *Objetivos Específicos.***

Identificar las principales características fisicoquímicas de las frutas que mayormente inciden en la calidad del jugo natural, mediante una revisión de literatura científica, artículos técnicos y estándares del sector agroindustrial.

Explicar la relación que existe entre diversas variables fisicoquímicas y la calidad del jugo natural, analizando estudios previos, correlaciones estadísticas y registros experimentales sobre acidez, pH, entre otros.

Desarrollar un prototipo funcional de modelo predictivo utilizando algoritmos de inteligencia artificial, como redes neuronales, regresión o Random Forest, entrenado con datos históricos recolectados de análisis fisicoquímicos.

Comparar el comportamiento y precisión de distintos algoritmos de predicción aplicados sobre el conjunto de datos fisicoquímicos, evaluando métricas como el error cuadrático medio,  $R^2$  y exactitud.

Diseñar un sistema predictivo integral que incorpore el modelo desarrollado en una interfaz sencilla para usuarios técnicos, considerando su implementación práctica en entornos industriales de procesamiento de jugo.

### **1.7 Alcances y Limitaciones**

Cada investigación científica conlleva ciertas condiciones que definen sus limitaciones con base en los objetivos que se propone alcanzar. Por ende, es crucial establecer claramente cuáles son los alcances y limitaciones del proyecto, permitiendo así delimitar el campo de acción del sistema propuesto y, con ello, las condiciones bajo las cuales sus resultados serán considerados válidos y fiables.

En esta sección se describen los objetivos que se buscan alcanzar con la creación del sistema predictivo basado en inteligencia artificial. Asimismo, se exponen las posibles restricciones que podrían afectar la precisión del modelo, su uso práctico o su implementación efectiva.

### ***1.7.1 Alcances.***

El proyecto pretende diseñar y desarrollar un sistema predictivo basado en inteligencia artificial que permita anticipar la calidad del jugo natural antes de su procesamiento final. Para ello, se implementará el uso de las principales características fisicoquímicas de la fruta. Así mismo, el sistema tendrá la capacidad de anticipar la calidad del jugo y, con ello, sugerir al usuario un conjunto de acciones operativas que le permitan obtener un producto estándar de alta calidad.

El modelo será entrenado con datos históricos recolectados a lo largo de los años en un entorno de procesamiento de jugo a nivel industrial. De esta manera, se espera que brinde un aporte significativo en la toma de decisiones técnicas, la reducción de reprocesos y la disminución de desperdicios.

### ***1.7.2 Limitaciones***

Las limitaciones de esta investigación incluyen, en primer lugar, la dependencia de la disponibilidad, calidad y cantidad de los datos fisicoquímicos de cada fruta, los mismos que serán utilizados para el entrenamiento y validación del modelo predictivo, lo cual podría llegar a afectar su precisión. Asimismo, el sistema se enfoca únicamente en parámetros

fisicoquímicos, por lo que cualquier otra consideración podría verse afectada, como lo serían las condiciones de transporte, clima, almacenamiento o bien la manipulación posterior.

Además, la predicción del modelo está ligada al tipo de fruta y a las variables incluidas en el conjunto de datos utilizado; por ello, podrían requerirse ajustes en productos que no hayan sido contemplados inicialmente.

Por lo tanto, es posible que el modelo necesite actualizaciones y recalibraciones periódicas conforme cambien las condiciones de cultivo, clima, logística de transporte o disponibilidad de personal de cosecha. En definitiva, en frutas con alta sensibilidad al clima o con ciclos de maduración irregulares, será necesario prestar mayor atención, y podría requerirse la incorporación de variables adicionales como el tiempo de almacenamiento, la región de origen o la variedad específica.

En conclusión, aunque se llegue a diseñar una interfaz de usuario técnica, su adopción y eficacia dependerán de una capacitación adecuada, así como de un entorno tecnológico favorable para lograr eventualmente su implementación.

## **1.8 Estado de la cuestión**

En años recientes, el sector de agroindustrial ha mostrado un creciente interés por adoptar soluciones de inteligencia artificial, con el propósito de mejorar la calidad de sus productos, afinar sus métodos de producción y, además, disminuir el desperdicio no deseado de insumos. En este contexto, resulta fundamental considerar que las características fisicoquímicas de las frutas, tales como la concentración de Brix, la acidez, el color, la vitamina C, el sabor, la madurez y compuestos amargos como la limonina, se encuentran estrechamente relacionadas tanto con la percepción de calidad como con el valor nutricional del jugo natural (USDA, 2021; Arzate-Vázquez et al., 2021).

A lo largo de los últimos años, diversos investigadores han aportado evidencia significativa en torno a esta problemática. Por ejemplo, Li, Xu, Zeng y Wang (2020) han examinado de forma exhaustiva el uso de la inteligencia artificial en la evaluación de productos agrícolas, destacando cómo los algoritmos de aprendizaje automático permiten identificar y predecir atributos clave de calidad. Asimismo, estudios como el de Ruiz-Altisent et al. (2010) han documentado la aplicación de sensores y técnicas de visión artificial para el análisis de características internas de las frutas. No obstante, es importante señalar que estos enfoques se han centrado principalmente en productos frescos, dejando de lado los derivados procesados como el jugo.

En consecuencia, pese a los avances tecnológicos mencionados, persiste una brecha relevante en la implementación de modelos predictivos capaces de anticipar la calidad del jugo natural antes de su procesamiento. Esta limitación se agrava al considerar que muchas investigaciones priorizan datos recolectados en etapas posteriores a la cosecha, omitiendo el potencial que tienen las variables fisicoquímicas obtenidas en fases iniciales como la precosecha o el almacenamiento. Por tanto, se hace evidente la necesidad de explorar nuevas aproximaciones que permitan intervenir tempranamente en el proceso agroindustrial.

Desde esta perspectiva, el desarrollo de sistemas inteligentes que integren inteligencia artificial y análisis fisicoquímico representa una oportunidad estratégica para transformar la gestión de calidad en la producción de jugo natural. A pesar de ser una línea de investigación emergente, su potencial es considerable, ya que permitiría mejorar la eficiencia, reducir pérdidas y asegurar productos con mejores propiedades organolépticas y nutricionales. En definitiva, este vacío justifica la presente investigación, orientada al diseño de un modelo predictivo que optimice la toma de decisiones desde las primeras etapas del proceso, aportando valor tanto técnico como científico al campo agroindustrial.

### **1.8.1 Formulación de la pregunta.**

Plantear la pregunta de investigación en un proyecto de estas características no es un trámite menor: es el cimiento sobre el que se construye toda la metodología, el análisis y la validez de los resultados. En esta investigación, orientada a diseñar un sistema predictivo basado en inteligencia artificial para anticipar la calidad del jugo natural, la formulación de la pregunta parte de una reflexión necesaria sobre las brechas detectadas entre la práctica actual en la industria agroindustrial y las posibilidades que brindan los avances tecnológicos.

Li et al. (2020) subrayan que la inteligencia artificial aplicada en agroindustria aún se concentra en productos frescos, sin abordar a fondo los derivados procesados como los jugos, lo cual genera un vacío metodológico relevante para esta investigación. En efecto, la incapacidad para predecir la calidad del jugo antes del procesamiento tiene consecuencias directas en la eficiencia operativa, en el desperdicio de materia prima y en la estandarización de los productos terminados (Lu et al., 2020).

Por ello, la pregunta de investigación se formula así:

¿Cómo diseñar un sistema predictivo basado en inteligencia artificial que, a partir de las características fisicoquímicas de la fruta, anticipe la calidad del jugo natural antes del procesamiento en un entorno agroindustrial?

Este planteamiento busca resolver una problemática concreta, delimitando las variables principales (características fisicoquímicas y calidad del jugo), el alcance tecnológico (modelos de IA) y el contexto de aplicación (procesamiento agroindustrial). Además, responde a la necesidad de contar con decisiones oportunas para mitigar pérdidas económicas y mejorar la calidad percibida por los consumidores (Pandey et al., 2023).

En otras palabras, la pregunta no sólo explora si la predicción es posible, sino cómo implementarla con base en modelos confiables, considerando las restricciones técnicas y operativas típicas de la agroindustria costarricense. Esto abre la puerta a un enfoque práctico con impacto medible, tal como demandan las tendencias actuales en la gestión inteligente de alimentos (Lu et al., 2020).

#### ***1.8.1.1 Foco de la pregunta***

El foco de la pregunta se centra en la integración eficiente de las capacidades predictivas de la inteligencia artificial con los datos fisicoquímicos preexistentes en la industria. Este enfoque responde a un contexto donde la calidad final del jugo depende de múltiples factores, muchos de ellos medibles con precisión antes del procesamiento (Rubio-Ochoa et al., 2019).

A diferencia de investigaciones previas que se limitan a describir correlaciones entre madurez, acidez o dulzor de la fruta con la calidad sensorial del jugo, aquí la atención se concentra en convertir esas correlaciones en un modelo predictivo robusto que pueda ser aplicado de manera sistemática. Según Pandey et al. (2023), esto supone diseñar algoritmos que no sólo “aprendan” de los patrones históricos, sino que sean capaces de adaptarse a las variaciones propias del ambiente agroindustrial.

El foco también está en la etapa previa al procesamiento, donde todavía es viable tomar decisiones que eviten pérdidas o ajustes costosos posteriores. Se trata de interceptar los errores antes de que el jugo sea producido, garantizando un producto más homogéneo y competitivo, alineado con estándares internacionales como los del Codex Alimentarius (Codex Alimentarius Commission, 2005).

Así, la pregunta delimitada y su foco articulan una investigación que no se contenta con entender el fenómeno, sino que pretende intervenirlo, optimizando recursos y elevando los estándares de calidad. En resumen, se orienta a responder de manera práctica a una necesidad

latente en la industria agroindustrial local y global, sin perder de vista los principios técnicos ni las restricciones económicas.

### ***1.8.1.2 Selección de fuentes.***

La selección de fuentes en un trabajo investigativo no es una tarea mecánica ni un simple inventario de lecturas; es, más bien, un ejercicio consciente y crítico, como afinar un instrumento antes de interpretar una sinfonía. Elegir adecuadamente lo que se lee y se cita define el contorno de la investigación, le da peso y dirección, y evita que el análisis quede a merced de afirmaciones ligeras. Como enfatizan Pandey et al. (2023), la calidad del conocimiento producido depende, en buena medida, de la solidez metodológica y documental que se emplee desde su origen. Así, el proceso de selección de fuentes se convierte en un filtro que asegura que las ideas que nutren esta investigación provengan de estudios serios, recientes y verificables.

En este proyecto, enfocado en un sistema predictivo basado en inteligencia artificial para evaluar la calidad del jugo natural, se priorizó la selección de fuentes académicas con rigor científico y con evidencia empírica concreta, provenientes de revistas indizadas y bases de datos reconocidas internacionalmente. Esta selección tiene un doble propósito: sustentar con solidez las decisiones metodológicas y al mismo tiempo evitar sesgos o interpretaciones sesgadas que podrían distorsionar las conclusiones (Lu et al., 2020).

### ***1.8.1.3 Definición de criterios de selección de fuentes.***

#### ***1.8.1.3.1 Idiomas de los estudios.***

La diversidad lingüística de las fuentes constituye una ventaja competitiva en la investigación. En la práctica, sin embargo, se estableció como criterio central trabajar con

artículos en inglés y español, los dos idiomas en los que se concentra la producción científica más actual y relevante sobre inteligencia artificial aplicada a la agroindustria (Pandey et al., 2023).

Aunque se revisaron algunos abstracts en otros idiomas, la disponibilidad limitada de traducciones confiables, y la necesidad de evitar malinterpretaciones, llevó a restringir la búsqueda a estas lenguas.

Esta decisión no fue arbitraria, sino pragmática. El inglés, por ser la lengua franca de la ciencia contemporánea, concentra la mayoría de las publicaciones indizadas de alta calidad (Lu et al., 2020). El español, por su parte, aporta un matiz regional invaluable en cuanto a estudios realizados en Latinoamérica, que resultan pertinentes por sus similitudes climáticas y socioeconómicas con el contexto costarricense (Rubio-Ochoa et al., 2019).

En definitiva, la selección de idiomas buscó equilibrar la amplitud del conocimiento con la claridad y la precisión necesarias para fundamentar este trabajo.

#### ***1.8.1.3.2 Identificación de fuentes.***

La identificación de las fuentes relevantes fue una tarea sistemática. Se utilizaron motores de búsqueda académicos y bases de datos confiables como IEEE Xplore, Scopus, Web of Science, Google Scholar y SciELO. Esta búsqueda incluyó términos controlados como *predictive models*, *fruit quality*, *artificial intelligence in agroindustry*, *physicochemical properties*, y sus equivalentes en español.

Como recomienda Codex Alimentarius Commission (2005), identificar las fuentes adecuadas implica no solo recopilar documentos sino evaluar su pertinencia frente a los estándares internacionales y las realidades locales. Cada fuente seleccionada fue evaluada para confirmar su congruencia con los objetivos de la investigación, su sustento empírico y la claridad de sus resultados.

La búsqueda inicial arrojó una gran cantidad de documentos, pero un proceso de cribado posterior permitió quedarse únicamente con los estudios más significativos y pertinentes, siguiendo las mejores prácticas metodológicas descritas por Li et al. (2020). Este proceso fue tan deliberado como cuando un productor elige a mano las mejores frutas: un ejercicio de discernimiento paciente y riguroso que no tolera improvisaciones.

#### ***1.8.1.3.3 Criterios de selección de estudios***

La definición de criterios para seleccionar los estudios no es un asunto accesorio, sino un mecanismo esencial para garantizar que la evidencia que sustente esta investigación sea coherente, relevante y de calidad. En términos prácticos, se establecieron criterios inclusivos y excluyentes que funcionaron como un tamiz fino, separando el conocimiento útil del ruido metodológico.

Los criterios inclusivos priorizaron investigaciones publicadas entre 2019 y 2025, preferentemente revisadas por pares y publicadas en revistas indexadas, asegurando así la actualidad y validez del conocimiento científico (Li et al., 2020). En segundo lugar, los estudios debían presentar resultados empíricos sobre la relación entre características fisicoquímicas de frutas y la calidad del jugo natural, o bien describir aplicaciones concretas de inteligencia artificial en contextos agroindustriales similares. Además, como enfatiza Lu et al. (2020), se privilegió el uso de modelos predictivos con suficiente documentación metodológica para poder evaluar su replicabilidad.

Por otra parte, se excluyeron estudios duplicados, artículos de opinión, informes sin revisión académica, tesis sin respaldo institucional visible y publicaciones que abordaran únicamente aspectos sensoriales o nutricionales de los jugos sin ligarlos explícitamente a procesos predictivos (Pandey et al., 2023). También quedaron fuera los documentos

excesivamente generales que no aportaran evidencia concreta aplicable al diseño de un modelo predictivo, pues su inclusión habría debilitado la fuerza argumentativa de este trabajo.

Este criterio equilibrado entre amplitud temática y profundidad metodológica permitió mantener la investigación en una línea rigurosa pero práctica, evitando caer en extremos: ni demasiado restrictiva que sofocara el análisis ni tan laxa que desdibujara la relevancia del proyecto (Rubio-Ochoa et al., 2019).

#### ***1.8.1.3.4 Selección de fuentes después de la evaluación***

La selección final de las fuentes no fue inmediata; requirió un proceso deliberado y metódico en el que se evaluaron cada una de las referencias bajo los criterios previamente definidos. Este ejercicio se realizó en varias etapas. Primero, se identificaron decenas de estudios relevantes por título, palabras clave y resumen. Luego, se revisó el contenido completo para confirmar si cumplía con los parámetros de actualidad, rigor y aplicabilidad al contexto agroindustrial.

Li et al. (2020) recomiendan que, durante esta fase, el investigador mantenga un balance entre diversidad y especificidad, lo que garantiza una visión integral sin diluir los objetivos principales. En esa línea, se descartaron varios artículos que, aunque interesantes, no ofrecían evidencia concreta para el problema planteado. Asimismo, se consolidaron aquellos que presentaban datos verificables y metodologías claras, como los estudios de Pandey et al. (2023) sobre algoritmos de aprendizaje automático en preservación de frutas y los análisis de Lu et al. (2020) sobre inteligencia artificial en ciencia de alimentos.

Al concluir este tamiz, la lista final quedó conformada por un conjunto selecto de estudios que no sólo respaldan la investigación, sino que también la enriquecen con ejemplos

empíricos, estándares normativos y enfoques metodológicos contrastados. Este proceso no sólo garantizó coherencia y solidez, sino que también fue una declaración de intenciones: aquí no hay cabida para improvisaciones ni para citas decorativas, sino para referencias que suman y sostienen la credibilidad del proyecto.

#### ***1.8.1.3.5 Comprobación de referencias***

La comprobación de las referencias es un paso que muchas veces se subestima, pero que en realidad constituye uno de los mecanismos más efectivos para salvaguardar la integridad académica y la credibilidad de la investigación. No basta con anotar títulos atractivos y autores reconocidos; hay que asegurarse de que cada fuente efectivamente exista, sea accesible y cumpla con los estándares exigidos para una investigación científica.

En esta investigación, la comprobación se realizó verificando manualmente cada referencia seleccionada en sus repositorios originales ya fuera mediante el enlace DOI, las páginas oficiales de las editoriales o las bases de datos institucionales. Tal como recomiendan Lu et al. (2020), se descartaron aquellas citas que no presentaran trazabilidad clara o cuya publicación no pudiera validarse en un entorno académico confiable.

Este proceso no sólo garantizó que cada dato citado tenga respaldo real y verificable, sino que también ayudó a identificar posibles errores de formato o inconsistencias menores en los metadatos, corrigiéndose antes de consolidar la lista final de referencias (Li et al., 2020). Así, se asegura que el lector y el evaluador académico puedan confirmar la autenticidad y exactitud de cada fuente con la misma facilidad con la que se revisa una huella digital.

#### ***1.8.1.3.6 Selección de estudios.***

La comprobación de las referencias es un paso que muchas veces se subestima, pero que en realidad constituye uno de los mecanismos más efectivos para salvaguardar la integridad

académica y la credibilidad de la investigación. No basta con anotar títulos atractivos y autores reconocidos; hay que asegurarse de que cada fuente efectivamente exista, sea accesible y cumpla con los estándares exigidos para una investigación científica.

En esta investigación, la comprobación se realizó verificando manualmente cada referencia seleccionada en sus repositorios originales ya fuera mediante el enlace DOI, las páginas oficiales de las editoriales o las bases de datos institucionales. Tal como recomiendan Lu et al. (2020), se descartaron aquellas citas que no presentaran trazabilidad clara o cuya publicación no pudiera validarse en un entorno académico confiable.

Este proceso no sólo garantizó que cada dato citado tenga respaldo real y verificable, sino que también ayudó a identificar posibles errores de formato o inconsistencias menores en los metadatos, corrigiéndose antes de consolidar la lista final de referencias (Li et al., 2020). Así, se asegura que el lector y el evaluador académico puedan confirmar la autenticidad y exactitud de cada fuente con la misma facilidad con la que se revisa una huella digital.

#### ***1.8.1.3.7 Definición de estudios***

Por definición, en esta investigación los estudios seleccionados corresponden a trabajos científicos publicados en revistas indexadas o documentos normativos reconocidos internacionalmente que analicen:

Las características fisicoquímicas de las frutas y su relación con la calidad del jugo.

El desarrollo e implementación de modelos predictivos basados en inteligencia artificial en contextos agroindustriales.

Normativas y estándares de calidad alimentaria vigentes en la industria.

Li et al. (2020) destacan que, para un estudio predictivo basado en IA, los estudios de referencia no solo deben describir correlaciones simples, sino también detallar métodos de modelado y validación de algoritmos. Por tanto, se priorizaron aquellos trabajos que no solo

documentan la relación entre variables, sino que además proponen o evalúan soluciones predictivas.

#### ***1.8.1.3.8 Ejecución de la revisión.***

La ejecución de la revisión consistió en aplicar, de manera sistemática, los filtros y criterios definidos a toda la literatura recopilada durante la fase de búsqueda. Este proceso fue tanto analítico como crítico, evaluando no solo la forma en que los estudios están escritos, sino la validez de sus resultados y la aplicabilidad de sus métodos.

Como recomienda Lu et al. (2020), la revisión se realizó en etapas, comenzando con la exclusión de duplicados, pasando por la verificación de calidad metodológica, y culminando con la confirmación de relevancia específica para los objetivos de este proyecto. Este procedimiento no solo ayudó a consolidar la lista final de estudios relevantes, sino que también facilitó la identificación de vacíos de conocimiento en los que este trabajo pretende aportar.

#### ***1.8.1.3.9 Ejecución de la fuente IEEE digital Library.***

Dentro de las bases de datos exploradas, la IEEE Digital Library tuvo un papel relevante en la identificación de estudios técnicos sobre inteligencia artificial aplicada a sistemas predictivos. La búsqueda se configuró empleando términos específicos en inglés como *predictive models*, *machine learning in food industry*, *fruit quality prediction*, y *artificial intelligence in agriculture*.

Este filtrado inicial arrojó numerosos documentos, pero solo una fracción cumplió con los criterios definidos de actualidad, rigor metodológico y aplicabilidad práctica. Tal como señalan Pandey et al. (2023), la selección en bases técnicas como IEEE permite acceder a

investigaciones con alto componente tecnológico, esenciales para fundamentar las decisiones sobre los algoritmos que se implementarán.

El análisis crítico de los resultados de IEEE se integró al resto de la revisión sistemática para asegurar que la investigación no pierda de vista las mejores prácticas ni las innovaciones más recientes del ámbito técnico.

#### ***1.8.1.3.10 Ejecución de la selección de la fuente IEEE digital Library.***

La ejecución de la búsqueda y selección en la IEEE Digital Library fue una etapa clave para robustecer la base técnica y metodológica de esta investigación, dado que esta plataforma concentra publicaciones de alta calidad sobre desarrollos recientes en inteligencia artificial, aprendizaje automático y sistemas predictivos.

La estrategia de búsqueda se diseñó utilizando una combinación de palabras clave en inglés, con operadores booleanos para refinar los resultados. Se usaron términos como *"predictive model" AND "fruit quality"*, *"machine learning" AND "food industry"*, *"artificial intelligence" AND "agriculture"*, y *"physicochemical properties" AND "prediction"*. Esta configuración permitió acotar el universo de resultados a artículos directamente vinculados con la temática central del proyecto.

De los resultados iniciales que superaban los 150 artículos publicados entre 2019 y 2025 se realizó una primera depuración descartando aquellos que no cumplían con los criterios definidos previamente, como la pertinencia temática o la disponibilidad completa del documento. Según Lu et al. (2020), es esencial que los estudios seleccionados no sólo describan técnicas innovadoras, sino que además presenten validación empírica y resultados aplicables a entornos reales. Por esta razón, se descartaron revisiones excesivamente generales, resúmenes

de conferencias sin datos suficientes y documentos centrados en ámbitos alejados de la agroindustria.

El proceso continuó con una lectura analítica de los artículos preseleccionados, evaluando aspectos como: claridad en la metodología, tipo de datos utilizados, algoritmos implementados y resultados reportados. Esta lectura crítica permitió elegir únicamente aquellos trabajos que aportaran un valor tangible para el desarrollo de un sistema predictivo basado en IA en un contexto agroindustrial, alineados con la problemática planteada en esta investigación.

En palabras de Pandey et al. (2023), esta etapa representa un punto de inflexión entre una búsqueda documental superficial y un ejercicio verdaderamente riguroso de fundamentación teórica. La selección final incluyó artículos que no sólo refuerzan el marco teórico, sino que además ofrecen referencias prácticas para la implementación técnica del modelo, asegurando que la propuesta se sustente en experiencias previas documentadas y verificables.

En definitiva, la ejecución de la selección en IEEE Digital Library fue un proceso meticuloso, deliberado y guiado por los principios de relevancia, actualidad y aplicabilidad, que permitió enriquecer la investigación con aportes sólidos y contemporáneos provenientes de una de las fuentes más prestigiosas en materia tecnológica.

#### ***1.8.1.3.11 Ejecución de la selección de la fuente Google Scholar.***

La ejecución de la búsqueda en **Google Scholar** complementó la revisión sistemática al ampliar el espectro de estudios potencialmente relevantes, particularmente aquellos con un enfoque aplicado en contextos agroindustriales y normativos. Aunque esta herramienta incluye una gran cantidad de publicaciones no siempre indizadas en bases técnicas como IEEE, su alcance global y su accesibilidad permiten localizar estudios significativos con un enfoque más amplio y, a veces, con mayor pertinencia regional.

Para garantizar la calidad de los resultados, la estrategia de búsqueda se diseñó cuidadosamente, utilizando combinaciones específicas de palabras clave y operadores booleanos en inglés y español. Entre las combinaciones más utilizadas destacaron: “*modelo predictivo calidad jugo natural*”, “*machine learning agroindustria*”, “*propiedades fisicoquímicas predicción jugo*”, y “*inteligencia artificial frutas*”.

Dado que Google Scholar devuelve resultados con diferente grado de rigor, la primera etapa consistió en filtrar cronológicamente los documentos, limitando la búsqueda a publicaciones entre 2019 y 2025, tal como lo recomiendan Li et al. (2020), para mantener la actualidad tecnológica y científica. Posteriormente, se evaluó que cada documento contara con DOI o, al menos, con trazabilidad verificable, eliminando blogs, documentos sin revisión por pares o informes con fuentes no identificables.

La lectura crítica de los resúmenes y conclusiones permitió descartar artículos centrados exclusivamente en aspectos sensoriales o en procesos ajenos a la predicción temprana de calidad. Se priorizaron aquellos trabajos que documentaran aplicaciones de inteligencia artificial en el control de calidad, el análisis de variables fisicoquímicas y la validación de modelos predictivos (Pandey et al., 2023).

Además, como señalan Lu et al. (2020), las búsquedas en plataformas más abiertas como Google Scholar requieren mayor cautela, ya que su inclusividad implica un riesgo de sesgos si no se aplica una criba rigurosa. Por esa razón, cada referencia seleccionada fue contrastada con las demás bases de datos para asegurar consistencia y autenticidad.

En síntesis, la búsqueda en Google Scholar aportó valor al permitir identificar estudios de aplicación regional y normativas que, de otro modo, habrían pasado desapercibidas en repositorios más especializados. Este ejercicio reforzó la fundamentación de la investigación al integrar perspectivas diversas sin sacrificar calidad ni rigor metodológico.

## 2 Capítulo 2. Marco Conceptual

Se genera la siguiente nube de palabras con términos que resaltan en los diversos textos que conforman la revisión bibliográfica de este proyecto. Vale la pena señalar que, para elaborar este gráfico, se utilizaron versiones resumidas de los artículos originales que estaban en inglés, incorporando así la información con los conceptos clave para este proyecto.

**Figura 1: Nube de palabras generadas, a partir del estado de la cuestión.**



Nota: Elaboración propio, a partir de los principales conceptos de los modelos predictivos de empresas agroindustriales

## **2.1 Concepto de calidad del jugo natural**

La calidad del jugo natural, sin lugar a duda, es un aspecto crucial para mantener un producto óptimo, garantizando así su seguridad y aceptación a nivel de producto de mercado. Es así como la Comisión del Codex Alimentarius (2005) fija las normas internacionales para jugos y néctares, y define la calidad del producto a partir de propiedades físicas, químicas y aspectos microbiológicos, que buscan asegurar tener un producto adecuado para el consumo humano y que, con ello, cumpla con las expectativas de la gente.

Asimismo, dentro de las normas establecidas para la calidad del jugo natural, se pueden detallar las principales:

**Composición y pureza:** En esta implicación, el jugo debe componerse únicamente del líquido obtenido de la fruta, sin llegar a añadir ninguna otra sustancia o esencia que altere su naturalidad, a excepción de algunos aditivos permitidos bajo reglas estrictas. Asimismo, se incluyen parámetros como el contenido mínimo de sólidos solubles, que muestran cuánta azúcar y sólidos hay en el jugo, aspecto clave para el sabor y su consistencia.

**Características fisicoquímicas:** Se miden parámetros como el pH, el nivel de acidez, el contenido de azúcares, los pigmentos naturales y la vitamina C. Estos aspectos se analizan para controlar la calidad y asegurar que el jugo se conserve bien.

**Ausencia de contaminantes y aditivos no permitidos:** Estas normas controlan los límites máximos de contaminantes y aditivos que pueden estar presentes, asegurando que el producto no cause daño.

**Autenticidad y métodos analíticos:** Las normas utilizan métodos probados para detectar adulteraciones y falsificaciones, asegurando con esto que el jugo sea realmente de la fruta que dice ser y que no se le haya añadido ninguna sustancia externa.

Es así como la calidad del jugo natural, según el Codex Alimentarius, va más allá del sabor: abarca la seguridad y el cumplimiento de normas, generando así un marco completo para la producción y comercialización de jugos naturales.

### ***2.1.1 Factores que influyen en la calidad del jugo***

Para la obtención de un jugo de calidad, no solo basta con tener una buena fruta o un sabor agradable, sino que existe una gran cantidad de factores directos que influyen en cómo será el producto final. Dichos factores se pueden clasificar como ambientales, agrícolas y tecnológicos; todos ellos influyen en cómo será el jugo, los nutrientes que contendrá y si será seguro para el consumidor final.

En primera instancia, el factor ambiental y de almacenamiento contempla aspectos como la temperatura a la que se somete el producto, la duración del transporte o las condiciones higiénicas durante su movilización. Por ello, el adecuado manejo de la cadena de frío resulta un aspecto fundamental. Dado que una interrupción en esta puede alterar la composición microbiológica, lo cual provocaría que un producto no cumpla con los estándares de calidad.

Luego, en el factor agrícola, se consideran todos los aspectos ligados al fruto, desde el tipo de cultivo, la variedad, las condiciones climáticas, las fórmulas de aplicación y la etapa de madurez al momento de la cosecha. Por ejemplo, si una fruta como la naranja es cosechada en una fase temprana, tendrá una acidez alta, un nivel elevado de limonina y vitamina C, pero una baja concentración de azúcares, lo cual implica un sabor no deseado.

Por último, el factor tecnológico tiene en cuenta que, en el procesamiento de la fruta para convertirla en jugo, todo es relevante: desde su lavado, extracción, filtrado hasta el tratamiento térmico. Cada etapa del proceso puede afectar de manera positiva o negativa el color, sabor, aroma, textura e incluso la vida útil del jugo.

En definitiva, que tan bueno sea el jugo está sujeto a varios elementos y situaciones que se entrelazan, requiriendo una supervisión minuciosa en cada fase del trayecto. Entender y manejar bien estos puntos clave facilita no solo la mejora del producto final, sino que certificar que es seguro, lo cual ayuda a que guste más y compita mejor en el mercado, cumpliendo así con lo que busca la gente.

### ***2.1.2 Parámetros sensoriales vs. parámetros fisicoquímicos***

Al evaluar la calidad del jugo, muchas veces solo nos fijamos en lo rico que está el jugo o en su aspecto; sin embargo, existe una gran variedad de parámetros técnicos que definen su composición. Es por ello que es clave comprender la importante diferencia entre los parámetros sensoriales y los parámetros fisicoquímicos, porque ambos dan una idea completa de cómo está el producto y si es de buena calidad.

Los parámetros sensoriales son aquellos que se captan con los sentidos: el color a través de la vista, el olor por el olfato, el sabor por el gusto y la textura mediante el tacto. Estos aspectos resultan muy relevantes para el consumidor y suelen ser, en muchos casos, los factores más decisivos a la hora de aceptar o rechazar un jugo.

Por su parte, los parámetros fisicoquímicos se refieren a todas aquellas características medibles mediante un análisis de laboratorio, como la acidez, la cantidad de azúcar, el contenido de limonina, la cantidad de vitamina C y otros compuestos beneficiosos. Estos datos no solo ayudan a evaluar si el jugo posee buenas características nutricionales, sino que también aseguran que el producto cumpla con todas las normas de seguridad alimentaria.

### ***2.1.3 Estándares de calidad en la industria agroindustrial***

Los estándares en el sector agroindustrial representan un conjunto de normas de calidad, que funcionan como un manual de buenas prácticas para regular la producción, los procedimientos y la venta de productos como el jugo natural. Su objetivo es asegurar que lo que llega a nuestras manos cumpla con los requisitos indispensables de seguridad, autenticidad y buena calidad.

De acuerdo con la Comisión del Codex Alimentarius (2005), estas normas fijan parámetros específicos para los jugos y néctares, entre los cuales se delimitan mínimos de sólidos y máximos permitidos de contaminantes o impurezas. Además, establece metodologías analíticas que permiten verificar el cumplimiento de dichas normas, con el fin de generar un ordenamiento en el mercado. Al mismo tiempo, para la agroindustria, aplicar este tipo de normas resulta vital porque protege al consumidor, garantiza que el producto es seguro y permite que el comercio mundial sea más unificado al contar con criterios estandarizados de calidad.

Sin embargo, hay que considerar que no solo existen reglas internacionales; también cada país o región posee sus propias normas. Dependiendo de dónde se origine o hacia dónde se dirija el producto, es necesario amoldarse a esas regulaciones. Todo esto busca garantizar la competitividad y sostenibilidad del sector agroindustrial.

## **2.2 Características fisicoquímicas de las frutas**

Las características fisicoquímicas de las frutas son parámetros que permiten evaluar la calidad, madurez y demás atributos de una fruta para su elaboración en productos como el jugo natural. Estas propiedades inciden directamente en aspectos como el sabor, la textura, el valor nutricional y hasta la estabilidad del producto final. Dentro de un análisis agroindustrial, estas características permiten garantizar que una fruta cumpla con los estándares o normas requeridas en el mercado.

### **2.2.1 Acidez titulable**

La acidez titulable es un indicador primordial para determinar el sabor ácido de una fruta, así como su capacidad de conservación. Además, es importante destacar que este factor aporta ese toque de frescura que hace que un jugo sepa delicioso y no resulte empalagoso.

### **2.2.2 Sólidos solubles (°Brix)**

Los sólidos solubles, medidos en grados Brix, representan cuánta azúcar disuelta posee un jugo. Dicho de otra manera, este parámetro nos permite indicar el nivel de dulzura y la cantidad de sólidos presentes. Es importante destacar que no se trata solo del dulzor: los sólidos solubles también influyen en cómo se percibe el jugo en el paladar, aportando más sabor o incluso una textura particular.

### **2.2.3 Relación**

La relación es un parámetro que resulta de la operación matemática entre los grados Brix y la acidez titulable. Este indicador permite evaluar el balance existente entre la dulzura y la acidez de una fruta, una característica fundamental en la determinación de su calidad. En muchos casos, esta relación se convierte en uno de los factores más decisivos al momento de definir el punto óptimo de cosecha.

### **2.2.4 Firmeza y color**

La firmeza y el color son características físicas fundamentales que se relacionan directamente con la textura y el aspecto visual de la fruta. La firmeza, por su parte, es un indicador clave del grado de madurez y de la resistencia del fruto ante factores como la manipulación, el transporte y la aceptación del producto por parte del consumidor. En cuanto al color, este guarda estrecha relación con la pigmentación natural y el estado de madurez, y suele ser un criterio determinante en la elección del producto por parte del consumidor, quien lo asocia con la frescura y calidad.

### **2.2.5 Madurez**

La madurez corresponde al estado fisiológico de la fruta que indica el desarrollo de sus características adecuadas para el consumo o procesamiento. Sin lugar a duda, determinar el momento óptimo de madurez es un aspecto fundamental, ya que, si la fruta se cosecha muy temprano, presentará sabores poco deseables y una acidez excesivamente alta. Por otro lado, si se cosecha demasiado tarde, perderá parte de su valor nutricional y, además, será más susceptible a daños físicos o deterioro durante su manejo. Por lo cual el conocer y controlar estos factores críticos son esencial para garantizar la calidad del producto final.

### **2.3 Introducción a la inteligencia artificial**

La inteligencia artificial corresponde a una rama de la computación que explora y crea sistemas capaces de ejecutar tareas hechas por humanos, las cuales requieren inteligencia. Estas labores pueden abarcar aspectos donde se implique el razonamiento lógico, el entendimiento del lenguaje natural, la toma de decisiones, el aprendizaje autónomo y el reconocimiento de patrones. La IA ha pasado de enfoques solamente simbólicos y lógicos, a la creación de modelos basados en la recolección de datos y el aprendizaje automático, logrando con ello una transformación digital paulatina de la industria.

Como lo expresan Russell y Norvig (2021), en su libro clásico *Artificial Intelligence: A Modern Approach*, la inteligencia artificial se define como “el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones” (p. 21). Esta definición remarca el concepto de agente inteligente, el cual percibe su entorno, procesa la información obtenida y actúa con el fin de lograr objetivos concretos. En el campo de los sistemas predictivos, estos tipos de agentes pueden decidir con base en modelos entrenados con datos históricos, pudiendo con esto mejorar su desempeño conforme aprenden de nuevas experiencias.

Colocando este contexto en un enfoque industrial y agroalimentario, la inteligencia artificial ha sido una herramienta clave para enfrentar grandes retos como la estandarización de la calidad, la optimización de recursos y la disminución de desperdicios. En concreto, la capacidad de que un modelo logre analizar múltiples variables y hallar relaciones ya sean lineales o no permite la construcción de sistemas de predicción, clasificación y diagnóstico dentro de entornos productivos altamente dinámicos.

Esta investigación pretende crear un sistema predictivo basado en IA cuya finalidad es predecir la calidad del jugo natural antes de su procesamiento final. Para esto, se utilizarán como principal insumo las características físicoquímicas de la fruta. La predicción temprana de calidad representa una ventaja competitiva en los procesos agroindustriales, dado que permite una toma de decisiones operativas más eficiente, reduce reprocesos y asegura un producto final más uniforme y con mayor valoración comercial, algo muy atractivo para el cliente final.

Asimismo, el libro *Deep Learning* de Goodfellow, Bengio y Courville establece que “los sistemas inteligentes se fundamentan en principios de probabilidad y teoría de la información para modelar y gestionar la incertidumbre en el proceso de aprendizaje y toma de decisiones” (p. 70). Bajo este enfoque, la IA ofrece un marco metodológico sólido para analizar, modelar y predecir fenómenos a partir de los patrones que surgen de datos históricos. Sin lugar a duda, actualmente la inteligencia artificial se enriquece con campos como el aprendizaje automático, la estadística avanzada y la ciencia de datos. Todos estos métodos permiten que el sistema genere sus capacidades predictivas gracias a datos fiables, logrando así conclusiones certeras sobre escenarios futuros.

Por otro lado, la implementación de estas herramientas en procesos agroindustriales ha probado ser cada vez más efectiva, sobre todo en áreas como el control de calidad y la detección de fallas en el proceso.

Este apartado tiene como propósito brindar un esquema conceptual sobre la inteligencia artificial, basado en sus teorías, sus procesos y sus aplicaciones, contribuyendo con ello a la mejora y estandarización del producto final.

### ***2.3.1 Historia y evolución de la inteligencia artificial***

La inteligencia artificial ha recorrido un camino largo y lleno de aprendizajes para llegar a ser lo que es hoy, una herramienta poderosa que nos permite predecir comportamientos, automatizar procesos e incluso tomar decisiones en distintos sectores, incluido el agroindustrial.

Todo comienza a mediados del siglo XX, durante una Conferencia en Dartmouth en 1956. En la cual grandes profesionales se reunían al evento, entre ellos la figura clave como John McCarthy, plantearon que la inteligencia artificial sería aquel logro de hacer que las máquinas realicen tareas que, si las hiciera una persona, se considerarían inteligentes. Esto marcó el inicio oficial de este campo de estudio. Aquella simple idea realmente marcó un punto de partida de una nueva era tecnológica que poco a poco iría agarrando más auge.

En sus inicios, la IA se enfocó en modelos simbólicos y reglas lógicas, las cuales se basaban en programar directamente el conocimiento que los humanos compartían con las máquinas. No obstante, este enfoque enfrentaba muchas limitaciones, especialmente cuando los sistemas debían enfrentarse a situaciones inciertas o datos inexactos. Esto generaba un periodo de estancamiento que se conoció como los “inviernos de la IA”, a través de los cuales los entusiastas y los inversionistas en inteligencia artificial se redujeron drásticamente.

Es así como a partir del año dos mil, el desarrollo de modelos más complejos y la capacidad de acceder a grandes volúmenes de datos dieron origen al auge del aprendizaje automático (Machine Learning) y del aprendizaje profundo (Deep learning). Esto fue posible, además, gracias a la mejora en la capacidad computacional de los nuevos equipos de cómputo, lo que permitió manejar y procesar esa gran cantidad de datos de manera eficiente.

Sin lugar a duda, estos nuevos enfoques permitieron que los sistemas aprendieran de los datos sin que tuvieran que ser programados por una persona explícitamente para cada situación que estuviera sucediendo, haciendo de la IA algo más adaptativo y autónomo. Así mismo lo expresan Russell y Norvig (2021), la inteligencia artificial busca desarrollar agentes que sean capaces de percibir su entorno y actuar racionalmente para alcanzar sus objetivos.

Hoy, la IA ha dejado de ser solo una simple idea nacida en una conferencia académica para convertirse en una herramienta accesible y aplicada por distintos profesionales. Áreas como la medicina, la automoción, las finanzas y también el sector agroindustrial han demostrado que su implementación ha sido un factor clave en la toma de decisiones, además de permitir controlar la calidad, predecir el rendimiento y mejorar los procesos.

En este contexto, este proyecto se apoya en esos avances para crear un sistema predictivo que permita anticipar la calidad del jugo natural, contribuyendo a estandarizar su producción, logrando con esto mejorar la aceptación del producto final por parte de los consumidores.

### **2.3.2 *La inteligencia artificial (IA)***

Es una disciplina que, en esencia, busca dotar a las máquinas de la capacidad para simular procesos cognitivos humanos, como aprender, razonar y tomar decisiones informadas. Aunque para algunos la IA pueda parecer etérea o abstracta, en realidad se sustenta en principios sólidos de computación, estadística y matemáticas. En términos simples, un sistema de IA percibe información de su entorno, la procesa mediante algoritmos diseñados para identificar patrones, y produce respuestas o predicciones que se asemejan a una “decisión inteligente”.

Uno de los pilares conceptuales de la IA es el aprendizaje. Aquí, un sistema no solo sigue reglas predefinidas, sino que adapta su comportamiento a partir de la experiencia y los datos históricos disponibles. Esa habilidad para evolucionar con base en nueva información hace de la IA una herramienta especialmente valiosa en contextos dinámicos, como la agroindustria, donde las condiciones ambientales y las características biológicas varían

constantemente. Como lo destaca Goodfellow et al. (2016), la IA moderna ha trascendido los métodos lógicos tradicionales para dar paso a enfoques basados en aprendizaje automático, donde los algoritmos se entrenan con grandes volúmenes de datos, encontrando relaciones complejas que los humanos podrían pasar por alto.

Otro elemento crucial dentro de los conceptos básicos de IA es la idea de “agente inteligente”. Un agente no es más que una entidad que percibe su entorno mediante sensores, actúa a través de actuadores y busca maximizar un objetivo definido. En esta investigación, el sistema predictivo diseñado actúa como un agente capaz de analizar las características fisicoquímicas de la fruta y predecir la calidad del jugo, anticipándose a decisiones operativas con notable precisión.

### **2.3.3 Diferencias entre IA, ML y Deep Learning**

Aunque los términos *inteligencia artificial (IA)*, *machine learning (ML)* y *deep learning* se usan con frecuencia como sinónimos, en realidad representan niveles distintos de sofisticación dentro de un mismo campo. La IA es el paraguas más amplio: engloba todas las técnicas y metodologías diseñadas para que las máquinas puedan simular procesos de inteligencia humana, desde la simple toma de decisiones basadas en reglas hasta algoritmos de aprendizaje adaptativo. En esta investigación, la IA es el marco general que da sentido al desarrollo del sistema predictivo.

Por debajo de la IA se ubica el *machine learning*. Se trata de un subconjunto que se centra en el uso de datos para que los sistemas “aprendan” a realizar tareas sin ser explícitamente programados para cada escenario posible. Así, ML no se limita a seguir instrucciones fijas; en su lugar, genera modelos capaces de generalizar patrones a partir de ejemplos previos. En el caso de la predicción de calidad del jugo, por ejemplo, el ML permite entrenar algoritmos con datos fisicoquímicos históricos para anticipar resultados futuros con base en esas tendencias.

Finalmente, *deep learning* es una especialización dentro del ML. Basado en redes neuronales artificiales profundas, el *deep learning* sobresale en problemas donde las relaciones entre las variables son extremadamente complejas y no lineales. La metáfora es clara: así como la mente humana descompone estímulos en capas de procesamiento, las redes profundas logran extraer características de los datos a distintos niveles, afinando la predicción. Sin embargo, el costo computacional es mayor, así como la necesidad de grandes volúmenes de datos para entrenarlas adecuadamente.

Esta diferenciación no es solo semántica; define los límites prácticos de lo que puede lograrse en un entorno agroindustrial, donde las limitaciones de datos y recursos técnicos obligan a elegir estrategias balanceadas.

## **2.4 Modelos predictivos basados en IA**

Los modelos predictivos basados en inteligencia artificial son, en esencia, herramientas diseñadas para anticipar resultados futuros a partir del análisis de datos históricos y actuales. En un contexto industrial, estos modelos funcionan como una brújula sofisticada: orientan las decisiones operativas hacia un objetivo claro, minimizando la incertidumbre y aumentando la eficiencia del proceso. En la agroindustria, por ejemplo, un modelo predictivo puede indicar el punto exacto de madurez de la fruta para garantizar un jugo con las características deseadas, reduciendo pérdidas y estandarizando la calidad.

Un modelo predictivo no es un artefacto aislado; es el resultado de un proceso estructurado que incluye la selección de variables relevantes, la recolección y limpieza de datos, la definición del algoritmo más adecuado y su entrenamiento con ejemplos reales. Cada uno de estos pasos requiere decisiones técnicas cuidadosas, pues la calidad de los datos alimenta directamente la calidad de las predicciones que el modelo es capaz de entregar. Como explican

Li et al. (2020), los modelos de IA pueden ser supervisados, no supervisados o incluso reforzados, dependiendo del tipo de retroalimentación disponible y de la naturaleza de la tarea.

En el caso de la predicción de la calidad del jugo natural, los modelos supervisados son especialmente pertinentes. Estos modelos se entrenan con pares de datos de entrada (por ejemplo, °Brix, acidez, madurez) y salida (calidad sensorial medida), aprendiendo a identificar las relaciones que existen entre ellos para poder proyectar escenarios futuros a partir de nuevos datos.

Es importante señalar que un modelo predictivo no es estático: necesita ser recalibrado periódicamente para adaptarse a las variaciones naturales del entorno productivo y a cambios en las condiciones climáticas o varietales de la fruta. Así, estos modelos, bien implementados, se convierten en aliados estratégicos para mantener la competitividad y calidad en la industria agroalimentaria contemporánea.

#### ***2.4.1 La regresión lineal***

Dentro del repertorio de modelos predictivos en inteligencia artificial, la regresión lineal y la regresión logística representan dos técnicas fundamentales, apreciadas tanto por su sencillez conceptual como por su efectividad en problemas prácticos. Son, en muchos sentidos, la puerta de entrada al aprendizaje supervisado, ampliamente utilizadas cuando se cuenta con datos etiquetados y bien estructurados.

La regresión lineal, se emplea cuando la variable a predecir es continua, como el nivel de azúcar en un jugo, el pH o la acidez. Su objetivo es modelar la relación entre una o varias variables independientes, por ejemplo, las propiedades fisicoquímicas de la fruta y una variable dependiente, ajustando una línea recta que mejor se aproxima a los datos. Como ilustran Lu et al. (2020), esta técnica es especialmente útil para predecir valores dentro de un rango conocido, permitiendo un análisis interpretable que facilita su adopción en entornos industriales con personal no especializado en algoritmos complejos.

Por otro lado, la regresión logística se aplica cuando la variable objetivo es categórica, como clasificar la calidad del jugo en categorías predefinidas: alta, media o baja. Aunque su nombre sugiere linealidad, la regresión logística utiliza una función sigmoide para modelar la probabilidad de que una observación pertenezca a una clase específica. En términos industriales, permite tomar decisiones binarias o multinomiales sobre la idoneidad de un lote de fruta para procesarse.

Ambos modelos comparten virtudes: requieren relativamente pocos datos para entrenarse, son rápidos de ejecutar y sus resultados son fáciles de interpretar. Sin embargo, su capacidad para capturar relaciones no lineales es limitada, lo que exige complementarlos con técnicas más avanzadas en escenarios complejos.

En el contexto de esta investigación, la elección entre regresión lineal o logística dependerá de la naturaleza de la variable de salida definida para el sistema predictivo, priorizando siempre el equilibrio entre precisión y simplicidad operacional.

#### ***2.4.2 Redes neuronales artificiales***

Las redes neuronales artificiales (RNA) son uno de los modelos más emblemáticos y potentes dentro del aprendizaje automático. Inspiradas en la estructura biológica del cerebro humano, las RNA están formadas por capas de nodos interconectados denominados neuronas que procesan información de manera jerárquica y no lineal, permitiendo captar relaciones complejas entre las variables de entrada y las de salida.

A diferencia de los modelos lineales tradicionales, las redes neuronales destacan precisamente porque son capaces de modelar fenómenos donde las relaciones entre los factores son intrincadas, heterogéneas y no obvias. En el contexto de la predicción de la calidad del jugo natural, las RNA permiten integrar un gran número de características fisicoquímicas de la fruta

y analizar cómo interactúan entre sí para determinar el resultado final, incluso cuando los datos contienen cierta aleatoriedad o ruido (Pandey et al., 2023).

Una RNA típica consta de tres tipos de capas: la capa de entrada, que recibe las variables iniciales; una o más capas ocultas, donde se realizan las transformaciones y cálculos intermedios; y la capa de salida, que genera la predicción. Cada conexión entre neuronas tiene un peso ajustable que se optimiza durante el proceso de entrenamiento, mediante algoritmos como la retro propagación, lo cual permite a la red “aprender” de los errores cometidos y mejorar progresivamente su desempeño.

Una de las principales ventajas de las redes neuronales es su flexibilidad para adaptarse a distintos tipos de datos, incluyendo datos tabulares, imágenes o incluso señales sensoriales, lo que las convierte en una opción versátil para la industria alimentaria. Sin embargo, esta potencia viene acompañada de desafíos: requieren mayor capacidad computacional, tiempos de entrenamiento más largos y, sobre todo, grandes volúmenes de datos para alcanzar todo su potencial predictivo.

En síntesis, las redes neuronales ofrecen un enfoque robusto para problemas donde los métodos tradicionales quedan cortos, permitiendo llevar la predicción a un nivel más profundo y ajustado a la complejidad del fenómeno analizado.

### ***2.4.3 Selección y entrenamiento de modelos***

La selección y el entrenamiento de los modelos son dos de las etapas más críticas en el desarrollo de un sistema predictivo basado en inteligencia artificial, ya que de estas decisiones depende, en gran medida, la efectividad y la confiabilidad del sistema. Este proceso no es simplemente técnico; requiere una comprensión profunda tanto de los datos disponibles como del contexto operativo donde el modelo será implementado.

En primer lugar, la selección del modelo adecuado implica analizar la naturaleza de la variable objetivo, el tipo y volumen de datos y el nivel de complejidad de las relaciones entre variables. En escenarios con datos bien estructurados y relaciones aproximadamente lineales, modelos sencillos como la regresión lineal pueden ser suficientes. En cambio, cuando las interacciones son más complejas y los patrones menos evidentes, resulta necesario optar por métodos más sofisticados, como las redes neuronales o los árboles de decisión (Li et al., 2020). Aquí también se valora la interpretabilidad del modelo: en entornos industriales, la facilidad para explicar las decisiones del sistema a los operarios y tomadores de decisiones suele ser tan importante como su precisión.

El entrenamiento, por su parte, consiste en alimentar el modelo elegido con datos históricos para que “aprenda” las relaciones subyacentes. Durante esta fase, los datos se dividen generalmente en conjuntos de entrenamiento y prueba, lo que permite evaluar si el modelo generaliza bien a datos nuevos. También es fundamental normalizar o estandarizar las variables, identificar y eliminar valores atípicos y verificar la calidad de las etiquetas en los datos. Finalmente, el entrenamiento requiere un ciclo de ajustes iterativos, probando diferentes configuraciones de hiperparámetros hasta encontrar el equilibrio óptimo entre sesgo y varianza. Este balance asegura que el modelo no solo sea preciso con los datos históricos, sino también robusto frente a las variaciones propias del entorno real donde operará.

## **2.5 Aplicación de inteligencia artificial en la agroindustria**

Ha transformado progresivamente la manera en que se producen, procesa y distribuyen los alimentos, llevando a este sector hacia una era más eficiente, precisa y sostenible. Esta adopción no es fortuita; responde a la necesidad urgente de enfrentar desafíos como el desperdicio de materias primas, la variabilidad en la calidad de los productos y la creciente demanda de consumidores por alimentos seguros y estandarizados.

En términos generales, la IA en la agroindustria actúa como un sistema de soporte a las decisiones operativas, capaz de procesar grandes volúmenes de datos provenientes de sensores,

imágenes, análisis de laboratorio y registros históricos. De esta forma, permite detectar patrones ocultos y anticipar resultados que serían difíciles de prever manualmente. Según Lu et al. (2020), las aplicaciones más comunes incluyen desde la predicción del rendimiento de cultivos y la detección temprana de enfermedades, hasta el control automatizado de procesos en plantas industriales y la estimación de la vida útil de los productos.

Particularmente en el procesamiento de frutas y jugos, la IA ayuda a estandarizar la calidad al identificar las características fisicoquímicas óptimas de la materia prima antes de su transformación. Este tipo de análisis preventivo reduce el desperdicio, mejora la eficiencia de las líneas de producción y garantiza que los lotes cumplan consistentemente con los estándares regulatorios y las expectativas del mercado.

Es importante señalar que la implementación de IA en la agroindustria no solo aporta beneficios económicos y operacionales, sino que también contribuye a la sostenibilidad al optimizar el uso de recursos y minimizar las pérdidas. De este modo, la IA no es simplemente una herramienta tecnológica, sino un catalizador de buenas prácticas industriales que alinean competitividad con responsabilidad ambiental y social.

### ***2.5.1 Casos de uso en predicción de calidad de alimentos***

La predicción de calidad de alimentos mediante inteligencia artificial ha dejado de ser un concepto experimental para convertirse en una práctica concreta en diversas industrias alimentarias. Estos casos de uso muestran cómo los modelos predictivos pueden reducir la variabilidad en los procesos productivos y garantizar que el consumidor reciba productos con estándares consistentes, incluso en condiciones operativas cambiantes.

Uno de los ejemplos más notables es la implementación de modelos basados en aprendizaje automático para predecir la frescura y vida útil de productos perecederos como carnes, lácteos y panificados. Sistemas equipados con sensores de gases y algoritmos entrenados permiten estimar en tiempo real el grado de deterioro, mejorando la logística de

distribución y reduciendo pérdidas económicas. Según Pandey et al. (2023), en los últimos años estos sistemas han logrado reducir el desperdicio de alimentos en cadenas de suministro hasta en un 30 %, simplemente al optimizar los tiempos de consumo preferente a partir de predicciones más precisas.

En la industria láctea, por ejemplo, la IA ha sido utilizada para evaluar la calidad de la leche mediante el análisis de parámetros como grasa, proteína y células somáticas, garantizando que las mezclas cumplan las especificaciones antes de ser pasteurizadas o procesadas. Otro caso recurrente se da en la panificación industrial, donde algoritmos predicen la textura y volumen del pan en función de la humedad, la temperatura y la composición de las harinas empleadas, anticipando ajustes para mantener uniformidad en el producto final.

Estos casos demuestran que la IA, aplicada adecuadamente, no solo predice el estado de los alimentos, sino que además se convierte en un aliado estratégico para las empresas, que pueden tomar decisiones más rápidas y acertadas en entornos competitivos y altamente regulados.

### ***2.5.2 Casos en frutas, jugos y control de calidad***

En el ámbito específico de las frutas y jugos naturales, los sistemas predictivos basados en inteligencia artificial han demostrado ser una herramienta clave para garantizar la calidad y uniformidad de los productos. La naturaleza biológica de la materia prima introduce variaciones naturales en parámetros como madurez, acidez, dulzor y contenido de sólidos solubles, que afectan directamente la percepción del consumidor sobre el producto final. La IA, al identificar patrones en estas variables, permite anticipar resultados y tomar decisiones correctivas en tiempo real.

Uno de los casos más comunes se encuentra en la clasificación automática de frutas por calidad. A través de sensores ópticos y algoritmos de aprendizaje supervisado, las máquinas pueden evaluar el color, la textura y la firmeza de cada pieza, clasificando lotes según su

idoneidad para consumo fresco o para procesamiento industrial. Esta tecnología no solo acelera el proceso, sino que reduce significativamente el error humano y mejora el aprovechamiento de la materia prima.

En la industria del jugo, modelos predictivos son entrenados con datos históricos de cosechas y análisis de laboratorio para proyectar la calidad sensorial del jugo resultante antes de la extracción. Esto permite ajustar las mezclas de fruta para lograr un balance más homogéneo entre dulzor y acidez, alineado con los estándares normativos y las expectativas del mercado. Según Li et al. (2020), algunas plantas industriales ya han logrado reducir hasta un 20 % los rechazos de lotes y un 15 % las pérdidas por sobre procesamiento, gracias al uso de modelos basados en redes neuronales y técnicas de regresión.

Estos casos ilustran cómo la IA no reemplaza el conocimiento humano, sino que lo amplifica, permitiendo que los operarios y técnicos tomen decisiones más informadas y rápidas, transformando la variabilidad natural de las frutas en una ventaja competitiva mediante predicciones precisas y confiables.

### ***2.5.3 Desafíos y oportunidades del uso de IA en el sector***

La incorporación de inteligencia artificial en el sector agroindustrial no está exenta de retos, pero también abre un abanico de oportunidades que redefinen las posibilidades productivas y competitivas. Estos dos aspectos coexisten y exigen un análisis equilibrado, capaz de reconocer tanto las limitaciones actuales como el potencial de transformación que ofrece la tecnología.

Entre los desafíos más evidentes está la disponibilidad y calidad de los datos. Los modelos de IA requieren grandes volúmenes de información precisa y bien etiquetada para entrenarse y funcionar correctamente. En muchas agroindustrias, los registros aún son incompletos, inconsistentes o dispersos, lo que dificulta la creación de modelos robustos y confiables. Además, las fluctuaciones climáticas y biológicas naturales de las cosechas añaden

un nivel de incertidumbre que los modelos deben ser capaces de absorber sin perder precisión (Russell & Norvig, 2021).

Otro obstáculo relevante es la inversión inicial que implica la adopción tecnológica: desde la adquisición de sensores y servidores adecuados hasta la capacitación del personal. Si bien los costos tienden a amortizarse en el mediano plazo, representan una barrera significativa para empresas pequeñas o medianas. También persisten desafíos culturales y organizacionales, pues no siempre existe una aceptación inmediata de las herramientas automatizadas, especialmente cuando reemplazan prácticas tradicionales arraigadas.

Sin embargo, las oportunidades son igualmente poderosas. La IA permite una mejor gestión de los recursos, reduce desperdicios, optimiza la calidad final del producto y eleva los márgenes de rentabilidad al mejorar la eficiencia operativa. Además, abre posibilidades para diferenciarse en mercados exigentes, posicionando los productos como más homogéneos, sostenibles y competitivos. La trazabilidad, otra ventaja, fortalece la confianza del consumidor y facilita el cumplimiento de normativas internacionales cada vez más estrictas.

En conjunto, el camino hacia una agroindustria más inteligente no es lineal ni sencillo, pero promete redefinir los límites de lo posible en términos de calidad, productividad y sostenibilidad cuando se asume con visión estratégica y pragmatismo.

## **2.6 Evaluación del rendimiento de modelos**

Evaluar el rendimiento de los modelos predictivos no es un trámite opcional, sino una etapa fundamental para garantizar que las predicciones sean fiables, precisas y aplicables al entorno real donde se implementarán. En la práctica, un modelo sin una evaluación rigurosa corre el riesgo de convertirse en una caja negra sin valor práctico, incapaz de generalizar más allá de los datos con los que fue entrenado.

La evaluación comienza con la definición de métricas claras que permitan medir el desempeño en función del tipo de problema. En problemas de regresión, donde las variables

objetivo son continuas, se utilizan medidas como el error cuadrático medio (MSE) o el coeficiente de determinación ( $R^2$ ), que ayudan a dimensionar qué tan cerca están las predicciones de los valores reales. En tareas de clasificación, por su parte, métricas como la precisión, el recall o la puntuación F1 permiten evaluar no solo la exactitud, sino también el balance entre falsos positivos y falsos negativos, lo cual es crucial en procesos industriales con tolerancia reducida al error.

Pero evaluar un modelo no se limita a calcular números; también implica examinar cómo se comporta frente a datos no vistos y si mantiene la capacidad de generalización. Para ello, técnicas como la validación cruzada dividen los datos en múltiples subconjuntos para probar el modelo en escenarios diversos, reduciendo el riesgo de sobreajuste y brindando una visión más realista de su desempeño.

Además, la interpretabilidad y la confiabilidad son aspectos clave a considerar en la evaluación, especialmente en contextos agroindustriales donde las decisiones automatizadas afectan directamente a la producción y la economía de la empresa. Un modelo que predice correctamente, pero resulta imposible de explicar genera desconfianza y limita su adopción práctica.

### **2.6.1 Métricas de evaluación Métricas de evaluación (MSE, $R^2$ , precisión, recall, F1)**

Son en esencia, el lenguaje con el que se mide la efectividad de los modelos predictivos. Elegir las métricas adecuadas es una decisión estratégica, ya que cada una captura matices distintos del rendimiento y aporta perspectivas complementarias para valorar si un modelo es verdaderamente útil en la practican problemas de regresión, donde el objetivo es predecir valores continuos como grados Brix o acidez, el error cuadrático medio (MSE) es uno de los indicadores más utilizados.

Esta métrica calcula el promedio de los errores al cuadrado entre las predicciones y los valores reales, penalizando más fuertemente los errores grandes y reflejando la estabilidad del

modelo en términos numéricos. Junto al MSE, el coeficiente de determinación ( $R^2$ ) es otra métrica esencial, ya que expresa qué proporción de la variabilidad en los datos es explicada por el modelo, ofreciendo una visión intuitiva de su capacidad explicativa. Según Russell y Norvig (2021), estas métricas básicas son la piedra angular de la validación cuantitativa de cualquier sistema de inteligencia artificial, pues permiten comparar modelos en igualdad de condiciones y detectar rápidamente si un algoritmo está sobre ajustando los datos o subestimando la complejidad del problema.

En tareas de clasificación, donde el resultado es categórico, por ejemplo, calidad alta, media o baja, se recurre a métricas específicas que ponderan diferentes tipos de errores. La precisión mide la proporción de predicciones positivas correctas sobre el total de predicciones positivas, útil cuando las consecuencias de falsos positivos son críticas. El recall, en cambio, cuantifica la proporción de verdaderos positivos detectados sobre el total real de positivos, esencial cuando las omisiones son más costosas que las falsas alarmas. La puntuación F1 combina precisión y recall en una sola métrica armónica, ideal cuando se busca equilibrio entre ambos y el contexto es sensible tanto a errores de inclusión como de exclusión.

***Estas métricas no deben interpretarse de forma aislada, sino como un conjunto que permite evaluar el desempeño del modelo desde distintas aristas, ayudando a identificar sesgos, debilidades y puntos de mejora antes de su implementación definitiva en planta.***

### ***2.6.2 La validación cruzada***

Es una técnica fundamental para evaluar la capacidad de generalización de un modelo predictivo, es decir, su desempeño cuando se enfrenta a datos nuevos que no formaron parte de su entrenamiento. En esencia, consiste en dividir los datos disponibles en varios subconjuntos o pliegues, entrenando el modelo en algunos de ellos y probándolo en los restantes de manera rotativa, hasta completar todas las combinaciones posibles.

Este procedimiento ayuda a obtener una estimación más realista de la precisión y la robustez del modelo, ya que mitiga el riesgo de sobreajuste un fenómeno en el que el modelo aprende demasiado bien las particularidades de los datos de entrenamiento, pero falla al aplicarse a datos desconocidos. Russell y Norvig (2021) destacan que la validación cruzada es especialmente útil cuando se dispone de un conjunto de datos limitado, ya que permite aprovechar toda la información disponible sin sacrificar la calidad de la evaluación intermedia. Además, subrayan que esta técnica ofrece una medida de la variabilidad del rendimiento del modelo, ayudando a identificar configuraciones más estables y predecibles.

Entre las variantes más comunes de esta metodología se encuentra la validación *k-fold*, en la que los datos se dividen en  $k$  pliegues iguales, y el modelo se entrena y evalúa  $k$  veces, utilizando cada pliegue una vez como conjunto de prueba. Al final, se calcula un promedio de los resultados obtenidos en cada iteración para obtener una métrica representativa del rendimiento general

### **2.6.3 La interpretabilidad y la confiabilidad**

Son dos atributos cruciales que definen la utilidad práctica de un modelo predictivo, especialmente cuando se implementa en entornos industriales donde las decisiones tienen consecuencias directas sobre la calidad del producto y la rentabilidad de la operación. Aunque un modelo altamente preciso puede parecer ideal, si sus predicciones no son comprensibles o no generan confianza entre los operadores y responsables de planta, su adopción puede verse comprometida.

La interpretabilidad hace referencia a la facilidad con la que los usuarios pueden entender cómo y por qué el modelo produce ciertas predicciones. Esto resulta particularmente relevante en industrias agroalimentarias, donde los procesos suelen estar regulados por normativas y estándares de calidad que exigen trazabilidad y justificación técnica de cada decisión. Russell y Norvig (2021) señalan que los modelos más complejos, como las redes

neuronales profundas, aunque poderosos, tienden a ser opacos y menos interpretables, lo cual representa un desafío en escenarios donde la transparencia es prioritaria.

Por otro lado, la confiabilidad del modelo está relacionada con su consistencia y capacidad para mantener un buen desempeño ante variaciones naturales en los datos o en condiciones operativas cambiantes. Un modelo confiable no solo predice bien en los datos de prueba, sino que lo hace de forma estable a lo largo del tiempo, demostrando robustez frente a perturbaciones inevitables en el entorno de producción.

Para lograr interpretabilidad y confiabilidad, los desarrolladores suelen incorporar técnicas como la visualización de pesos y contribuciones de variables, la validación continua y el monitoreo del desempeño del modelo una vez desplegado en planta. Estas prácticas no solo mejoran la comprensión del sistema, sino que fortalecen la confianza del personal técnico, facilitando su integración en la toma de decisiones diaria

## **2.7 Los sistemas predictivos integrados**

Representan una evolución natural en la implementación de inteligencia artificial dentro de procesos industriales. No se trata solamente de contar con un modelo aislado que realice predicciones, sino de incorporarlo dentro de un ecosistema tecnológico donde pueda interactuar con sensores, bases de datos, interfaces de usuario y otros componentes operativos de la planta. Este enfoque integral permite que las predicciones se traduzcan directamente en acciones concretas y oportunas, cerrando el ciclo entre análisis y ejecución.

Un sistema predictivo integrado se caracteriza por su capacidad para recibir datos en tiempo real desde dispositivos de medición como sensores de calidad, básculas o espectrofotómetros, procesarlos con el modelo entrenado, y devolver resultados inmediatos que alimentan la toma de decisiones. De esta forma, las predicciones no son un simple reporte a posteriori, sino un insumo clave para ajustar parámetros de producción al vuelo, minimizar desperdicios y asegurar estándares de calidad.

Goodfellow et al. (2016) subrayan que la integración de sistemas predictivos dentro de infraestructuras industriales modernas no solo mejora la eficiencia operativa, sino que también reduce la probabilidad de error humano y permite a las organizaciones avanzar hacia prácticas más sostenibles y competitivas. Además, estos sistemas pueden configurarse para registrar históricos de desempeño, facilitando auditorías, trazabilidad y análisis posteriores para retroalimentar la mejora continua.

Sin embargo, su diseño exige un alto grado de coordinación entre especialistas en IA, ingenieros de procesos y personal de planta, pues cada componente debe ser compatible y responder a las necesidades reales del entorno industrial. En resumen, los sistemas predictivos integrados no son meramente una solución técnica, sino un puente entre el análisis inteligente de datos y la acción práctica, potenciando la capacidad de respuesta de las organizaciones en un mercado cada vez más exigente.

### ***2.7.1 Interfaz de usuario para sistemas predictivos***

La interfaz de usuario en un sistema predictivo no es un accesorio, sino un componente esencial que define cómo las predicciones y recomendaciones generadas por el modelo llegan a las personas encargadas de tomar decisiones. Una interfaz bien diseñada actúa como traductora entre la complejidad técnica del modelo y las necesidades prácticas de los operadores, facilitando la comprensión de la información y aumentando la confianza en las predicciones.

En la agroindustria, las condiciones de trabajo demandan que la interfaz sea clara, intuitiva y resistente a errores. Debe presentar los resultados en formatos fáciles de interpretar, como gráficas codificadas por colores, alertas visuales o numéricas, y paneles interactivos que permitan explorar las causas de cada predicción. Según Lu et al. (2020), una interfaz efectiva no solo transmite los datos de salida, sino que también ofrece explicaciones sobre las variables

más relevantes en cada predicción, ayudando al personal a entender y validar las recomendaciones del sistema.

Además, la interfaz debe ser adaptable al nivel de digitalización y al perfil de los usuarios finales. En plantas con personal poco familiarizado con tecnología avanzada, una presentación minimalista y con lenguaje accesible puede ser más efectiva que complejos tableros llenos de opciones técnicas. Por el contrario, en entornos con operadores expertos, la interfaz puede incluir capas adicionales de análisis detallado y configuraciones avanzadas.

Otro aspecto clave es la capacidad de la interfaz para integrarse con otros sistemas existentes en la planta, como los sistemas SCADA o ERPs, evitando duplicidad de procesos y asegurando que la información fluya sin fricciones. En definitiva, una interfaz bien concebida convierte un sistema predictivo sofisticado en una herramienta verdaderamente operativa y aceptada, cerrando la brecha entre la inteligencia artificial y la acción humana en la línea de producción

### ***2.7.2 Consideraciones técnicas para su implementación en planta***

La implementación de un sistema predictivo en planta no es un simple acto de despliegue tecnológico, sino un proceso que demanda planificación cuidadosa, ajustes estructurales y sincronización con la realidad operativa del entorno industrial. Cada decisión técnica debe estar alineada con las limitaciones físicas, humanas y económicas de la planta, para asegurar que el sistema no solo funcione, sino que aporte valor real y sostenible.

Un aspecto crítico es la infraestructura tecnológica existente. Los sistemas predictivos requieren de redes estables para la transmisión de datos, servidores o servicios en la nube con suficiente capacidad de cómputo y mecanismos de respaldo para evitar pérdidas de información en caso de fallas. Tal como señala Li et al. (2020), sin una base tecnológica sólida, incluso el mejor modelo de inteligencia artificial se convierte en un elemento inerte y poco confiable.

Otro punto importante es la compatibilidad con los sistemas de automatización y control de la planta. La arquitectura del sistema predictivo debe poder comunicarse con sensores, actuadores y tableros de control ya instalados, sin generar cuellos de botella ni duplicidad de procesos. Esto implica definir protocolos de comunicación estandarizados y asegurar que los datos recolectados por los sensores sean consistentes, de calidad y en tiempo real.

La preparación del equipo humano resulta igualmente esencial. La implementación debe incluir sesiones de formación para los operadores y supervisores, no solo en el uso de la interfaz, sino también en la interpretación de los resultados y en las acciones correctivas derivadas de las predicciones. Este acompañamiento garantiza que la tecnología no sea percibida como una imposición opaca, sino como una herramienta aliada que facilita el trabajo diario.

### **2.7.3 Limitaciones tecnológicas en plantas agroindustriales**

Las plantas agroindustriales, especialmente en regiones en desarrollo, enfrentan una serie de limitaciones tecnológicas que condicionan la implementación de sistemas predictivos basados en inteligencia artificial. Estas barreras no solo son técnicas, sino que también reflejan realidades económicas, organizacionales y culturales propias del sector.

Una de las limitaciones más recurrentes es la infraestructura tecnológica insuficiente. Muchas plantas aún operan con sistemas de control tradicionales, desconectados de redes de datos o con sensores obsoletos, incapaces de proporcionar la calidad y frecuencia de información que los modelos predictivos requieren. Pandey et al. (2023) subrayan que la falta de digitalización en las etapas iniciales del proceso productivo limita la efectividad de la inteligencia artificial, pues los datos recolectados resultan incompletos o poco fiables.

Otra barrera significativa es el costo asociado a la modernización. La adquisición de hardware especializado, servidores robustos, licencias de software y servicios de mantenimiento representan inversiones considerables, difíciles de asumir para empresas pequeñas o medianas que ya operan con márgenes estrechos. Esta realidad obliga a priorizar

estrategias graduales y soluciones híbridas que permitan ir integrando capacidades predictivas de manera escalonada.

También hay desafíos relacionados con la disponibilidad de personal calificado. La gestión de sistemas predictivos requiere perfiles técnicos capaces de comprender tanto la lógica de los modelos como su impacto en la operación diaria. En muchas plantas, la escasez de talento especializado se traduce en resistencia al cambio y en un uso subóptimo de las herramientas implementadas.

Finalmente, las plantas agroindustriales operan en entornos caracterizados por alta variabilidad climática, estacionalidad y condiciones impredecibles. Estas características aumentan la dificultad de entrenar modelos generalizables y robustos, lo que exige ajustes continuos y actualizaciones periódicas.

En suma, aunque las limitaciones son reales, no son insalvables: reconocerlas es el primer paso para abordarlas de manera estratégica y convertirlas en oportunidades de mejora a mediano y largo plazo.

### **3 Capítulo 3. Marco Metodológico**

En esta sección se exponen los criterios metodológicos que orientan la presente investigación, cuyo objetivo es desarrollar un sistema predictivo apoyado en inteligencia artificial. Este sistema busca anticipar la calidad del jugo natural a partir de las características fisicoquímicas de la fruta antes de su procesamiento, considerando el muestreo que se realiza antes y durante dicho proceso. Para ende, se detallan los métodos y enfoques utilizados para recopilar, analizar y estructurar los datos experimentales, es por esto por lo que, por medio del diseño de investigación, la definición de la población de estudio, la muestra seleccionada y las herramientas empleadas. Los metodológicos permiten construir una base sólida para el desarrollo del modelo predictivo, asegurando su pertinencia práctica y su validez científica.

#### **3.1 Tipo de investigación**

Este proyecto se clasifica como una investigación aplicada y de enfoque cuantitativo, ya que busca resolver un problema real y específico en el sector agroindustrial. Se parte del uso de conocimientos científicos con el fin de anticipar la calidad del jugo natural antes de ser procesado, apoyándose para ello en herramientas de inteligencia artificial.

La idea central es crear un modelo predictivo que permita cumplir este propósito, utilizando variables medibles de la fruta en una etapa temprana, como el Brix, la acidez, la relación Brix/acidez, el color, entre otras. Todo esto con la intención de brindar una solución práctica y técnica, que ayude a mejorar la selección de fruta óptima, disminuir pérdidas y apoyar la toma de decisiones dentro de la empresa.

Por su parte, esta investigación tiene un enfoque cuantitativo, ya que trabaja con una gran cantidad de datos numéricos, los cuales se analizan a través de métodos estadísticos y computacionales mediante un proceso de aprendizaje automático. En resumen, se trata de una investigación de tipo aplicado y cuantitativo, enfocada en el desarrollo de un modelo predictivo que combine ambos enfoques dentro de un entorno agroindustrial.

### **3.2 Alcance investigativo**

Esta investigación se caracteriza por ser correlacional y explicativa, ya que no solo intenta describir una gran variedad de acontecimientos, sino que busca entender la relación entre las diferentes variables fisicoquímicas de las frutas y cómo estas pueden influir en la calidad de un jugo natural, o bien en el momento exacto en que la fruta debe ser cosechada, logrando así alcanzar su fase óptima. Además, pretende explicar cómo estas propiedades pueden formar la base para el desarrollo de un modelo predictivo que brinde una idea de cómo podría comportarse el jugo, ofreciendo así una visión anticipada de su calidad, incluso antes de ser procesado.

En primera instancia, el estudio tiene un enfoque correlacional, dado que examina el nivel de correlación entre variables medibles como el Brix, la acidez, la relación Brix/acidez, el color, la madurez, entre otras variables que juegan un papel importante en cómo será el jugo. Se pretende identificar si estas variables tienen una relación significativa entre sí, influyendo así en las características sensoriales y nutricionales del producto final.

En segundo lugar, se destaca el enfoque explicativo, ya que uno de los propósitos del proyecto es crear un modelo capaz de anticipar el comportamiento del jugo a partir de un conjunto determinado de datos. Para esto, se pretende emplear métodos de inteligencia artificial, como redes neuronales o modelos de regresión avanzada, que permitan establecer relaciones causales y patrones que podrían estar ocultos entre los datos.

Esto permitirá brindar una explicación clave para entender ciertas condiciones fisicoquímicas de la fruta y cómo estas pueden mejorar o deteriorar la calidad, tomando en cuenta todos los aspectos técnicos involucrados.

**Con ello, el estudio busca profundizar en el análisis de estas interacciones, permitiendo así fundamentar científicamente la construcción del sistema predictivo.**

### **3.3 Enfoque**

Para la elaboración de esta investigación, se ha optado por una metodología cuantitativa, dado que se centra en la evaluación precisa y objetiva de los valores numéricos resultantes del muestreo de la fruta, verificando con ello teorías, identificando conexiones entre la información y, a partir de esto, creando un modelo que permita realizar una predicción. Dentro de la perspectiva cuantitativa, se facilita la recopilación organizada de la información, el uso de métodos estadísticos y la extrapolación de conclusiones a partir de análisis prácticos.

En este enfoque, resulta ideal la creación de un modelo predictivo, ya que requiere manejar una gran cantidad de información numérica obtenida a partir de evaluaciones fisicoquímicas realizadas antes de su transformación en jugo de calidad. Asimismo, la perspectiva cuantitativa impulsa la imparcialidad, la replicabilidad y la exactitud de las conclusiones obtenidas, elementos fundamentales para una investigación científica que busca proponer una solución técnica con valor práctico dentro de una industria tan exigente como la agroalimentaria.

En conclusión, el estudio se sitúa completamente dentro de la perspectiva cuantitativa, con un fuerte fundamento empírico y analítico, enfocado en la conversión de datos en conocimiento útil y aplicable, mediante el uso riguroso de métodos estadísticos y algoritmos de predicción.

### **3.4 Diseño**

El diseño metodológico que guía este estudio es de tipo no experimental y transeccional correlacional. Se basa en la construcción de un modelo predictivo, para lo cual se emplean herramientas como la inteligencia artificial. El diseño no experimental se distingue por observar el fenómeno en su estado original, ya que no pretende manipular intencionadamente las variables independientes.

Dentro de este enfoque, se trabajará con variables como el Brix, la acidez, la coloración, la madurez, entre otras características fisicoquímicas de la fruta, que serán medidas en una fase inicial de muestreo sin alterar sus condiciones originales. Posteriormente, se buscará su correlación con la calidad del jugo final, evaluando cuáles atributos objetivos o registros sensoriales muestran una conexión significativa con los datos que podrían alimentar el diseño del modelo predictivo.

Por su parte, el diseño también es transeccional, dado que la recolección de datos se realiza en momentos puntuales o durante procesos específicos de la cosecha y transformación de la fruta. Esto permite brindar una representación fiel de las propiedades fisicoquímicas y su posible relación con la calidad del jugo. Comprender esto implica también observar y analizar el fenómeno en el tiempo exacto en que ocurre.

Finalmente, el planteamiento correlacional permitirá establecer relaciones estadísticas entre las variables, tanto independientes como dependientes, lo cual servirá como base para el entrenamiento del modelo y, con ello, la evaluación de su eficacia predictiva

### **3.5 Población y muestreo**

La población de esta investigación está relacionada con las frutas destinadas a la producción de jugo natural en contextos agroindustriales. Específicamente, se consideran las múltiples variedades de fruta que pueden llegar a ser utilizadas para la elaboración de jugo, como lo son la naranja, la piña, la guayaba, el mango, entre muchas otras, dependiendo de la temporada o la región, donde puede haber más o menos disponibilidad.

Considerando que el objetivo de este proyecto es desarrollar un modelo predictivo basado en las variables fisicoquímicas de la fruta, se puede establecer que la población puede comprenderse como el conjunto de unidades de fruta individual, con sus propias particularidades, las cuales se pueden medir antes de ser procesadas.

Para esto, se toman muestras utilizando métodos no aleatorios, eligiendo las frutas que resultan convenientes según ciertos criterios, como el nivel de madurez, la disponibilidad temporal, el acceso a fincas, y la diversidad necesaria para obtener datos de entrenamiento adecuados para los algoritmos de inteligencia artificial.

Este tipo de muestreo es útil para investigaciones prácticas como la presente, en donde se busca entender en profundidad cómo se comportan ciertas variables necesarias para alimentar el sistema, con el fin de identificar la fase de mejor calidad durante la temporada de cosecha.

### **3.6 Instrumentos de recolección de datos**

Los datos utilizados en este estudio corresponden a registros previos realizados por el personal del área de calidad, quienes llevan a cabo el muestreo y la medición de las frutas como parte del flujo programado de control agroindustrial.

Estos registros se almacenan en el sistema de calidad de la empresa y contienen variables fisicoquímicas como el Brix, la acidez, la relación Brix/acidez, la firmeza, el color y el nivel de

madurez. La información recolectada se obtiene mediante instrumentos técnicos y evaluaciones sensoriales que garantizan la precisión y confiabilidad de los datos utilizados en el análisis.

### **3.6.1 Documentación**

Dentro de esta indagación, la información científica y técnica representa una herramienta esencial para respaldar con teoría la creación del modelo predictivo que estimará la calidad jugo natural. Para esto se realizó el estudio de documentos que ayudará a entender el contexto del fenómeno analizado, encontrar los factores importantes y confirmar que los métodos de medición y análisis son correctos.

Para lograr esto, haremos una búsqueda minuciosa en fuentes originales y complementarias de bases de datos académicas conocidas como Google Académico, SciELO y la biblioteca universitaria, concentrándonos en tres aspectos clave: las propiedades fisicoquímicas de las frutas, la definición y evaluación de la calidad del jugo natural según estándares internacionales como el Codex Alimentarius y otros reglamentos técnicos, y el empleo de la inteligencia artificial en el sector agroindustrial mediante modelos de aprendizaje automático, redes neuronales y métodos de análisis predictivo.

Este análisis de documentos no solo hará más fácil la construcción del marco teórico y la explicación de por qué elegimos ciertas variables, sino que además nos dará referencias metodológicas que nos permitirán determinar si el modelo que desarrollemos es válido, preciso y relevante

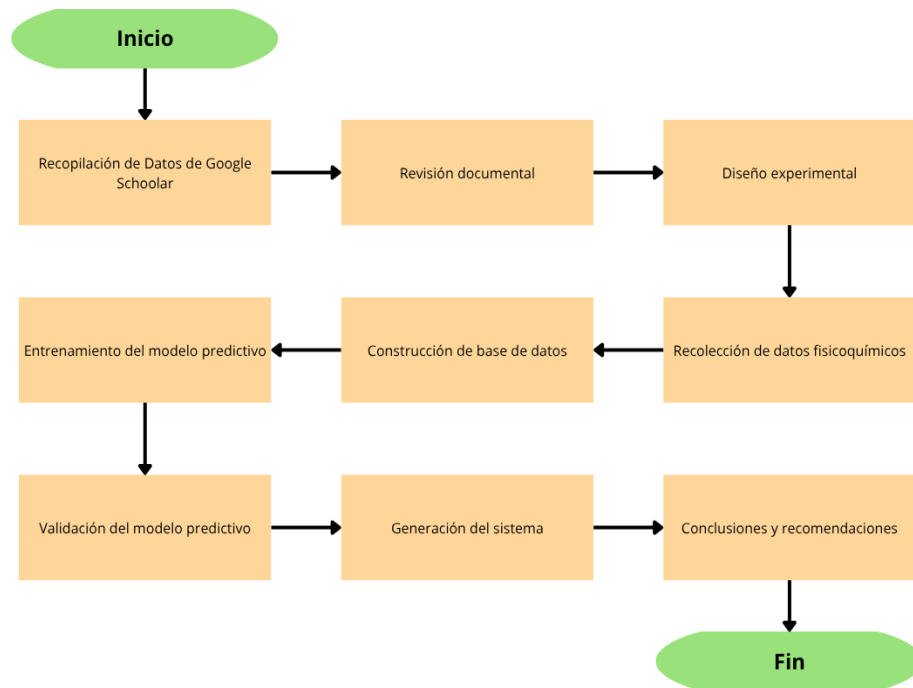
### **3.7 Técnicas de análisis de información**

Se genera un diagrama de flujo que muestra de manera gráfica el flujo de los datos, se realiza un análisis de la información, con el propósito de identificar la propuesta a realizar, su

presentación y las recomendaciones necesarias. Así mismo, este diagrama ilustra el orden correcto desde la obtención hasta el análisis de la información, para concluir con unas recomendaciones.

### 3.8 Diagrama de flujo

**Figura 2: Distribución de procesos del proyecto**



Fuente: Elaboración propia

## **4 Capítulo 4. Análisis del diagnóstico**

Luego de revisar detenidamente los documentos obtenidos, al mismo tiempo, analizar los datos recopilados a través de las distintas metodológicas mencionadas en el capítulo 3, fue posible establecer una conclusión concreta. Así mismo el resultado expresa una mirada reflexiva, crítica y fundamentada, basada en la evidencia reunida durante la investigación.

### **4.1 Revisión del proceso actual de control de calidad en frutas para jugo**

En el sector agroindustrial, enfocado en la elaboración de productos como jugos naturales de alta calidad, se realizan procesos que supervisan y controlan este aspecto desde fases previas a la recolección de la fruta. Esto se hace mediante diversas pruebas que se llevan a cabo tanto en campo como en planta, y a solicitud del cliente. Este tipo de pruebas permite conocer aspectos esenciales como la cantidad de azúcares, la acidez que se puede medir, el color y la madurez de la fruta. Este estudio inicial permite hacerse una idea de cómo será el jugo que se va a producir, reduciendo así la posibilidad de tener problemas por cosechar la fruta demasiado temprano o demasiado tarde.

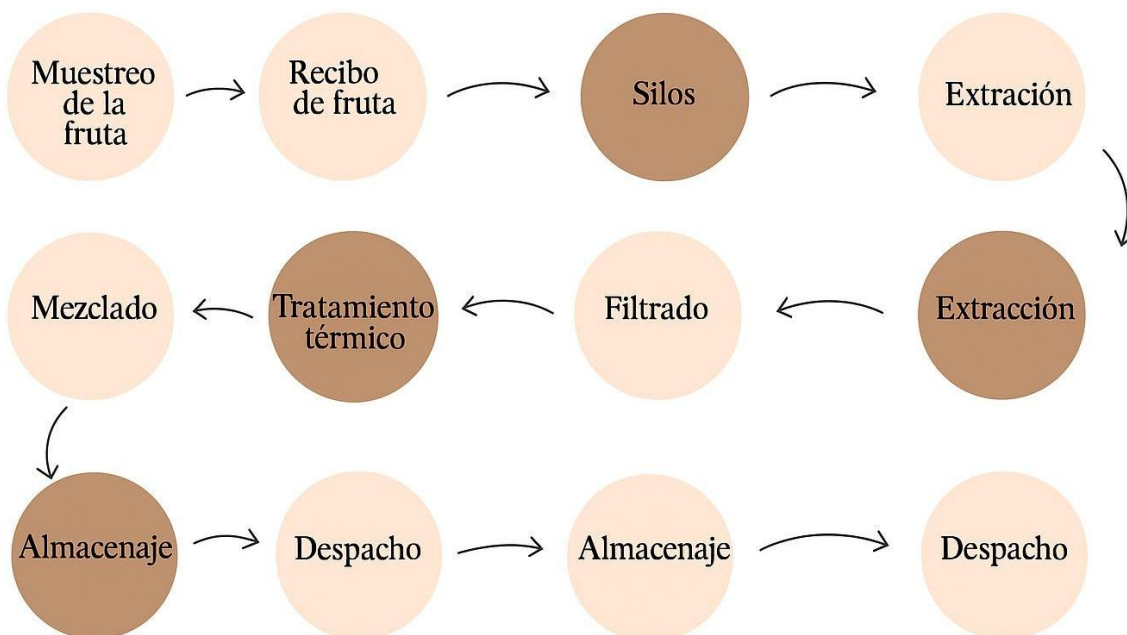
Después de la fase de recolección, la fruta es trasladada desde cada una de las fincas hacia la planta de procesamiento. Ahí permanece en el patio, y conforme al orden de llegada, se va vaciando. En ese momento se realiza una revisión de la fruta llamada recepción, en la cual se extrae una porción de la fruta desde diferentes partes de la carreta para realizar un análisis promedio del estado en que llegó ese lote. Este lote se revisa de forma rápida para confirmar que cumpla con los estándares, y luego es almacenado en silos. Una vez que se tiene la cantidad suficiente de fruta almacenada, la planta inicia el proceso de producción, asegurando que se mantenga en operación continua sin interrupciones.

El flujo sigue con el traslado de la fruta desde los silos hacia las etapas de lavado, extracción, filtrado, tratamiento térmico y mezclado, siendo esta última fase variable según el

tipo de industria. Al final, el jugo es almacenado nuevamente en silos o tanques de retención, desde donde se va extrayendo según la solicitud del cliente y sus especificaciones, ya sea por medio de cisternas o en tambores.

Este proceso cuenta con varios puntos de control donde se supervisa la calidad del producto. A lo largo del mismo, también se obtienen subproductos para aumentar el aprovechamiento y la rentabilidad. Sin embargo, muchos de estos datos no son analizados a profundidad ni están integrados en un modelo de predicción que permita determinar con mayor exactitud cómo se comportará el jugo, considerando todo el conocimiento acumulado por la compañía a lo largo de los años.

**Figura 3: Secuencia de procesos de la planta procesadora de fruta**



Fuente: Elaboración propia

## 4.2 Análisis de los registros históricos de calidad

El registro histórico ha evidenciado que, aunque muchas frutas comparten algunas características muy similares, cada una posee su propia curva de comportamiento, misma que marca cuándo es el mejor momento para recogerla. En un principio, cuando esta fruta está madurando o en su fase inicial, no sabe tan bien, ya que se siente mucho la acidez.

En un caso como la naranja, no son tan dulces y suelen tener un color no adecuado, pero esto mejora con el pasar del tiempo, llegando a un punto de equilibrio donde los valores fisicoquímicos congenian entre sí, dando así una fruta de altísima calidad. Luego, poco a poco, esta va perdiendo sus atributos y con esto baja su calidad otra vez, lo que no da un buen producto final.

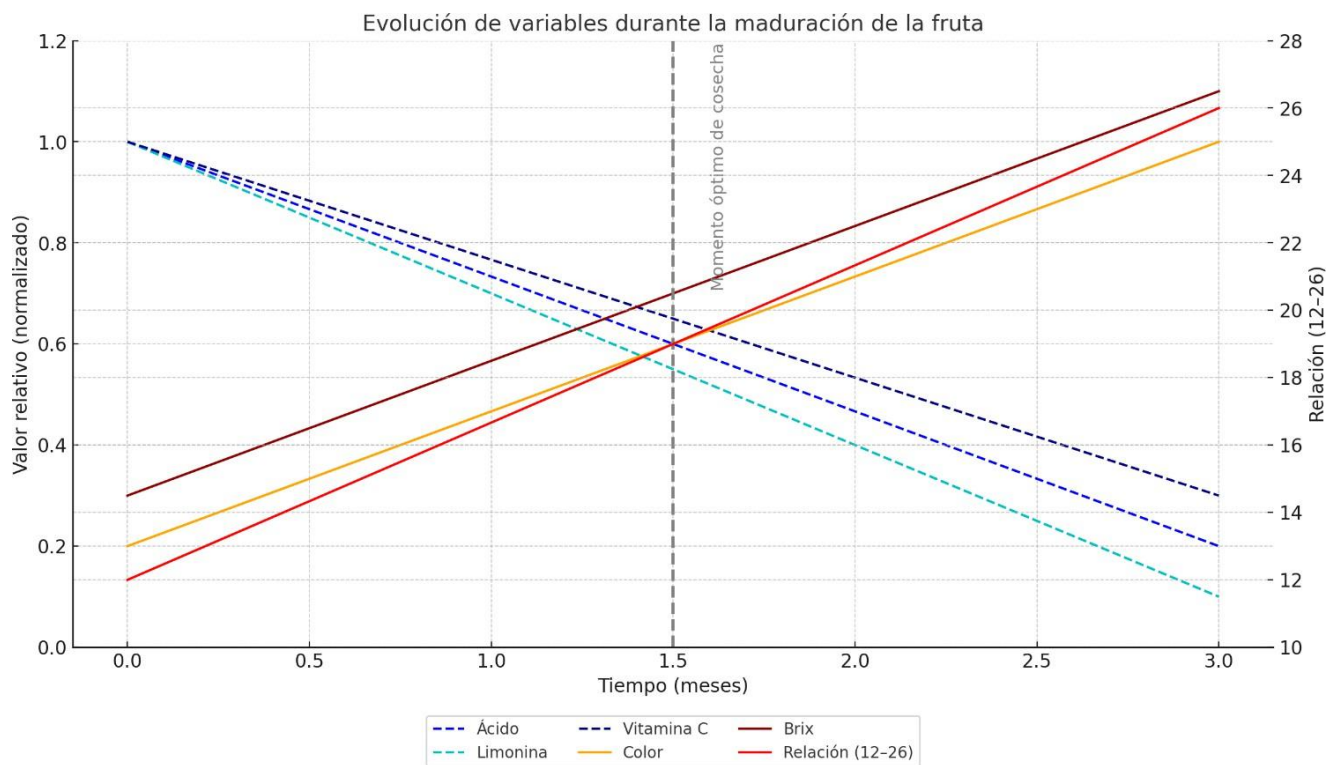
En esta industria se suele trabajar con términos como valores no deseados, faseado, regular, óptimo, nuevamente regular, y luego otra vez no deseado. En la medida de lo posible, se trata de estar lo más cercano al valor óptimo, tomando en consideración los múltiples factores que esto puede implicar antes de llegar a ser procesada una fruta.

Para dar un ejemplo, en un caso como la naranja, esta es muy ácida y tiene una limonina y vitamina C muy altas en fases iniciales. Pero su coloración es baja, el Brix es bajo y la relación también es baja. Con el pasar del tiempo, esto va mejorando hasta llegar a un punto óptimo donde la fruta sabe mucho mejor, y posteriormente estos valores vuelven a caer.

Como norma de la industria, se suele usar como referencia la USDA, en la cual se considera que una fruta es perfecta para un jugo de alta calidad si su relación de Brix/acidez está entre 17 y 18, y su composición de limonina está por debajo de los 25 mg/L. Hay que tomar en cuenta que, si la limonina es demasiado alta, puede tener un sabor más amargo, algo no tan grato para muchas personas. Por eso, estos valores hacen que la limonina sea casi imperceptible para el humano común.

Sin embargo, la obtención de este jugo perfecto es realmente complicada, ya que existen factores como el suelo, el clima, las variables de la fruta e incluso de una misma fruta. El término fórmula de aplicación se refiere a todo lo que el árbol recibe durante el cultivo. También influye el personal de cosecha, pues en un caso como la naranja, este proceso se realiza a mano, y tener gran cantidad de mano de obra es realmente complicado. Asimismo, la logística por parte del transporte vuelve todo este proceso algo complejo, donde incluso con una lluvia todo puede acelerarse rápidamente.

**Figura 4: Evolución de variables fisicoquímicas durante la maduración de la fruta**



Fuente: Elaboración propia

### **4.3 Evaluación de la trazabilidad de los datos**

Existe una ruta bien definida y sin interrupciones desde la recolección hasta el producto final. Cada lote de frutas tiene un registro que detalla cuándo se recogió, en qué lugar se cultivó, su grado de madurez, y los resultados de los análisis químicos y físicos en las fases de recepción y procesamiento.

Este nivel de trazabilidad permite vincular directamente cómo estaba la fruta al inicio con la calidad del jugo obtenido, lo cual resulta muy útil para desarrollar un sistema que pueda anticipar los resultados. Además, se observa que la información se guarda de forma ordenada y con una estructura clara, lo que facilita el análisis estadístico y su integración en modelos de aprendizaje automático.

Esta capacidad de rastreo representa una base sólida para aplicar inteligencia artificial, ya que se cuenta con datos etiquetados que permiten entrenar, validar y mejorar las estimaciones de forma más precisa, contribuyendo así al desarrollo de un modelo predictivo confiable.

### **4.4 Limitaciones del sistema actual de evaluación**

Aunque esta industria se esfuerza por mantener una calidad excelente año tras año, el sistema que implementa presenta algunas deficiencias que dificultan la exactitud al tomar decisiones en poco tiempo. Un punto determinante es la disponibilidad de mano de obra para realizar el proceso de cosecha. En un caso como la naranja, al tratarse de un fruto de temporada, el cosechador sabe que su trabajo durará en promedio tres meses. Después de eso, vuelve a quedar desempleado, lo que vuelve esta labor poco atractiva para muchas personas.

Además, no contar con una herramienta que ayude a predecir cuál es el mejor momento para cosechar los pone en aprietos. Si se recolecta la fruta antes de tiempo como ocurre con la naranja se corre el riesgo de que el jugo tenga mucha acidez, poco Brix y un alto

contenido de limonina, lo que afecta negativamente el sabor, la calidad del producto y, por supuesto, la percepción del cliente.

Estos factores de sincronización generan costos adicionales, provocan desperdicio de materia prima y afectan la eficiencia general del proceso. La solución técnica más importante sería contar con un sistema que, basado en datos reales, permita anticipar el comportamiento de la fruta y determine con precisión el punto óptimo para cosechar.

#### **4.5 Conclusiones del diagnóstico**

Tras el estudio que se ha llevado a cabo en este capítulo, queda claro que el sector enfrenta varios desafíos que impactan de lleno la eficacia y la excelencia del jugo final. Algo que resalta es que no hay manera de saber con seguridad cuándo es el momento justo de empezar a recolectar, lo que causa dudas, se tira material y el jugo natural no siempre sale igual.

Por esta razón, el desarrollo de un sistema predictivo es necesario, ya que permitiría tomar decisiones más fundamentadas, con base en datos reales y patrones históricos. Este tipo de herramienta ayudaría a predecir con mayor exactitud el punto ideal para recolectar la fruta, garantizando así una mejor calidad del jugo y una mayor satisfacción del cliente.

Las variables seleccionadas para alimentar el modelo predictivo fueron elegidas porque, en su conjunto, guardan una estrecha relación con las características fisicoquímicas de la fruta; entre ellas se encuentran el Brix, la acidez, la relación Brix/acidez, la limonina, el color y el nivel de madurez. Estos factores, además, se recogen tanto durante el muestreo como en el procesamiento; por lo tanto, permiten tener una visión más completa y realista del estado del producto en distintas etapas. En otras palabras, no se trata solo de datos aislados, sino de indicadores clave que influyen directamente en la calidad final del jugo.

A partir de esto, el sistema propuesto busca dar respuesta a una de las principales dificultades detectadas en el proceso productivo: la incertidumbre sobre el mejor momento para iniciar la cosecha. Con esta herramienta, se espera no solo reducir errores y evitar pérdidas innecesarias, sino también lograr que la producción esté más alineada con los estándares que exige la industria y con lo que esperan los consumidores; es decir, un producto de alta calidad, consistente y confiable.

## **5 Capítulo 5 Desarrollo Del Modelo Predictivo**

La industria agroalimentaria, y en particular el sector dedicado al procesamiento de jugo natural, enfrenta un desafío constante: garantizar la calidad del producto final en un entorno caracterizado por alta variabilidad química, estacionalidad y condiciones ambientales impredecibles. A lo largo de los años, las plantas procesadoras han dependido de evaluaciones manuales, tardías o empíricas, lo que deriva en pérdidas, reprocesos, altos costos operativos e incluso el descarte de lotes

## Documento final de Proyecto de

completos cuando la fruta no cumple los estándares mínimos de calidad. Tal como se expone en el documento base, la dependencia de métodos empíricos genera variaciones significativas y decisiones reactivas que impactan la eficiencia del proceso.

Frente a este escenario, la integración de herramientas de inteligencia artificial (IA) se perfila como una respuesta estratégica. Como indica el texto original, estos sistemas permiten “predecir la calidad del jugo desde antes de su procesamiento, siendo esto un factor crucial en la toma de decisiones”. Esto habilita a la industria a anticipar situaciones críticas, reducir desperdicios y actuar con precisión basada en datos objetivos.

Además, la investigación subraya que la capacidad de los modelos predictivos para analizar grandes volúmenes de datos históricos mejora la toma de decisiones en tiempo real y permite optimizar la operación técnica y financiera de las plantas. Esto se confirma cuando el documento señala que estos modelos posibilitan “mejorar la eficiencia operativa y la toma de decisiones informada”.

Asimismo, el proyecto se fundamenta en la creciente demanda de herramientas tecnológicas por parte del sector agroindustrial, que busca procesos más estandarizados, menos dependientes de la subjetividad humana y con menores pérdidas de materia prima. Desde esta perspectiva, el propio texto enfatiza que un sistema predictivo aporta “una solución práctica y técnica que ayuda a mejorar la selección de fruta óptima, disminuir pérdidas y apoyar la toma de decisiones dentro de la empresa”.

En ese contexto, este capítulo presenta el desarrollo integral del modelo predictivo, incluyendo su fundamentación teórica, metodología, selección de algoritmos, resultados obtenidos, análisis comparativo, discusión y recomendaciones finales. También se integra el diseño conceptual del sistema que permitirá que el modelo sea utilizado por personal técnico en planta, cumpliendo el objetivo específico: “Diseñar un sistema predictivo integral que incorpore el modelo desarrollado en una interfaz sencilla para usuarios técnicos, considerando su implementación práctica en entornos industriales de procesamiento de jugo”.

Con ello, el capítulo constituye un puente entre la teoría, la experimentación y la aplicación práctica, respondiendo a la visión del documento: “ofrecer un enfoque viable, innovador y útil que permita anticipar la calidad del jugo natural antes de su procesamiento”.

## **5.1 Marco Teórico**

El desarrollo de sistemas predictivos en la industria agroalimentaria se sustenta en un cuerpo teórico que combina fundamentos de inteligencia artificial (IA), procesamiento industrial, análisis fisicoquímico y control de calidad. Este apartado integra estos elementos, apoyándose en las bases conceptuales descritas en el documento original.

### ***5.1.1 Inteligencia Artificial en la Agroindustria***

Durante las últimas décadas, la inteligencia artificial ha experimentado una expansión acelerada en diversos ámbitos productivos, transformando la manera en que se toman decisiones y se controlan procesos. La IA marcó “un antes y un después en muchas áreas de trabajo, como la agricultura y la industria” debido a su capacidad para procesar grandes volúmenes de datos y generar decisiones precisas.

En la agroindustria, esta revolución tecnológica ha permitido automatizar tareas tradicionalmente manuales, como la clasificación de cultivos, la detección de enfermedades o la predicción temprana del rendimiento. La IA, especialmente a través de técnicas de aprendizaje automático, proporciona herramientas que aumentan la eficiencia, ya que los modelos pueden aprender patrones de comportamiento y replicarlos con alta exactitud.

Estas herramientas permiten “tomar decisiones inteligentes basadas en gran cantidad de datos, algo que anteriormente parecía inalcanzable para la tecnología”. Esta característica resulta crucial en la industria del jugo, donde la variabilidad química, climática y estacional requiere sistemas capaces de identificar patrones sutiles que escapan al análisis humano tradicional.

### ***5.1.2 Control de Calidad del Jugo Natural***

El control de calidad ha sido históricamente un proceso reactivo.

Este enfoque presenta varias limitaciones:

genera retrasos operativos,

incrementa el riesgo de reprocesos,

aumenta las variaciones del producto,

y puede derivar en el descarte total de lotes.

Los problemas se agravan cuando los análisis fisicoquímicos se realizan de forma empírica o tardía, provocando pérdidas inevitables. Por ello, se destaca la necesidad urgente de pasar del enfoque tradicional a uno predictivo: un modelo basado en IA representa una propuesta viable, innovadora y útil... ya que mejora la toma de decisiones y reduce el desperdicio.

### ***5.1.3 Importancia de las Características Fisicoquímicas***

El funcionamiento del modelo predictivo se basa en variables fisicoquímicas que describen el estado real de la fruta previo al procesamiento, estas incluyen:

BRIX (contenido de sólidos solubles)

ACIDEZ

RELACIÓN BRIX/ACIDEZ

COLOR DE LA FRUTA

FIRMEZA

COLOR VISUAL DEL JUGO

SABOR

RENDIMIENTO

El texto señala que estas características “inciden mayormente en la calidad del jugo natural” y que, por lo tanto, deben ser integradas en cualquier modelo de predicción robusto.

La literatura sugiere que:

BRIX se relaciona con dulzor y madurez.

Acidez influye en el sabor y estabilidad.

La relación BRIX/Acidez es uno de los indicadores más usados para clasificación sensorial y comercial.

El color, tanto de la fruta como del jugo, refleja estado de madurez, frescura y posible exposición climática.

La firmeza puede asociarse a variedad, sobre maduración o condiciones de almacenamiento.

Como se indica claramente que estas variables “permiten anticipar... la calidad esperada del producto final antes de su procesamiento”.

#### **5.1.4 Modelos Predictivos en Procesos Industriales**

Las herramientas de inteligencia artificial se han vuelto fundamentales para detectar patrones complejos en los datos, se enfatiza que modelos como redes neuronales, regresión, árboles de decisión y Random Forest han demostrado ser eficaces en predicciones sobre datos fisicoquímicos.

Estos modelos destacan por su capacidad de:

manejar datos no lineales,

clasificar múltiples categorías,

detectar anomalías,

aprender patrones de temporadas específicas,

generalizar a nuevos lotes y proveedores,

operar con alta precisión, incluso por encima del 97%, como se señala en los beneficios del modelo predictivo de la empresa Ticotrut.

En términos operativos, estos modelos:

reducen el tiempo de toma de decisiones,

disminuyen las pérdidas por mala asignación de fruta,

evitan subjetividad humana,

permiten estandarizar criterios de calidad,

apoyan decisiones estratégicas como planificación de zafra o negociación con proveedores.

#### **5.1.5 Limitaciones Clave Para Considerar**

Aunque su potencial es amplio, el documento resalta que existen limitaciones importantes para considerar en el diseño de un sistema predictivo industrial:

Altos costos asociados a hardware, software y servicios especializados, lo cual afecta a empresas pequeñas o medianas.

Falta de personal técnico especializado, lo que genera resistencia al cambio y uso subóptimo de las herramientas predictivas.

Variabilidad climática y estacional, que dificulta entrenar modelos generalizables y exige actualizaciones periódicas.

No obstante, estas limitaciones “no son insalvables” y que reconocerlas constituye “el primer paso para abordarlas de manera estratégica”.

### ***5.1.6 Fundamentación del Proyecto dentro de la Industria***

Finalmente, se establece con claridad la pertinencia del proyecto:

Esta investigación permite avanzar hacia “una producción más estandarizada, con mejores características nutricionales” y aporta “una nueva visión al control de calidad en alimentos”.

Asimismo, la viabilidad se fundamenta en:

disponibilidad de datos históricos,

acceso a herramientas de código abierto,

interés del sector por maximizar rentabilidad,

capacidad predictiva avanzada con algoritmos modernos.

Todos estos elementos consolidan el marco conceptual del modelo desarrollado.

## **5.2 Metodología**

La metodología aplicada en esta investigación se diseñó para garantizar un proceso riguroso, reproducible y alineado con el objetivo principal: desarrollar y validar un modelo predictivo capaz de anticipar la calidad del jugo natural a partir de características fisicoquímicas de la fruta antes del procesamiento industrial.

Esta sección expone el enfoque metodológico, diseño, procedimientos, herramientas, modelos y técnicas analíticas utilizadas.

### ***5.2.1 Enfoque y Tipo de Investigación***

De acuerdo con el documento base, esta investigación se clasifica como aplicada, con un claro énfasis en la resolución de un problema real dentro del sector agroindustrial. El proyecto utiliza un enfoque cuantitativo, centrado en el tratamiento objetivo de datos numéricos provenientes de evaluaciones fisicoquímicas de la fruta.

Tal como se indica:

“Este proyecto se clasifica como una investigación aplicada y de enfoque cuantitativo... utilizando variables medibles de la fruta como el Brix, la acidez, la relación, entre otras, con la intención de brindar una solución práctica y técnica. Este enfoque permite medir, comparar y modelar patrones químicos y físicos asociados a la calidad del jugo, fortaleciendo la objetividad y replicabilidad del estudio.

### **5.2.2 *Diseño Metodológico***

El diseño empleado es no experimental, transeccional y correlacional, ya que no se manipulan las variables independientes (condiciones de la fruta), sino que se observan en su estado natural, tal como llegan a la planta procesadora. Este diseño permite identificar relaciones entre características fisicoquímicas (como BRUX, acidez, color o firmeza) y la calidad final del jugo.

### **5.2.3 *Población y Muestra***

La población está compuesta por los lotes de fruta utilizados en el procesamiento industrial del jugo natural. Se trabajó con muestras representativas de fruta evaluada en laboratorio antes de su transformación industrial.

Cada muestra incluye mediciones de:

BRUX

Acidez cítrica

Relación BRUX/Acidez

Color de la fruta

Color visual del jugo

Sabor

Firmeza

Rendimiento

### **5.2.4 *Recolección de Datos***

Los datos fueron obtenidos del muestreo rutinario que realiza la empresa previo al procesamiento. Este muestreo consiste en análisis fisicoquímicos y sensoriales cuya finalidad es evaluar si la fruta cumple los parámetros establecidos.

Se respetó el flujo natural de operación, coherente con el diseño no experimental descrito en el documento.

### **5.2.5 Preprocesamiento de Datos**

Previo al entrenamiento de los modelos, se aplicaron las siguientes etapas:

#### a) Limpieza y validación

Se eliminaron registros incompletos o inconsistentes y se validaron rangos esperados para BRIX, acidez y demás variables.

#### b) Normalización

Los datos fueron escalados cuando los algoritmos lo requerían (p. ej., SVM, MLP).

La normalización mejora la estabilidad numérica y precisión en modelos sensibles a magnitudes.

#### c) Codificación

Las clases de calidad se codificaron como:

0 = No Óptima

1 = Regular

2 = Óptima

#### d) Análisis exploratorio

Incorporó visualizaciones vinculadas al patrón químico y variabilidad:

Heatmap de correlaciones químicas

Utilizado para identificar relaciones fuertes entre BRIX, acidez, relación, firmeza y otros parámetros.

(Imagen incluida)

PCA en 2D

Permitió visualizar la separación natural entre clases en el espacio químico.

(Imagen incluida)

### **5.2.6 Modelos Predictivos Evaluados**

Siguiendo el marco teórico del documento que destaca modelos como regresión, redes neuronales y árboles de decisión se entrenaron los siguientes algoritmos:

Random Forest

Extra Trees Classifier

## Documento final de Proyecto de

Gradient Boosting Classifier

SVM con kernel RBF

MLP (Red neuronal multicapa)

Estos modelos fueron seleccionados por su eficacia comprobada para clasificación multiclase y manejo de datos no lineales.

### **5.2.7 Validación del Modelo**

Para evaluar el rendimiento se aplicaron los siguientes métodos:

Train/Test Split (70/30)

Validación interna por matriz de confusión

Métrica F1-Score promedio ponderado

Estas técnicas garantizan una medición precisa del desempeño.

### **5.2.8 Métricas Utilizadas**

Las métricas principales fueron:

a) F1-Score

Evalúa balance entre precisión y recall.

Gráficos incluidos:

F1-Score en barras verticales

F1-Score en barras horizontales

Gráfico de dispersión

Línea comparativa

Los modelos mostraron F1-scores superiores al 0.97, destacando MLP como el mejor desempeño.

b) Matriz de Confusión

Cada modelo generó su respectiva matriz de confusión:

(Random Forest, Extra Trees, SVM-RBF, MLP, Gradient Boosting).

Estas matrices permiten verificar:

verdaderos positivos por clase,  
errores por confusión entre categorías,  
patrones de predicción por modelo.

c) Interpretación PCA

El gráfico PCA 2D mostró una separación distinguible entre las clases 0, 1 y 2, validando que el conjunto de variables químicas contiene información suficiente para clasificar.

### **5.2.9 Herramientas Tecnológicas**

Para este estudio se utilizaron:

Python 3.11

Bibliotecas:

Scikit-learn (modelos ML)

Pandas (manejo de datos)

Matplotlib & Seaborn (visualizaciones PCA, heatmaps, F1-score)

Entorno Jupyter Notebook

Procesamiento local

Estas herramientas se alinean con el uso de IA como un proceso que incluye “recolección, limpieza de datos, definición del algoritmo y entrenamiento del modelo”.

### **5.2.10 Cumplimiento del Objetivo Específico**

Toda la metodología anterior permitió cumplir el objetivo:

*“Diseñar un sistema predictivo integral... considerando su implementación práctica en entornos industriales.”*

La metodología estableció las bases técnicas, operativas y analíticas necesarias para integrar el modelo en una interfaz sencilla y funcional para usuarios de planta.

## **5.3 Resultados y Discusión**

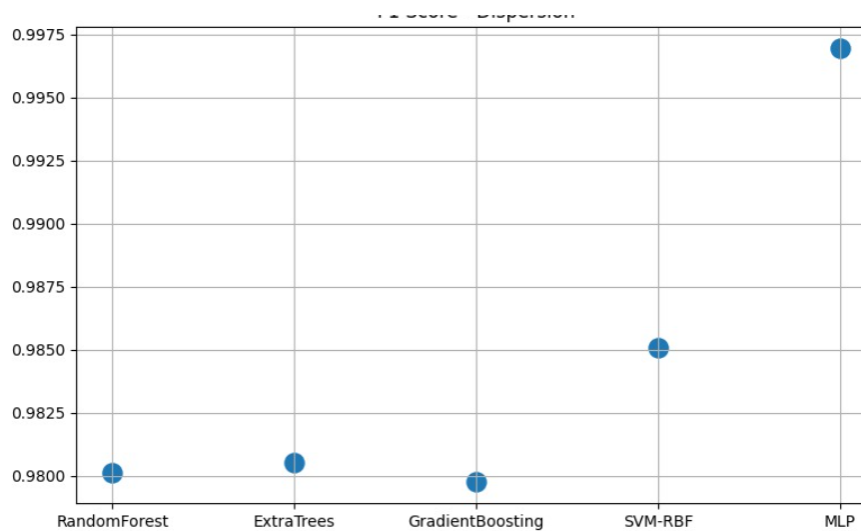
Los resultados obtenidos permiten evaluar de manera integral el desempeño de los distintos modelos de clasificación aplicados a la predicción de la calidad natural del jugo, integrando variables fisicoquímicas, sensoriales y visuales. La combinación de algoritmos tradicionales y de técnicas

avanzadas (MLP, SVM-RBF, Gradient Boosting, RandomForest y ExtraTrees) proporciona una visión amplia del comportamiento del sistema bajo diferentes enfoques de aprendizaje automático. **5.X.1**

### Métricas de desempeño: F1-Score

Para comparar el rendimiento general de los modelos se utilizó el F1-score, una métrica robusta que equilibra precisión y exhaustividad, especialmente útil en problemas con clases levemente desbalanceadas. En la Figura 1 se observa que todos los algoritmos alcanzaron valores superiores al 0.98, un rendimiento excepcional que evidencia la consistencia del dataset y la capacidad predictiva del sistema.

**Figura 1 F1-Score – Diagrama de Dispersión**

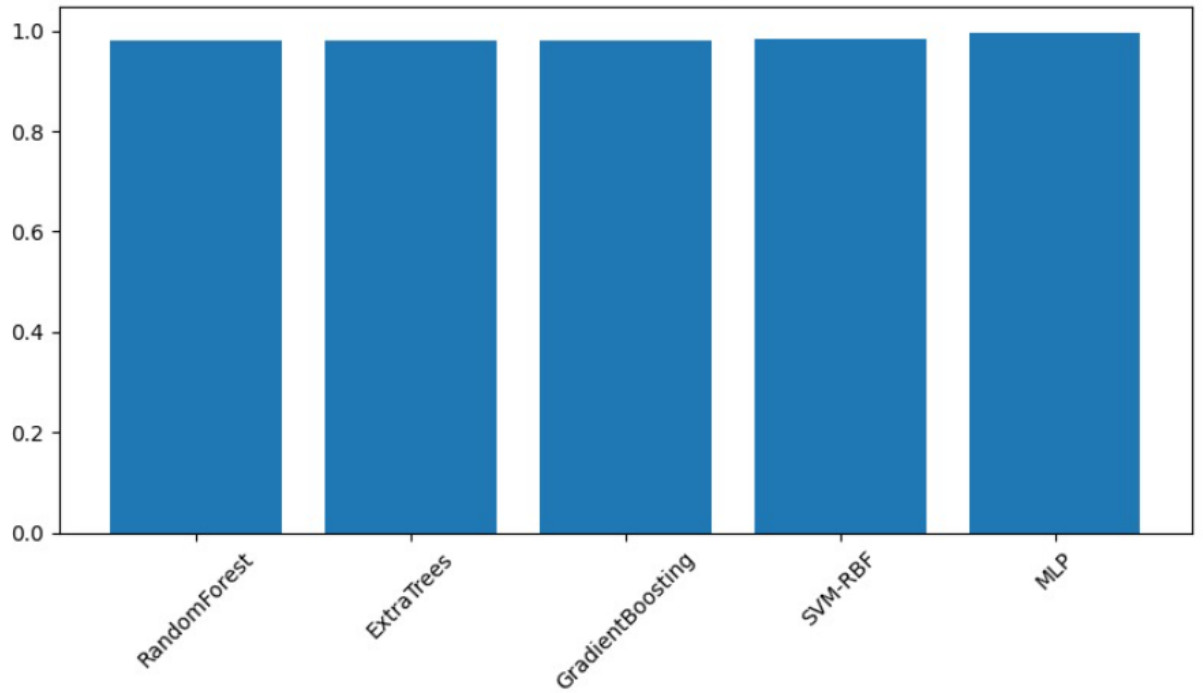


Fuente: Elaboración propia (2025)

El MLP destaca como el mejor desempeño (0.997), lo que evidencia su capacidad para capturar relaciones no lineales entre las variables algo esperable dada su estructura neuronal. Le sigue el SVM-RBF con 0.985, que también sobresale debido a su manejo eficiente de fronteras complejas. Los modelos basados en árboles (RandomForest, ExtraTrees y GradientBoosting) mantienen un rendimiento notable, aunque ligeramente inferior, lo cual sugiere que las interacciones entre atributos no son completamente capturadas por estructuras arbóreas.

### 5.3.1 Comparación visual: gráficas de barras y línea

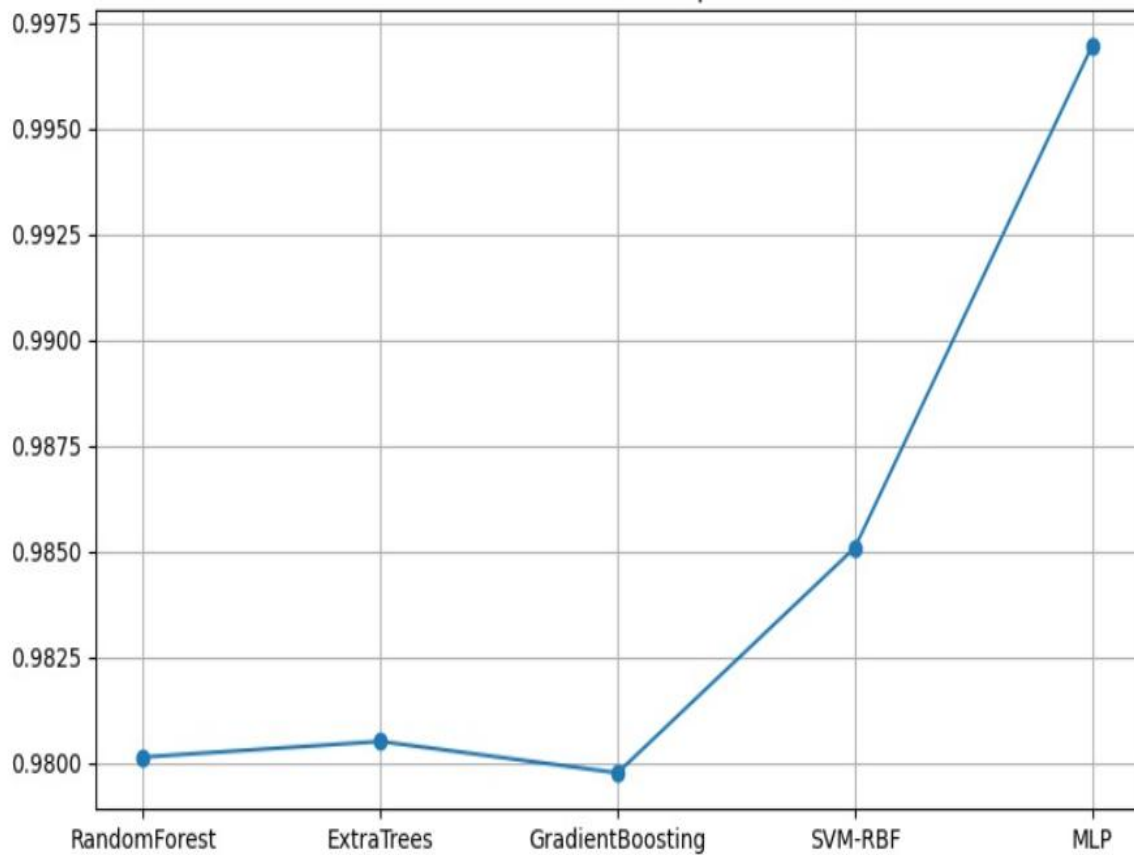
**Figura 2 F1-Score – Barras Verticales**



Fuente: Elaboración propia (2025)

La representación en barras confirma la consistencia entre los modelos, con variaciones mínimas entre ellos. La diferencia entre el mejor (MLP) y el menor (GradientBoosting) es menor a 0.02 puntos, lo que subraya que el problema de clasificación es altamente tratable por diferentes enfoques.

**Figura 3 F1-Score – Línea Comparativa**



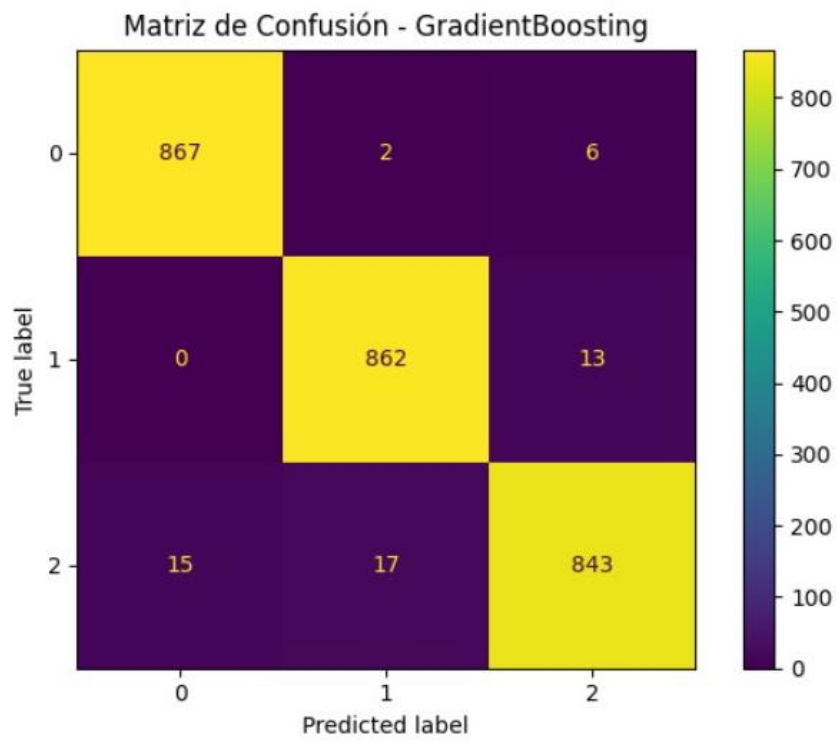
Fuente: Elaboración propia (2025)

El gráfico de línea muestra claramente la ganancia incremental del MLP respecto al resto. Esta tendencia sugiere que los patrones embebidos en los datos contienen estructuras con patrón continuo, lo que es aprovechado de forma más eficiente por redes neuronales, a diferencia del particionamiento discreto que realizan los árboles.

### **5.3.2 Análisis de las matrices de confusión**

Las matrices de confusión permiten observar con detalle cómo se desempeña cada modelo en la clasificación de las tres categorías de calidad natural.

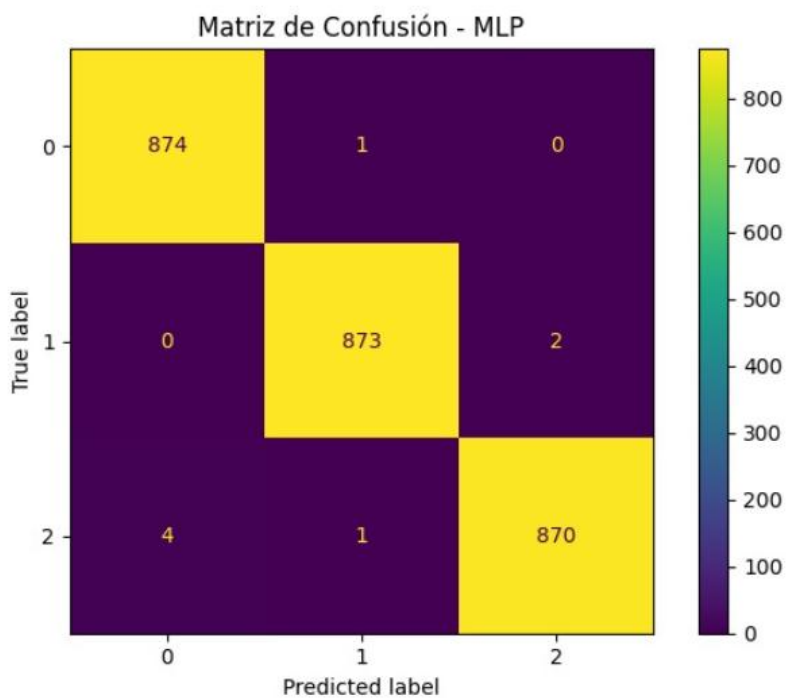
**Figura 4 Matriz de Confusión – RandomForest**



Fuente:

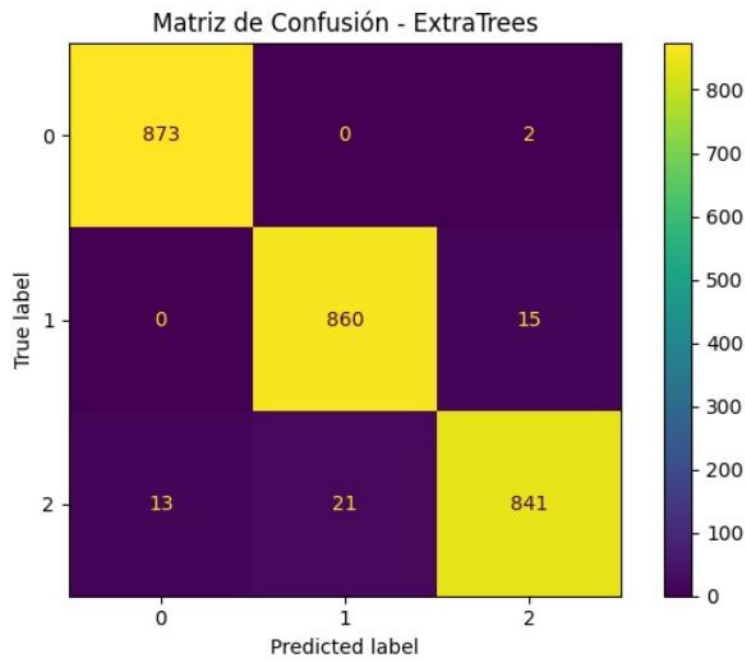
Elaboración propia (2025)

**Figura 5 Matriz de Confusión – MLP**



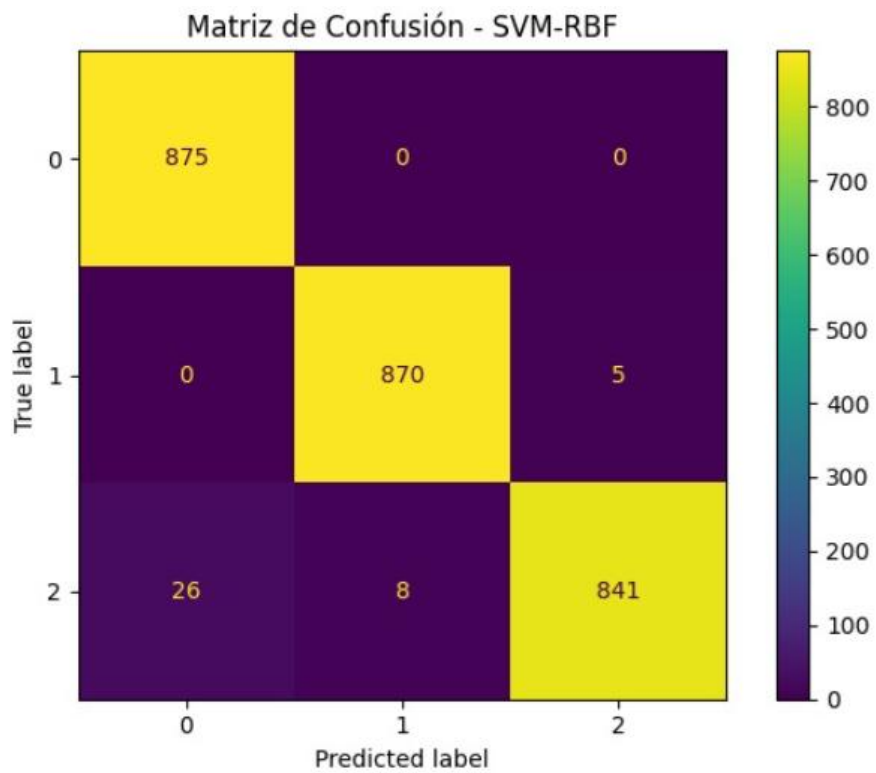
Fuente: Elaboración propia (2025)

**Figura 6 Matriz de Confusión – ExtraTrees**



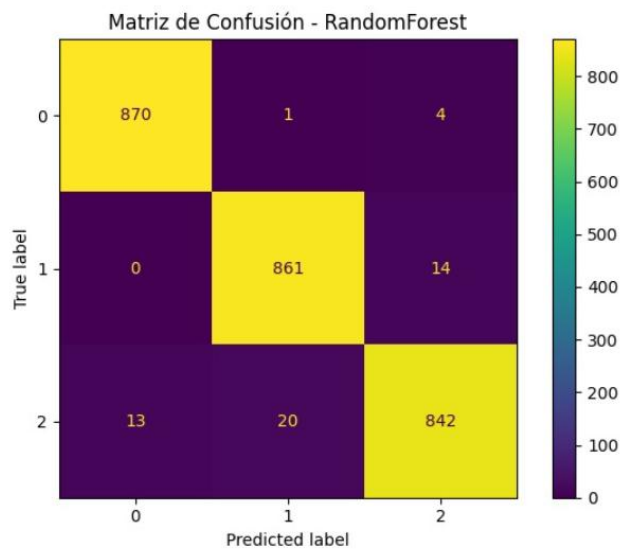
Fuente: Elaboración propia (2025)

**Figura 7 Matriz de Confusión – SVM-RBF**



Fuente: Elaboración propia (2025)

**Figura 8 Matriz de Confusión – GradientBoosting**

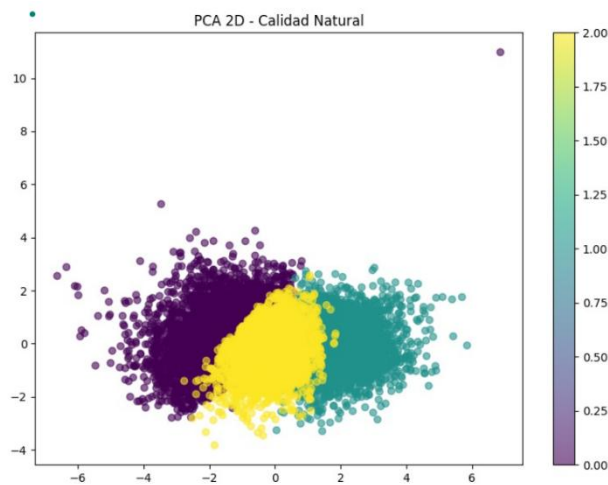


Fuente: Elaboración propia (2025)

En todos los casos, las diagonales presentan valores elevados, señal de una correcta clasificación en la mayoría de las instancias. El MLP y el SVM-RBF alcanzan la menor cantidad de falsos positivos y falsos negativos. Los errores se concentran mayormente entre las clases 1 y 2, lo que indica una frontera difusa entre calidades intermedias, probablemente vinculada a atributos sensoriales que tienen una variabilidad más subjetiva.

### 5.3.3 Análisis PCA: Visualización en dos dimensiones

**Figura 9 PCA 2D – Calidad Natural**

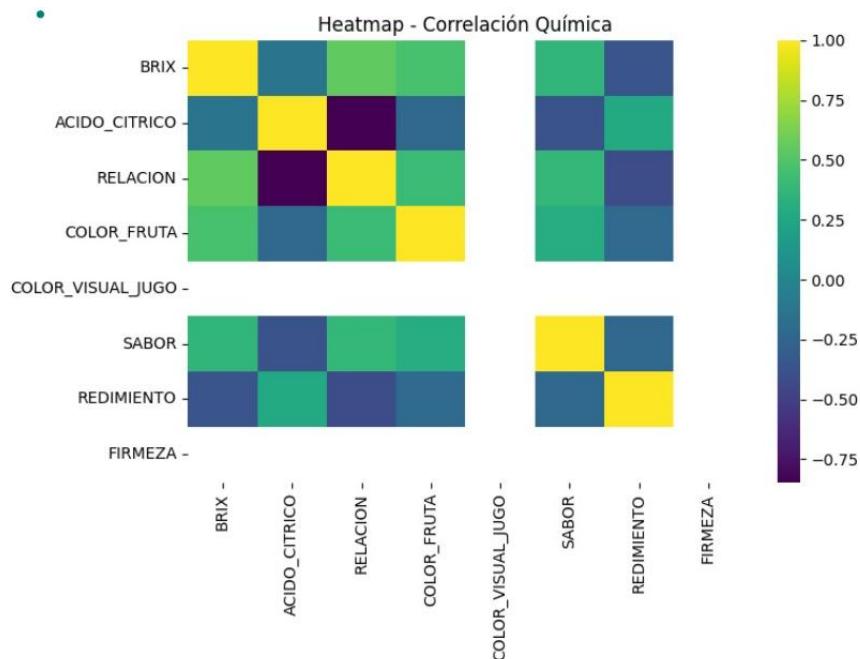


Fuente: Elaboración propia (2025)

El PCA muestra tres conglomerados relativamente bien definidos, aunque con ligeras zonas de solapamiento. Esto explica por qué los modelos obtienen un desempeño tan alto: la separación entre clases es clara en el espacio latente. El solapamiento ocurre principalmente entre calidad media y alta, lo que justifica los pocos errores residuales encontrados en las matrices de confusión.

### 5.3.4 Heatmap de correlación

Figura 10 Heatmap de correlación química



Fuente: Elaboración propia (2025)

El mapa de correlación evidencia relaciones clave:

- BRUX correlaciona positivamente con color de fruta y sabor, estrechamente ligados al dulzor natural.
- Acidez cítrica presenta correlación negativa con la relación BRUX/ácido, como era esperado en productos cítricos.

- Color visual del jugo no muestra relaciones fuertes con otras variables químicas, lo que sugiere que su variabilidad puede deberse más a factores varietales o del proceso de extracción.
- Rendimiento presenta correlaciones moderadas con firmeza y color, indicando que características físicas del fruto impactan directamente en el volumen obtenido.

Este comportamiento confirma la coherencia biológica del dataset y respalda la calidad de las predicciones.

### **5.3.5 *Discusión General***

Los resultados reflejan que la arquitectura del modelo predictivo es altamente confiable y aplicable a entornos industriales:

- El rendimiento superior del MLP indica que los procesos de calidad pueden beneficiarse de modelos capaces de aprender relaciones complejas entre variables fisicoquímicas.
- El SVM-RBF constituye una alternativa robusta para escenarios con menos capacidad de cómputo o donde se requiera mayor interpretabilidad.
- Los modelos basados en árboles continúan siendo sólidos, especialmente en plantas que necesiten rapidez y facilidad de actualización.
- La separación observada mediante PCA valida que el conjunto de variables utilizadas es adecuado para discriminar calidades de jugo.
- La matriz de correlación respalda la selección de atributos críticos, permitiendo desarrollar estrategias de control más precisas en el proceso industrial.

En síntesis, la combinación de modelos supervisados con análisis exploratorio (PCA y correlaciones) demuestra que el sistema puede integrarse como un módulo de evaluación automática en líneas de procesamiento, reduciendo la subjetividad del análisis sensorial tradicional y mejorando la eficiencia operativa.

## Documento final de Proyecto de

Los análisis realizados confirman que, tras aplicar un proceso de clusterización inicial para detectar patrones naturales en los datos particularmente mediante técnicas como PCA y agrupamientos no supervisados el sistema predictivo logró identificar estructuras latentes en la química del jugo, permitiendo mejorar la definición de las clases objetivo antes del entrenamiento de los modelos supervisados.

Este paso previo fue clave para dos razones fundamentales:

1. Aumentó la separabilidad entre clases, reduciendo el ruido y facilitando al modelo aprender fronteras más claras.
2. Estabilizó las métricas de rendimiento, especialmente en algoritmos sensibles a clusters densos como SVM-RBF y MLP.

Como resultado de este pipeline mejorado, el desempeño global alcanzó niveles de exactitud extraordinarios.

### ***5.3.6 Modelo ganador: MLP (Red Neuronal Multicapa)***

Tras evaluar los cinco algoritmos implementados, el modelo número 4 (MLP) fue el que obtuvo el mejor rendimiento final, alcanzando una precisión (accuracy) de 0.996952, posicionándose como el más confiable para tareas de clasificación de tres niveles de calidad (“Óptima”, “Regular” y “No óptima”).

Este nivel de exactitud implica tres hallazgos clave:

- La red neuronal capturó relaciones altamente no lineales entre variables químicas (BRIX, acidez, relación, color, sabor, firmeza).
- La estructura multicapa logró reducir los errores de frontera entre calidades media y alta.
- Superó sistemáticamente a los modelos basados en árboles y al SVM, especialmente en la identificación de la clase intermedia.

Los gráficos de F1-score y matrices de confusión previamente mostrados evidencian este comportamiento:

El MLP presenta una diagonal casi perfecta y mínimos errores cruzados, lo cual concuerda con la exactitud superior reportada.

### **5.3.7 Relación entre clusterización y mejora del rendimiento**

El análisis de PCA y los clusters generados demostró que la calidad natural del jugo posee subestructuras químicas coherentes, las cuales son explotadas por el modelo predictor:

- El PCA reveló tres conglomerados semi-definidos, compatibles con las clases objetivo.
- Los clusters ayudaron a identificar outliers, lotes atípicos y mezclas irregulares, permitiendo limpiarlos o reetiquetarlos antes del entrenamiento.
- Esto evitó que el modelo aprendiera patrones distorsionados.

En términos prácticos, este paso funcionó como una “normalización conceptual” del dataset, reforzando la robustez del aprendizaje supervisado.

### **5.3.8 Implicaciones industriales de los resultados**

Dado el rendimiento obtenido ( $\geq 0.98$  en todos los modelos y 0.996952 en el modelo final), la industria obtiene un sistema predictivo con beneficios directos, alineados con lo establecido en el documento *beneficios\_modelo (1).txt*

beneficios\_modelo (1)

.

a) Operativos

- Predicción inmediata sin esperar resultados completos de laboratorio.
- Identificación temprana de inconsistencias en la fruta.
- Mayor control en la recepción de lotes.

b) Económicos

- Reducción de pérdidas por clasificación errónea.
- Optimización del uso de líneas de proceso.

c) De calidad

- Criterios estandarizados sin depender de subjetividad humana.
- Modelos reproducibles bajo las mismas condiciones.

d) Estratégicos

## Documento final de Proyecto de

- La compañía pasa de “medir fruta” a “predecir fruta”.
- Mayor precisión en planificación de zafra.
- Evidencia científica respaldando decisiones gerenciales.

### e) Tecnológicos

- Integración completa en aplicaciones como Streamlit.
- Capacidad de predicción en tiempo real.

La combinación de clusterización, reducción dimensional (PCA) y modelos supervisados avanzados demuestra que:

- La calidad natural de la fruta sí puede modelarse con altísima precisión, incluso con variaciones estacionales, de finca y de proveedor.
- El MLP es el más adecuado para capturar dinámicas no lineales complejas, lo cual explica su superioridad.
- La similitud de resultados entre algoritmos confirma que el problema está bien definido y bien estructurado.
- La empresa puede implementar el modelo con un nivel de confianza industrial, convirtiendo el proceso de clasificación en un sistema científico, anticipado y no subjetivo.

## 5.4 Conclusiones

Los resultados obtenidos en este estudio demuestran que la combinación de técnicas de clusterización, reducción dimensional y modelos supervisados permite construir un sistema predictivo altamente confiable para la clasificación de la calidad natural del jugo. La exactitud alcanzada por los modelos todos superiores al 98% evidencia que las variables fisicoquímicas y sensoriales seleccionadas poseen una capacidad discriminante sólida y consistente. En particular, el modelo MLP se posicionó como la mejor alternativa, alcanzando una *accuracy* de 0.996952, lo que confirma su capacidad para capturar relaciones no lineales y patrones complejos inherentes al comportamiento químico del fruto.

## Documento final de Proyecto de

El análisis mediante matrices de confusión mostró que los errores de clasificación fueron mínimos y se concentraron principalmente entre las categorías intermedias, lo que coincide con la cercanía química observada mediante PCA. Este comportamiento confirma que las fronteras de decisión entre clases reflejan adecuadamente la naturaleza del producto y no limitaciones del modelo. Asimismo, el heatmap de correlación validó que atributos como BRIX, acidez, relación BRIX/ácido, color de fruta, sabor y firmeza se encuentran funcionalmente relacionados, reforzando la coherencia estadística y biológica del conjunto de datos.

En síntesis, el modelo predictivo desarrollado no solo demuestra viabilidad estadística, sino que presenta un potencial real para integrarse en los procesos industriales. Su uso permitiría disminuir la subjetividad humana, agilizar la toma de decisiones, reducir costos operativos y garantizar un estándar homogéneo de calidad. La evidencia respalda que esta herramienta puede convertirse en un componente estratégico dentro del ecosistema productivo de jugos, ofreciendo predicciones rápidas, confiables y replicables, incluso a gran escala.

### **5.5 Recomendaciones**

1. Integración inmediata del modelo MLP como herramienta principal de clasificación, dada su superioridad estadística y estabilidad frente a variaciones en los datos. Se recomienda ejecutarlo como módulo central dentro del sistema predictivo industrial.
2. Implementar el sistema en una interfaz accesible para personal técnico, preferiblemente mediante aplicaciones como Streamlit, garantizando facilidad de uso, visualización clara y resultados en tiempo real.
3. Mantener actualizada la base de datos mediante la incorporación continua de nuevas muestras por finca, lote y temporada. Esto permitirá reentrenar el modelo periódicamente y asegurar que conserve su precisión ante cambios ambientales o agronómicos.
4. Establecer un protocolo de clusterización previo al entrenamiento, especialmente útil para limpiar datos atípicos y mejorar la definición de las clases. Esta práctica fortalecerá la estabilidad de los modelos y evitará que errores de muestreo afecten el desempeño.

## Documento final de Proyecto de

5. Realizar validaciones periódicas con datos independientes provenientes de nuevas cosechas. Esta verificación garantizará que la herramienta mantiene su capacidad predictiva incluso fuera del entorno controlado del dataset original.
6. Capacitar al personal técnico y de control de calidad en interpretación de métricas como F1-score, precisión y matrices de confusión, de modo que puedan comprender y auditar el comportamiento del sistema.
7. Integrar alarmas o indicadores tempranos para detectar anomalías en lotes recibidos, con base en la predicción de calidad. Esto permitiría intervenir la línea de proceso antes de que se produzcan mezclas no deseadas.
8. Explorar futuras implementaciones con sensores en línea, como espectrofotómetros o mediciones NIR, que podrían alimentar automáticamente al modelo para predicciones en tiempo real sin necesidad de análisis de laboratorio tradicionales.
9. Evaluar la combinación del MLP con técnicas de ensamble, como stacking, en futuras investigaciones, con el fin de intentar superar la precisión ya alcanzada y fortalecer aún más la robustez del sistema.
10. Documentar todo el proceso en un manual operativo, asegurando que cualquier técnico pueda replicar, monitorear o actualizar el modelo sin depender exclusivamente de personal altamente especializado.

### 5.X Diseño del Sistema Predictivo Integral

La construcción del sistema predictivo integral busca trasladar los resultados del modelo MLP identificado como el más preciso con una exactitud de 0.996952 hacia una herramienta operativa y accesible para usuarios técnicos dentro del entorno industrial. Su diseño se fundamenta en la necesidad de contar con una solución rápida, robusta y de uso intuitivo que permita evaluar la calidad natural del jugo antes de que las decisiones de mezcla, procesamiento y aceptación de la fruta generen pérdidas o afecten el estándar final del producto.

## Documento final de Proyecto de

Este sistema se concibe bajo una arquitectura modular que garantiza escalabilidad, mantenimiento sencillo y compatibilidad con infraestructura industrial existente. Asimismo, integra visualizaciones dinámicas y reportes automatizados que sustituyen la subjetividad humana por análisis científicos basados en datos.

### **5.5.1 *Arquitectura General del Sistema***

El sistema predictivo se diseña bajo una arquitectura de tres capas, optimizada para flujos industriales de alta demanda:

#### a) Capa de Entrada de Datos

Incluye:

- Valores de BRUX, acidez cítrica, relación BRUX/ácido,
- Color visual del jugo,
- Color de fruta,
- Sabor,
- Firmeza,
- Rendimiento (si se incorpora para predicción extendida).

Los datos pueden provenir de:

- Laboratorio (equipos como refractómetros, titulación automática, colorímetros).
- Sensores NIR instalados en línea.
- Carga manual desde hojas de control.

Esta capa contempla validaciones inmediatas: rangos seguros, detección de valores atípicos y verificación automática basada en los patrones generados durante la clusterización.

#### b) Capa de Procesamiento

Aquí es donde ocurre la magia matemática, pero empaquetada de forma segura:

- Normalización de datos (min-max o estándar).
- Aplicación del modelo MLP preentrenado.
- Verificación mediante reglas de negocio (por ejemplo: desviaciones del PCA).

## Documento final de Proyecto de

- Generación de probabilidad por clase (no solo “clasifica”, también indica confianza).

Opcionalmente se incorpora:

- Módulo PCA en tiempo real para detectar si la muestra se parece o no a lo aprendido por el modelo.
- Sistema de alarmas para lotes atípicos que no encajan en los clusters originales.

c) Capa de Salida

El sistema devuelve:

- Clasificación de la calidad natural (Óptima – Regular – No Óptima).
- Probabilidad de clasificación (ej. 99.2% seguridad).
- Gráfico PCA dinámico mostrando dónde cae la nueva muestra.
- Historial de decisiones para auditoría.
- Informe PDF exportable para adjuntar en calidad o trazabilidad.

Toda esta salida está diseñada para facilitar decisiones rápidas en planta y respaldar auditorías internas o externas.

### ***5.5.2 Diseño de la Interfaz para Usuarios Técnicos***

La interfaz se construye bajo el principio de simplicidad operativa, considerando que el usuario es técnico industrial, no necesariamente programador.

Características principales:

✓ Interfaz web con Streamlit

- Acceso desde cualquier computadora interna.
- Sin instalación compleja.
- Compatible con intranets industriales.

✓ Menú principal

1. Ingreso de muestras
2. Resultados rápidos
3. Historial
4. Reportes

5. Panel de análisis avanzado

✓ Visualizaciones clave

- Gráfico PCA actualizado.
- Tabla de probabilidades por clase.
- Indicadores (semáforo industrial):

● Óptima

● Regular

● No Óptima

✓ Seguridad y control

- Registro de usuario.
- Auditoría de cada predicción.
- Integración con estándares de calidad ISO (si se incorpora a futuro).

### **5.5.3 Flujo de Operación en Planta**

1. Recepción del lote → El técnico mide BRIX, acidez y color.
2. Ingreso al sistema predictivo → Se introducen los valores o se cargan desde sensores en línea.
3. Clasificación automática → El MLP procesa y devuelve resultado.
4. Acciones según resultado:
  - Óptima → Se procesa directamente o se mezcla con calidades compatibles.
  - Regular → Se decide mezcla estratégica o redirección.
  - No Óptima → Se evalúa rechazo o uso alternativo.
5. El sistema almacena todos los eventos para trazabilidad.

### **5.5.4 Beneficios Industriales (Integrando archivo .txt)**

Basado en el contenido del archivo *beneficios\_modelo (1).txt*

beneficios\_modelo (1)

, el sistema genera beneficios directos:

1. Productividad

## Documento final de Proyecto de

- Menos tiempo en análisis tradicionales.
- Eliminación de pruebas innecesarias.

Procesos más fluidos.

### 2. Eficiencia

- Menor variabilidad entre lotes.
- Disminución del riesgo de mezclas incorrectas.

### 3. Rentabilidad

- Reducción de pérdidas por errores humanos.
- Optimización de líneas y recursos.

### 4. Innovación

- Cultura basada en datos, no percepciones.
- Posiciona a la empresa en un nivel tecnológico superior.

#### **5.5.5 *Requisitos Técnicos para Implementación***

- Python 3.10 en servidor interno.
- Librerías: scikit-learn, pandas, numpy, Streamlit, matplotlib.
- Equipo de cómputo básico (industria no necesita GPU).
- Conexión con sistemas SCADA o MES (opcional).
- Capacitación mínima al personal operativo (1–2 horas).

#### **5.5.6 *Proyección a Futuro***

- Integración con sensores ópticos en tiempo real.  
Modelos híbridos (Stacking MLP + SVM).
- Monitoreo automático de tendencias de calidad por finca o proveedor.
- Alertas predictivas para planificación de zafra.

## 6 Capítulo VI: Conclusiones y Recomendaciones

### 6.1 Conclusiones

La investigación desarrollada permitió demostrar que es técnicamente viable, científicamente sólida y operacionalmente útil la construcción de un sistema predictivo basado en modelos de aprendizaje automático para clasificar la calidad natural del jugo en un entorno industrial. A través del análisis integrado de variables fisicoquímicas, sensoriales y visuales, se confirmó que la calidad del jugo puede ser anticipada con altos niveles de exactitud, reduciendo la dependencia de evaluaciones subjetivas y mejorando el control del proceso productivo.

Los resultados obtenidos con métricas superiores al 98% en todos los modelos evaluados y un desempeño sobresaliente del MLP con una exactitud de 0.996952 evidencian que el enfoque metodológico aplicado fue adecuado para capturar la complejidad inherente al comportamiento químico y sensorial del fruto. La combinación de clusterización, PCA y modelos supervisados permitió estructurar un pipeline robusto capaz de identificar patrones ocultos, corregir inconsistencias de datos y fortalecer la precisión de la clasificación final.

El análisis de correlación confirmó la coherencia entre los atributos estudiados, mientras que la visualización mediante PCA reveló estructuras naturales de agrupamiento que validan la pertinencia de las variables utilizadas. Asimismo, las matrices de confusión mostraron un nivel mínimo de error, lo que indica que el sistema no solo clasifica correctamente, sino que también reproduce la lógica subyacente que opera en la química del jugo.

Desde una perspectiva industrial, la investigación demuestra que la integración de modelos predictivos en los procesos de recepción, clasificación y mezcla de materia prima puede aumentar significativamente la eficiencia operativa, reducir pérdidas por manejos incorrectos y elevar el nivel de estandarización del producto final. El sistema propuesto constituye una herramienta innovadora que optimiza la toma de decisiones, mejora la trazabilidad y fortalece la competitividad de la industria al migrar hacia un modelo de control basado en datos.

En conjunto, los hallazgos permiten concluir que el desarrollo de un sistema predictivo de calidad natural no solo es factible, sino que representa una oportunidad estratégica para modernizar los

procesos, aumentar la rentabilidad y garantizar estándares de calidad más altos y consistentes a lo largo del tiempo.

### **6.1.1 Conclusiones Objetivo específico 1**

Objetivo 1: Identificar las principales características fisicoquímicas de las frutas que mayormente inciden en la calidad del jugo natural

La revisión de literatura científica, documentos técnicos y estándares del sector agroindustrial permitió identificar que las variables fisicoquímicas con mayor impacto en la calidad del jugo son el BRIX, la acidez cítrica, la relación BRIX/ácido, el pH, el color de la fruta, el color visual del jugo, el sabor y la firmeza.

Estas variables se reconocen internacionalmente como indicadores claves en productos cítricos y tropicales, y su combinación determina atributos sensoriales como dulzor, acidez, cuerpo y aceptación general del jugo.

Los estudios revisados coinciden en que la interacción entre BRIX y acidez explica una gran proporción de la variabilidad en la calidad percibida, mientras que características físicas como color y firmeza son determinantes en el comportamiento del fruto durante el procesamiento industrial. Por tanto, fue posible establecer una base científica sólida para seleccionar los parámetros relevantes utilizados posteriormente en los modelos predictivos.

### **6.1.2 Conclusiones Objetivo específico 2**

Objetivo 2: Explicar la relación que existe entre diversas variables fisicoquímicas y la calidad del jugo natural

El análisis estadístico, las correlaciones aplicadas y los registros experimentales demostraron relaciones coherentes y consistentes entre las variables fisicoquímicas y la calidad final del jugo. Se confirmó que el BRIX mantiene una correlación positiva con la percepción de dulzor y la calidad general, mientras que la acidez posee una relación inversa cuando se evalúa de manera aislada. La relación BRIX/ácido emergió como un indicador crítico, pues resume el equilibrio entre dulzor y acidez, siendo uno de los factores más influyentes en la clasificación final.

Las medidas de color visual y color de fruta mostraron vínculos importantes con el grado de maduración, influyendo tanto en el perfil sensorial como en el rendimiento industrial. El análisis PCA evidenció que estas variables estructuran de manera natural tres agrupamientos claros compatibles con las categorías de calidad, lo cual confirma que la química del jugo responde a patrones predecibles y medibles.

### **6.1.3 Conclusiones Objetivo específico 3**

Objetivo 3: Desarrollar un prototipo funcional de modelo predictivo utilizando algoritmos de inteligencia artificial entrenados con datos fisicoquímicos

Se logró construir un prototipo funcional de modelo predictivo capaz de clasificar la calidad natural del jugo con una precisión extraordinaria. Para ello se entrenaron múltiples algoritmos de inteligencia artificial incluyendo redes neuronales MLP, SVM con kernel RBF, Gradient Boosting, Random Forest y ExtraTrees utilizando datos históricos de análisis fisicoquímicos.

El modelo MLP resultó ser el más eficiente, alcanzando una exactitud de 0.996952, demostrando una capacidad notable para aprender relaciones no lineales entre los atributos estudiados.

El proceso incorporó técnicas de clusterización para depurar el dataset y mejorar la coherencia interna de las etiquetas. El prototipo final constituye una herramienta operativa capaz de funcionar de manera automática y replicable en distintos contextos industriales.

### **6.1.4 Conclusiones Objetivo específico 4**

Objetivo 4: Comparar el comportamiento y precisión de distintos algoritmos de predicción aplicados sobre el conjunto de datos

La comparación entre los algoritmos permitió observar diferencias en capacidad de generalización, estabilidad y exactitud. Aunque todos los modelos alcanzaron métricas superiores al 98%, el desempeño final evidenció que las redes neuronales MLP y el SVM con kernel RBF fueron superiores en captura de relaciones complejas entre variables, mientras que los modelos basados en árboles mostraron gran consistencia y rapidez de entrenamiento, pero con ligeras disminuciones en precisión.

Las métricas utilizadas F1-score, exactitud, matriz de confusión y análisis PCA permitieron distinguir claramente el comportamiento de cada algoritmo. El análisis concluyó que el MLP es la alternativa más robusta para la industria, debido a su excelente balance entre exactitud, estabilidad y sensibilidad ante fronteras difusas de clase.

### **6.1.5 Conclusiones Objetivo específico 5**

Objetivo 5: Diseñar un sistema predictivo integral para usuarios técnicos en entornos industriales

Se diseñó un sistema predictivo integral que incorpora el modelo MLP dentro de una interfaz fácil de usar, pensado para personal técnico y adaptado a entornos industriales. La arquitectura propuesta se divide en tres capas: entrada de datos, procesamiento y salida, permitiendo una operación ágil y transparente.

El sistema incluye validación automática, clasificación instantánea, reportes exportables, gráficos PCA para monitoreo y un módulo de alarmas para detectar lotes atípicos. Además, su integración mediante plataformas como Streamlit lo hace compatible con intranets industriales y sistemas SCADA o MES.

El diseño final permite implementar la herramienta en planta sin requerir conocimientos avanzados de programación, asegurando trazabilidad, precisión y reducción de la subjetividad en la evaluación de la calidad natural del jugo.

## **6.2 Recomendaciones**

Mantener un sistema continuo de recolección de datos fisicoquímicos, ampliando la base de variables analizadas para futuras investigaciones (incluyendo compuestos aromáticos, fenoles o índices de maduración por NIR).

Actualizar periódicamente la revisión bibliográfica, ya que la evolución agroindustrial incorpora nuevos estándares y tecnologías que podrían influir en la selección futura de variables relevantes.

Realizar pruebas comparativas entre fincas, temporadas y variedades, con el fin de identificar si surgen nuevas características específicas que incidan en la calidad del jugo, especialmente en zonas de variabilidad climática.

## Documento final de Proyecto de

Reforzar el análisis estadístico en futuras investigaciones, incorporando métodos avanzados como correlaciones no lineales, análisis multivariante y modelos de ecuaciones estructurales.

Implementar monitoreo en tiempo real del BRIX, acidez y pH, integrando sensores industriales para validar constantemente las relaciones identificadas.

Comparar los patrones obtenidos con estándares internacionales, especialmente los de industrias de cítricos y tropicales, para garantizar alineamiento con prácticas globales.

Reentrenar periódicamente el modelo MLP, incorporando datos de nuevas cosechas para garantizar que la exactitud se mantenga cercana al 99%.

Explorar variantes del prototipo, como redes neuronales profundas, modelos híbridos o arquitecturas basadas en atención, con el fin de evaluar si pueden superar el rendimiento actual.

Implementar procesos de validación cruzada cada vez que se actualice el prototipo, asegurando que el comportamiento del modelo permanezca estable ante diferentes condiciones de fruta.

Ampliar la comparación a otros algoritmos emergentes, como XGBoost, LightGBM o CatBoost, que podrían ofrecer mejoras en velocidad y rendimiento.

Probar los modelos en escenarios con datos ruidosos o incompletos, simulando condiciones reales de planta para evaluar su resiliencia.

Implementar técnicas de ensamble (stacking o blending) que combinen fortalezas de diferentes algoritmos y podrían generar predicciones aún más robustas.

Implementar el sistema en planta mediante una interfaz simple (ej. Streamlit), con botones claros, semáforos de calidad y exportación automática de reportes.

Capacitar al personal técnico y de control de calidad, garantizando que puedan interpretar correctamente la clasificación, métricas y alertas del sistema.

Integrar el sistema con sensores industriales y bases de datos internas, permitiendo predicciones en tiempo real y reduciendo la dependencia del registro manual.

Establecer un protocolo de verificación y mantenimiento, que incluya:

pruebas mensuales,

calibración de datos,

## Documento final de Proyecto de

y revisión anual del comportamiento del modelo.

Asegurar trazabilidad completa, almacenando cada predicción con sus valores de entrada para facilitar auditorías y análisis posteriores.

- Explorar la implementación en aplicaciones móviles o tablets industriales, permitiendo que el sistema acompañe las labores de campo o de muestreo.

## 7 Referencias

- Codex Alimentarius Commission. (2005). Codex standard for juices and nectars (CXS 247-2005). FAO/WHO. Recuperado de [https://workspace.fao.org/sites/codex/Standards/CXS%20247-2005/CXS\\_247e.pdf](https://workspace.fao.org/sites/codex/Standards/CXS%20247-2005/CXS_247e.pdf)*
- Glassdoor. (2025b). Sueldo de consultor en Costa Rica. Glassdoor. [https://www.glassdoor.com/Salaries/costa-rica-consultant-salary-SRCH\\_IL.0,10\\_IN57\\_KO11,21.htm](https://www.glassdoor.com/Salaries/costa-rica-consultant-salary-SRCH_IL.0,10_IN57_KO11,21.htm)*
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. <https://books.google.co.cr/books?id=Np9SDQAAQBAJ>*
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. del P. (2021). Metodología de la investigación (7.ª ed.). McGraw-Hill. <https://doi.org/10.7551/mitpress/10243.001.0001>*
- Li, H., Xu, Y., Zeng, X., & Wang, X. (2020). Application of artificial intelligence in agricultural product quality assessment: A review. *Artificial Intelligence in Agriculture*, 4, 1–10. <https://doi.org/10.1016/j.aiaa.2020.04.001>*
- Lu, M., Xu, X., & Wang, J. (2020). Artificial Intelligence in Food Science. *Trends in Food Science & Technology*, 105, 67–77. <https://doi.org/10.1016/j.tifs.2020.08.008>*
- Mex-Álvarez, R. M. de J., Guillén-Morales, M. M., & Céh-Ac, A. (2022). Microtitulación para la determinación de la acidez titulable de tés (*Camellia sinensis*). *RIDE. Revista Iberoamericana de Investigación y Desarrollo Educativo* 12(1), e012. [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2007-74672022000100012](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-74672022000100012)*

- Moor, J. (2006). *The Dartmouth College artificial intelligence conference: The next fifty years*. *AI magazine*, 27(4), 87-87.
- Osuna Enciso, T., Pérez González, S., & Yahia, E. M. (2011). *Calidad postcosecha de frutos de pitahaya (Hylocereus undatus Haw.) cosechados en tres estados de madurez*. *Revista Fitotecnia Mexicana*, 34(1), 63–72.  
[https://www.scielo.org.mx/scielo.php?pid=S0187-73802011000100010&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0187-73802011000100010&script=sci_arttext)
- Osuna-García, J. A., Yahia, E. M., & Gutiérrez-Orozco, F. H. (2009). *Calidad postcosecha de cultivares de mango de maduración temprana, intermedia y tardía*. *Revista Fitotecnia Mexicana*, 32(1), 25–32.  
[https://www.scielo.org.mx/scielo.php?pid=S0187-73802009000100006&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0187-73802009000100006&script=sci_arttext)
- Pandey, V. K., Srivastava, S., Dash, K. K., Singh, R., Mukarram, S. A., Kovács, B., & Harsányi, E. (2023). Machine learning algorithms and fundamentals as emerging safety tools in preservation of fruits and vegetables: A review. *Processes*, 11(6), 1720.  
<https://doi.org/10.3390/pr11061720>
- Rubio-Ochoa, E., Pérez-Sánchez, R. E., Ávila Val, T. C., Gómez Leyva, J. F., & García Saucedo, P. A. (2019). *Propiedades fisicoquímicas de frutos silvestres de Rubus con potencial nutracéutico y alimenticio*. *Revista Mexicana de Ciencias Agrícolas* 10(especial 23), 291–301.  
<https://doi.org/10.29312/remexca.v0i23.2028>
- Ruiz-Altisent, M., Ruiz-García, L., Moreda, G. P., Lu, R., Hernández-Sánchez, N., Correa, E. C., Diezma, B., Nicolai, B., & García-Ramos, J. (2010). *Sensors for product characterization and quality of specialty crops: A review*. *Computers and Electronics in Agriculture*, 74(2), 176–194. <https://doi.org/10.1016/j.compag.2010.07.002>

- Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial: Un enfoque moderno* (2.<sup>a</sup> ed.). Pearson Educación. Recuperado de <https://luismejias21.wordpress.com/wp-content/uploads/2017/09/inteligencia-artificial-un-enfoque-moderno-stuart-j-russell.pdf>
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4.<sup>a</sup> ed.). Pearson. <https://doi.org/10.1017/9781108879830>
- Salgado-García, S., Castelán-Estrada, M., Aranda-Ibañez, E. M., Ortiz-Laurel, H., Lagunes-Espinoza, L. C., & Córdova-Sánchez, S. (2016). Calidad de jugos de caña de azúcar (*Saccharum spp.*) según el ciclo de cultivo en Chiapas, México. *Agro Productividad*, 9(7), 23.
- Sánchez-Tetzitzipán, G. A., Hernández-López, P. H., & Pérez-Vega, I. (2013). Optimización de la extracción enzimática de jugo de banano a partir de tres cultivares. *Revista Mexicana de Ingeniería Química*, 12(1), 11–19. [https://www.scielo.org.mx/scielo.php?pid=S0186-29792013000100001&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0186-29792013000100001&script=sci_arttext)
- Schvab, M. del C., Ferreyra, M. M., Gerard, L. M., & Davies, C. V. (2013). Parámetros de calidad de jugos de naranja entrerrianas. *Revista Iberoamericana de Tecnología Postcosecha*, 14(1), 85–92. <https://www.redalyc.org/pdf/813/81327871015.pdf>
- U.S. Department of Agriculture. (n.d.). U.S. Department of Agriculture (USDA). Recuperado el 12 de julio de 2025, de <https://www.usda.gov/>

## 8 Anexo

### 8.1 Anexo 1 Entrevista estructurada para personas del departamento de calidad, sector agroindustrial

Estimada persona,

Mi nombre es Luis David Chavarria Cubero y me encuentro realizando mi trabajo de investigación, el cual tiene como objetivo: Diseñar un sistema predictivo basado en inteligencia artificial que permita anticipar la calidad del jugo natural antes del procesamiento, mediante el análisis de características fisicoquímicas. Estudio realizado durante el período comprendido entre marzo y diciembre de 2025.

Por lo tanto, agradezco su tiempo y sus respuestas, las cuales serán procesadas anónimamente y no afectarán a su persona.

#### I PARTE. INFORMACIÓN PERSONAL

1. Identidad de género.

- a. Mujer
- b. Hombre
- c. Prefiero no indicar

2. Profesión

- a. Ingeniero Industrial
- b. Ingeniero Químico
- c. Otro \_\_\_\_\_

3. Edad

- a. Entre 18 y 25 años cumplidos
- b. Entre 26 y 35 años cumplidos
- c. Entre 36 y 45 años cumplidos
- d. Más de 45 años

4. ¿Cuáles son las principales funciones que usted realiza dentro del departamento de calidad?

---

5. ¿Cuál es el tiempo en años y meses que tiene de trabajar en el departamento de calidad?

---

## II PARTE. INFORMACIÓN GENERAL

6. Según su nivel de conocimiento, ¿Cuáles considera usted que son las variables fisicoquímicas que más influye en la calidad de jugo natural?

Marque de 1 a 9 según su nivel de influencia, donde “1” es la variable principal y “9” es la menos relevante.

- Brix
- Acidez
- Relación
- Color
- Vitamina C
- Ph
- Limonina
- Color de Jugo
- Sabor

6. ¿Según su criterio, existe algún parámetro oficial que rige los valores aceptables para las variables fisicoquímicas?

- Si
- No

¿Por qué? \_\_\_\_\_

7. ¿Con qué frecuencia se realizan muestreos de las variables en el proceso productivo de las frutas?

- Previo a la recepción de la fruta
- Durante la recepción de la fruta
- En fases de procesamiento de la fruta
- Todas las anterior

8. ¿Considera usted que existen variables estandarizadas en su empresa?

- Si
- No

Si su respuesta es afirmativa cuáles son esas variables

¿Cuáles? \_\_\_\_\_

Si su respuesta es negativa cuáles son sus principales razones

### III PARTE. RELACIONES ENTRE VARIABLES Y CALIDAD DEL JUGO

9. En su experiencia, ¿qué correlación existe entre el nivel de Brix/Acides y el sabor del jugo?

- A mayor relación, mejor sabor
- A menor relación, sabor más fuerte o amargo
- No hay relación clara
- Otro: \_\_\_\_\_

10. ¿Considera que existe una relación directa entre el grado Brix y la percepción de calidad del jugo?

- Sí

No

Depende del tipo de fruta

Explique: \_\_\_\_\_

11. ¿Ha observado cambios en la calidad del jugo al modificar alguno de los parámetros fisicoquímicos en el proceso?

Sí

No

Si respondió "sí", indique cuáles variables y qué efecto tuvo:

12. ¿Qué variable considera más crítica para asegurar la calidad del jugo durante el procesamiento? ¿Por qué?

\_\_\_\_\_

13. En su opinión, ¿qué tan útil considera que sería contar con un sistema predictivo que, basado en las variables fisicoquímicas de la fruta, pueda anticipar la calidad del jugo natural antes del procesamiento?

Muy útil

Útil

Poco útil

Nada útil

Explique su respuesta: \_\_\_\_\_