

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**Sistema de predicción de calidad del producto en el área de producción de
una empresa manufacturera de la región Lambayeque**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Josue Giancarlos Cumpa Chancafe

ASESOR

Karla Cecilia Reyes Burgos

<https://orcid.org/0000-0003-3520-5076>

Chiclayo, 2025

**Sistema de predicción de calidad del producto en el área de
producción de una empresa manufacturera de la región
Lambayeque**

PRESENTADA POR
Josue Giancarlos Cumpa Chancafe

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Guadalupe Teresa Lip Curo
PRESIDENTE

Segundo Jose Castillo Zumaran
SECRETARIO

Karla Cecilia Reyes Burgos
VOCAL

Dedicatoria

A Dios, por ser mi guía constante y la fuente de fortaleza que me ha permitido avanzar y cumplir cada uno de mis objetivos.

A mis padres, Jorge y Luz, por su amor, comprensión y por enseñarme, día a día, los valores que me han hecho crecer como persona. Su apoyo incondicional ha sido la base sobre la cual he podido construir y culminar esta etapa.

Agradecimientos

Agradezco a la universidad por brindarme una experiencia de formación profesional enriquecedora, que ha sido fundamental en mi desarrollo académico y personal.

Expreso mi más sincero agradecimiento a mi asesora, Reyes Burgos Karla, por su guía experta y su apoyo constante durante todo este proceso. Su orientación y experiencia han sido pilares clave en la realización de este trabajo.

Gracias también a la empresa donde desarrollé la presente investigación, por permitirme aplicar mis conocimientos en un entorno real. Su colaboración y los recursos proporcionados fueron indispensables para llevar a cabo este proyecto con éxito.

Sistema de predicción de calidad del producto en el área de producción de una empresa manufacturera de la región Lambayeque

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	tesis.usat.edu.pe Fuente de Internet	6%
2	repositorio.utn.edu.ec Fuente de Internet	1%
3	revistas.ubiobio.cl Fuente de Internet	1%
4	revistas.ulasalle.edu.pe Fuente de Internet	1%
5	hdl.handle.net Fuente de Internet	1%
6	Submitted to Universidad Internacional de la Rioja Trabajo del estudiante	1%
7	Submitted to utn Trabajo del estudiante	1%
8	tesis.pucp.edu.pe Fuente de Internet	1%

Índice

Resumen	6
Abstract	7
Introducción	8
Revisión de literatura	11
Materiales y métodos	17
Resultados y discusión	19
Conclusiones	34
Recomendaciones.....	35
Referencias	36
Anexos.....	39

Resumen

La presente investigación se centra en el desarrollo de un sistema de predicción de calidad para la producción de bandas modulares en una empresa manufacturera de Lambayeque. La metodología empleada fue CRISP-DM, permitiendo una estructuración eficiente desde la comprensión del problema hasta la implementación de soluciones. Se utilizó un data set de 3469 registros, de los cuales el 80% se destinó para el entrenamiento del modelo, garantizando así una base sólida para su desarrollo. Para la construcción del modelo predictivo, se utilizó la inteligencia artificial, específicamente redes neuronales con Tensor Flow, la cual, tras el análisis y ajuste de parámetros, alcanzó una precisión del 94%. Los resultados obtenidos demostraron la capacidad del sistema para prever información que permite corregir parámetros antes de la producción, mejorando la eficiencia y reduciendo defectos. Además, el sistema fue validado conforme a la norma ISO 25010, logrando un alto nivel de adecuación funcional. Estos hallazgos son relevantes para futuras investigaciones en el campo de la manufactura industrial.

Palabras clave: Modelo de Predicción, CRISP DM, Inteligencia artificial, Redes Neuronales

Abstract

The present research focuses on the development of a quality prediction system for the production of modular belts in a manufacturing company in Lambayeque. The methodology used was CRISP-DM, allowing efficient structuring from understanding the problem to implementing solutions. A data set of 3469 records was used, of which 80% was used for training the model, thus guaranteeing a solid foundation for its development. To build the predictive model, artificial intelligence was used, specifically neural networks with Tensor Flow, which, after analyzing and adjusting parameters, reached an accuracy of 94%. The results obtained demonstrated the system's ability to predict information that allows parameters to be corrected before production, improving efficiency and reducing defects. In addition, the system was validated in accordance with the ISO 25010 standard, achieving a high level of functional adequacy. These findings are relevant for future research in the field of industrial manufacturing.

Keywords: Prediction model, CRISP DM, Artificial Intelligence, Neural Networks

Introducción

Durante décadas el nivel de eficiencia y capacidad de respuesta en el área de producción de las empresas manufactureras de todo el mundo estuvo limitado por las capacidades tecnológicas que requerían rigor humano y carecía de disponibilidad oportuna de datos; sin embargo, solo el 5% de ellas está satisfecha con sus enfoques hacia la digitalización [1]. Según [2] la producción es un trabajo interesante pero desafiante, una pequeña interrupción en el lugar y momento equivocado puede provocar grandes pérdidas de producción y retrasos en las entregas, además debido a la poca adopción de tecnología se estima que las PYMES indias producen el 70% de la carga de contaminación industrial, lo que se traduce en aproximadamente 32 000 millones de dólares en costos ambientales. Por lo tanto, se debe implementar tecnología para mejorar de manera sostenible la calidad del producto y el desempeño general de la empresa [3].

En la actualidad, la gran parte de las organizaciones industriales hacen uso de algún tipo de sistema informático para gestionar sus operaciones, controlar la producción, administrar el inventario, automatizar procesos, recopilar datos y analizarlos para obtener información valiosa. Según un informe de CEPAL [4] del 2022, el sector manufacturero en Brasil, México, Argentina, Costa Rica, Perú y otros países de Latinoamérica están experimentando un estancamiento en la adopción de tecnología y productividad del 12% al 14%, Esto se debe en gran medida a que las soluciones tecnológicas disponibles están diseñadas principalmente para grandes empresas, dejando a las PYMES con opciones limitadas y, a menudo, inaccesibles, resaltando la necesidad de desarrollar e implementar tecnologías adaptadas a sus capacidades y necesidades. La empresa donde se realizará la investigación enfrenta a una problemática en la predicción de la calidad de su producto debido a varios factores. En primer lugar, la empresa ha realizado una inversión inadecuada en la infraestructura tecnológica para la predicción de la calidad debido a la falta de conocimiento en el tema. Como resultado, el software utilizado para la predicción cuenta con herramientas limitadas, lo que dificulta la identificación de los factores que afectan la calidad del producto. Además, la baja calidad de la capacitación al personal encargado del análisis de información y predicción de la calidad del producto es otro problema importante. Esto ha llevado a una sobrecarga de información y a un uso excesivo de tiempo para la predicción de la calidad del producto, lo que afecta la eficiencia y la productividad de la empresa.

En los últimos meses, la fábrica enfrenta desafíos significativos debido a la sobrecarga de información generada por un aumento considerable en la cantidad de ordenes de producción para la predicción. Este incremento del 14% en comparación con los años anterior como indica el anexo 4; ha tenido un impacto negativo en la eficiencia y productividad de la empresa para gestionar la información como lo indica el anexo 5, este proceso es de suma importancia debido a que se tiene que analizar detalladamente e identificar patrones con órdenes de producción anteriores para obtener una hipótesis de la predicción precisa y confiable; al ser este un proceso manual se tiene el riesgo de obtener datos incompletos en la extracción de la información e incoherencia en los datos que dificultarían más el proceso. Por otro lado, el uso de software con herramientas limitadas para la predicción ha obstaculizado la capacidad del personal para realizar la predicción, calificando como ineficiente la herramienta actual como lo indica el anexo 10. Esto debido a que la herramienta que se utiliza es Microsoft Excel un software que no está pensando para el análisis de datos y la predicción; en la herramienta se ingresa los datos y se busca por cada variable coincidencias con otras ordenes de fabricación ya realizadas para extraer sus valores óptimos y a través de procesos estadísticos se ajustan para obtener la predicción. Estas limitaciones de la herramienta causan la demora de 5 horas aproximadamente para realizar una predicción de la calidad del producto según el anexo 6; esta demora genera retrasos en la toma de decisiones importantes para la planificación de la producción y la entrega puntual de los productos a los clientes. A pesar de que la herramienta actual brinda resultados claros y fáciles de interpretar como lo indica el anexo 7, los empleados con esta herramienta no pueden identificar y medir de manera precisa las características clave que afecten la calidad del producto, además no se puede realizar ajustes proactivos para mejorar la predicción de la calidad como lo indican el anexo 8 y 9, Esto afecta tanto la calidad del producto como la eficiencia y competitividad de la empresa en el mercado.

Los factores que influyen en la calidad de la predicción es la sobre carga de información, el uso de herramientas limitadas y la mala configuración de las máquinas por parte del operario, lo que puede afectar la precisión de los datos obtenidos. Estos factores dificultan la identificación de las causas raíz de los problemas de calidad y la implementación de soluciones efectivas. En resumen, la empresa se enfrenta a múltiples desafíos en la predicción de la calidad de sus productos, desde la falta de infraestructura tecnológica adecuada hasta la capacitación insuficiente del personal encargado del análisis de información y la predicción. Estos factores se combinan para dificultar la identificación de las causas de los problemas de calidad y la

implementación de soluciones efectivas para mejorar la calidad del producto [5].

A todo esto, se suma, que la empresa se ve presionada al incrementar sus volúmenes de producción para satisfacer la creciente demanda de sus productos, por ello, al apresurarse en sus procesos de producción y ensamblaje, cometen errores y tropiezos que generan productos defectuosos. Cuando se carece de las herramientas tecnológicas y los conocimientos adecuados, como en el presente caso de estudio, se suele descuidar la calidad cometiendo fallos que les obligan a desaprovechar inversiones, ocupar tiempos excesivos para el cálculo de material y emplear aditamentos innecesarios para la producción de las piezas de las bandas modulares, gastos excesivo de material, gasto excesivos en el pago al personal, pérdida de clientes por las demoras en la producción y las ventas, siendo esta última el motor de la empresas manufacturera [6].

Actualmente, con la industria 4.0 las expectativas de estas empresas deben ser ambiciosas. Aplicando Sistemas Inteligentes, Minería de Datos o Aprendizaje automático que ayuden a detectar, predecir y alertar problemas en el área de producción. Sin embargo, no se enfocan en las MYPES donde se podría brindar un solución rápida y económica utilizando la data histórica y las reglas generales del negocio, para predecir de manera rápida la calidad el producto, el tiempo y la configuración necesaria en las máquinas con una precisión elevada para obtener óptimos resultados. Como consecuencia conlleva una gran pérdida económica como de tiempos para las empresas y con ello es necesario volver desde cero a realizar todo su proceso, además que algunos de sus resultados negativos pueden verse después de semanas o meses dirigiendo a muchas bajas organizacionales [7]. De acuerdo con la problemática se planteó la siguiente pregunta ¿Qué variables clave debe tener un sistema de predicción de calidad del producto en el área de producción de una fábrica de bandas modulares de la región Lambayeque?

Desde el punto de vista empresarial, fue crucial implementar un sistema de predicción de calidad para bandas modulares que permitiera a la empresa detectar y prever posibles problemas antes de que se convirtieran en inconvenientes graves. La implementación de este sistema predictivo de calidad asistió a la empresa en la identificación de las causas fundamentales de los problemas de calidad, posibilitando la adopción de medidas preventivas y correctivas que mejoraron la calidad del producto y redujeron los costos de producción.

A nivel tecnológico, la propuesta estuvo dirigida a una empresa B2B, que dependía en gran medida de la calidad del producto final. Con esta solución, se buscó innovar al integrar la tecnología de predicción con minería de datos, como una propuesta novedosa para el sector industrial.

Desde el punto de vista científico, después de una revisión sistemática de antecedentes relacionados con esta investigación, no se encontraron precedentes locales de sistemas predictivos en el ámbito industrial. Por lo tanto, se concluyó que con esta investigación se aspiraba a desplegar un sistema de predicción de calidad en el proceso de producción, con el objetivo de incentivar a futuros investigadores a explorar más en este campo y proponer soluciones relacionadas con inteligencia artificial y minería de datos.

Para abordar este desafío, se propuso como objetivo general la implementación de un sistema de predicción de calidad del producto en el área de producción de una fábrica de bandas modulares en la Región Lambayeque. Para el desarrollo de la investigación, se eligió la técnica de imputación que mejor se adaptó al balanceo de datos, lo que permitió mejorar la calidad de estos. Esto facilitó la evaluación de diversos algoritmos predictivos, utilizando métricas seleccionadas para determinar cuál de ellos ofrecía los mejores resultados con la data de entrenamiento. Posteriormente, se elaboró un modelo predictivo para la predicción de la calidad de las bandas modulares. Finalmente, se validó el nivel de adecuación funcional del sistema de predicción conforme a la norma ISO/IEC 25010, garantizando que el sistema cumpliera con los criterios de calidad establecidos.

Revisión de literatura

Antecedentes

Heng et al en su investigación [8], abordar la problemática de problemas en la producción debido a procedimientos que dependen gran medida de enfoques experimentales. Por ello, se estableció como objetivo diseñar un modelo asistido por ordenador para predecir y optimizar las propiedades de olor, intensidad y la persistencia para el diseño de nuevas moléculas de fragancias para lavavajilla. En la investigación se implementaron modelos con técnicas de aprendizaje automático como redes neuronales y modelo de árbol de decisión. El modelo predictivo implementado se desarrolló con redes neuronales obteniendo como logro un modelo eficaz para el diseño de nuevas moléculas de fragancia teniendo un 95% de precisión, Así como

también se logró la clasificación y detección de moléculas potenciales para otro tipo de aplicación.

En el estudio de Kumari et al [9], se tuvo como problemática la demora en la fabricación de piezas de polímero para autos, los cuales salían de baja calidad debido al poco tiempo y los procesos experimentales de las que depende. Para solucionar ello, se planteó desarrollar un modelo predictivo con minería de datos, aprendizaje automático y metodología CRISP-DM que trabaje sobre los parámetros de calidad en la producción de componentes de polímeros; además recomiende los respectivos valores para una producción eficiente y de calidad. El modelo predictivo que se desarrolló fue un algoritmo de aprendizaje automático de regresión logística, además se tomaron parámetros tales como: el diseño del molde, el espesor, la temperatura, tiempo de ciclo, material, etc; siendo el factor de decisión de calidad el ciclo de moldeo. Finalmente, se obtuvo una precisión de 95% para predecir la calidad de los componentes de las piezas de polímero antes de que se fabriquen reduciendo así los defectos de fabricación, aumentando la eficiencia de la producción.

C. Urrea y M. Ramos en su investigación [10], se tuvo como problemática que, en la fabricación de tableros, la calidad del producto es función de múltiples variables, especialmente de la variabilidad de la madera. Esta calidad depende, entre otros factores, de la adherencia entre chapas o resistencia a la tracción perpendicular. Como solución se planteó como objetivo desarrollar un modelo de aprendizaje automático que permita predecir la adherencia entre chapas en la fabricación de tableros. Se identificaron los parámetros clave que se consideraron en el estudio, los cuales incluyeron: tiempo de operación, cantidad de adhesivo utilizado, condiciones ambientales y temperatura de la chapa; Además se usó la metodología KDD y se evaluaron algoritmos de Redes Neuronales, árbol de decisión y Random Forest. obteniendo como resultado del modelo de redes neuronales una precisión de 69.95 % para predecir la adherencia entre chapas en la fabricación de tableros.

Gonzales en su investigación [11] describió la problemática de que la empresa farmacéutica objeto de estudio, ubicada en Chiclayo, ha enfrentado desafíos en la planificación de sus compras debido a su corta trayectoria y la falta de procedimientos claros para determinar qué medicamentos adquirir, en qué cantidad y cuándo hacerlo. El objetivo principal de la investigación fue desarrollar un sistema de soporte para la toma de decisiones que mejore la

planificación de compras en la droguería. Específicamente, el sistema busca optimizar la rotación de productos mediante un modelo de predicción basado en redes neuronales, logrando una alta precisión en la predicción de ventas y validando su usabilidad con los usuarios finales. El estudio utilizó una combinación de la metodología CRISP-DM para la construcción del modelo predictivo y SCRUM para el desarrollo de la aplicación web que interactúa con dicho modelo. La implementación técnica se realizó con Python y Laravel, y el sistema permitió gestionar clientes, proveedores, productos, y realizar operaciones CRUD. Para el desarrollo del modelo de predicción, se trabajó con datos de ventas y compras proporcionados por la empresa en formato Excel, los cuales fueron procesados y analizados para seleccionar los campos clave. Tras comparar diferentes técnicas, se optó por utilizar redes neuronales para la predicción de ventas. El modelo predictivo basado en redes neuronales alcanzó un 90.23% de precisión al eliminar las ventas con valores cercanos o iguales a cero. Se desarrollaron un total de 727 simulaciones, de las cuales el 90.23% cumplió con los criterios de aceptación establecidos. El sistema implementado permitió una mejora significativa en la planificación de compras, reduciendo las pérdidas económicas y mejorando la disponibilidad de productos en stock. Además, las pruebas de caja blanca y negra confirmaron la eficiencia de la interfaz, y las encuestas de los usuarios finales indicaron una alta probabilidad de adopción del sistema.

Delgado en su investigación [12] relato la problemática del Banco Azteca en Chiclayo que se enfrentó a serios problemas en la identificación de clientes morosos, lo que ha generado pérdidas significativas debido a la inclusión inexacta de clientes en listas de alto riesgo. El objetivo principal de esta investigación fue implementar una aplicación móvil utilizando un algoritmo de redes neuronales para predecir el comportamiento de los clientes del Banco Azteca en base a su historial crediticio, identificando a los clientes de alto riesgo y validando el sistema de detección según la norma ISO 25010. El sistema permitirá a los empleados monitorear a los clientes, simular su riesgo de morosidad y facilitar el contacto con ellos mediante la integración de un sistema de mensajería. Se desarrolló una aplicación móvil utilizando redes neuronales artificiales para analizar el historial crediticio de los clientes y predecir su comportamiento futuro. El sistema fue validado bajo los estándares de calidad ISO 25010 y se implementó con la capacidad de simular escenarios de morosidad y enviar notificaciones a los empleados del banco. Además, se diseñó una arquitectura tecnológica adecuada para soportar el procesamiento y almacenamiento de la información crediticia. En este estudio se implementó un modelo de redes neuronales para predecir el comportamiento crediticio de los clientes del Banco Azteca.

Se utilizó un algoritmo de aprendizaje supervisado, específicamente el modelo de redes neuronales multicapa, el cual fue elegido por su capacidad para identificar patrones en los datos históricos de los clientes y prever su riesgo crediticio con alta precisión. Entre otros algoritmos revisados, también se evaluaron Árboles de Decisión y modelos no supervisados como Apriori, aunque el modelo final seleccionado fue el de redes neuronales debido a su mayor precisión en este contexto.

Lopez en su investigación [13] describe la problemática del cultivo de café en la región Cajamarca enfrenta desafíos significativos debido a la escasa capacidad técnica de los agricultores para identificar deficiencias nutricionales en las plantas, lo que repercute en la calidad del producto final y afecta la producción. Los agricultores suelen depender de métodos empíricos y visuales para detectar deficiencias, lo que puede llevar a errores y diagnósticos inexactos. El objetivo principal de esta investigación fue desarrollar un sistema inteligente basado en redes neuronales convolucionales para apoyar la identificación de deficiencias nutricionales en las hojas de café mediante el reconocimiento de imágenes, implementado en una aplicación móvil híbrida. Se implementó un modelo de red neuronal convolucional entrenado con imágenes de hojas de café clasificadas según cinco tipos de deficiencias nutricionales (nitrógeno, fósforo, potasio, magnesio y hojas sin deficiencia). El modelo se desarrolló utilizando las librerías Keras y TensorFlow en Python, mientras que la aplicación móvil híbrida se creó con HTML, CSS, y JavaScript, empaquetada con Framework 7 y Phonegap. La metodología CommonKADS fue utilizada para la gestión del conocimiento y RUP para las iteraciones funcionales del sistema. El sistema inteligente desarrollado utilizó un modelo de red neuronal convolucional, entrenado con imágenes de hojas de café clasificadas por deficiencias nutricionales. Este modelo fue implementado con las librerías Keras y TensorFlow. El proceso de entrenamiento incluyó la normalización y procesamiento de las imágenes, lo que permitió una correcta clasificación de cinco clases de deficiencias: nitrógeno, fósforo, potasio, magnesio y hojas sin deficiencia. El modelo alcanzó una precisión del 95% en la identificación de deficiencias nutricionales durante las pruebas de validación. Los expertos en agronomía que evaluaron el sistema confirmaron la eficacia del modelo, destacando la utilidad de las recomendaciones sobre los fertilizantes adecuados proporcionadas por la aplicación.

Bases teóricas científicas

En el desarrollo de esta investigación es fundamental comprender los siguientes términos:

Inteligencia artificial (IA): Se fundamenta en la reproducción de la capacidad cognitiva humana a través del diseño y la implementación de algoritmos en un contexto computacional en constante cambio. Para lograr este objetivo, son necesarios tres elementos esenciales: sistemas computacionales, datos y su gestión, y algoritmos avanzados de IA, la cantidad de datos y capacidad de procesamiento requerida aumenta a medida que se busca una mayor similitud en el comportamiento humano [14].

Minería de Datos: Proceso para descubrir patrones y relaciones útiles y significativas en extensos conjuntos de datos utilizando técnicas estadísticas y computacionales avanzadas. Se utiliza para extraer información valiosa de datos complejos y estructurados, para apoyar a las empresas y organizaciones a tomar decisiones fundamentales y respaldarlas con información precisa y basadas en datos. La minería de datos utiliza una variedad de técnicas, como la clasificación, la regresión, la asociación y el clustering, para analizar grandes cantidades de datos y encontrar patrones que puedan ayudar a entender mejor el comportamiento de los clientes, predecir tendencias futuras, y mejorar la toma de decisiones empresariales [15].

- **Tipos de Datos**

- **Magnitudes:** se refieren a una forma de información que se puede medir o cuantificar numéricamente, es decir, son datos cuantitativos. Estos datos pueden ser continuos o discretos y se caracterizan por tener una escala o unidad de medida [15].
- **Catagóricos:** Los datos catagóricos son un tipo de información que se identifica y almacena por medio de sus etiquetas o nombres. Son datos cualitativos que se organizan en categorías en lugar de ser medidos numéricamente. Este tipo de datos está compuesto por variables catagóricas, como el género o la ciudad de origen de una persona, que se expresan en lenguaje natural en lugar de números [15].

- **Métodos**

- **Métodos supervisados:** También conocidos como aprendizaje supervisado, implican la construcción de modelos que utilizan conjuntos de datos etiquetados para predecir o clasificar nuevos datos. El algoritmo se instruye con un conjunto de datos de entrenamiento que incluye ejemplos etiquetados que indican la respuesta deseada. A partir de esta información, el algoritmo construye un modelo que puede utilizarse para hacer predicciones precisas sobre nuevos datos no etiquetados [16].
- **Métodos no supervisados:** Se utilizan para descubrir patrones o estructuras en un

conjunto de datos sin etiquetar y agruparlos en función de características comunes, los métodos no supervisados incluyen técnicas como el clustering o agrupamiento, la reducción de dimensionalidad y la detección de anomalías [16].

- **Modelos predictivos**

- **Modelos basados en redes neuronales:** Esta técnica de aprendizaje automático se aplica en la minería de datos con el fin de analizar y clasificar amplios y complejos conjuntos de datos. Los modelos de redes neuronales están inspirados en la estructura y funcionamiento del cerebro humano, operando mediante una red de nodos interconectados que procesan información. [17].
- **Modelo basado en árbol de decisión:** Este tipo de modelo de aprendizaje automático utiliza una estructura similar a un árbol para facilitar la toma de decisiones. Su funcionamiento se basa en reglas de decisión organizadas en una estructura jerárquica de nodos. En esta configuración, cada nodo representa un atributo o característica del conjunto de datos de entrada, mientras que cada rama simboliza una posible respuesta o salida [17].
- **Modelo basado en regresión logística:** Es un método estadístico que se utiliza para analizar y predecir la relación entre una variable dependiente binaria y una o más variables objetivo o independientes continuas o categóricas. En otras palabras, es un modelo que se emplea para predecir la probabilidad de que ocurra un evento o resultado binario en función de otras variables [17].
- **Técnicas de imputación:** Existen varias técnicas de imputación de datos que se pueden utilizar para manejar valores faltantes en conjuntos de datos. Aquí hay algunas técnicas comunes de imputación que se pueden implementar en Python [17]:
 - Imputación por valor medio: Esta técnica consiste en reemplazar los valores faltantes con la media de los valores existentes en esa columna. Se puede implementar utilizando la función `fillna()`.
 - Imputación por valor mediano: En esta técnica, los valores faltantes se reemplazan con la mediana de los valores existentes en esa columna. También se puede realizar utilizando la función `fillna()` de pandas.
 - Imputación por valor más frecuente: En esta técnica, los valores faltantes se reemplazan con el valor más común (moda) en esa columna. Esto se puede hacer utilizando la función `fillna()` de pandas y la función `mode()` para calcular la moda.

- Imputación mediante regresión: Esta técnica utiliza un modelo de regresión para predecir los valores faltantes en función de las variables predictoras disponibles. Puedes utilizar bibliotecas como scikit-learn para implementar modelos de regresión en Python y, luego, utilizar esos modelos para predecir y reemplazar los valores faltantes.

Materiales y métodos

A continuación, se presenta el tipo, métodos y metodología de investigación utilizados en este proyecto:

Tipo de investigación

Para este proyecto el tipo de investigación es aplicada, dado que tiene como objetivo desarrollar una solución práctica.

Según el Manual de Frascati [18], una investigación aplicada se define como aquella que tiene como objetivo la aplicación práctica de conocimientos o la resolución de problemas concretos.

Métodos de investigación

Se detalla a continuación, métodos y técnicas de investigación que se utilizaran:

TABLA I
MÉTODOS DE INVESTIGACIÓN

Método	Descripción
Analítico	Este método se utiliza para analizar en profundidad las causas que afectan la calidad de los módulos plásticos, identificando variables críticas utilizadas.
Deductivo	Se emplea para desarrollar una estrategia que guíe la construcción del sistema de predicción de calidad.
Implementación	Este método abarca la ejecución práctica de la solución desarrollada, que incluye la construcción y entrenamiento del modelo de predicción.
Prueba/Testeo	Se lleva a cabo para evaluar la efectividad y precisión del sistema de predicción implementado.

TABLA II
TÉCNICAS DE INVESTIGACIÓN

Técnicas	Instrumentos
Entrevista	Guía de entrevista
Encuestas	Cuestionarios a personal de la empresa
Observación	Ficha de observación
Análisis de documentos	Informes técnicos de producción

Metodología de desarrollo

La metodología escogida para este proyecto de investigación fue CRISP-DM [19] la cual es una metodología de desarrollo de minería de datos ampliamente utilizada, esta proporciona un enfoque estructurado y sistemático para la elaboración de proyectos de minería de datos, desde la fase de comprensión del problema hasta la fase de implementación de soluciones.

La metodología se basa en un marco de trabajo compuesto por seis fases:

- **Comprensión del problema:** En esta fase se reconoció la necesidad de un sistema enfocado a la predicción de la calidad de los módulos de plástico de las bandas modulares. Es por ello, que se tomo como objetivo principal el uso de minería de datos para predecir la calidad del producto en función de variables clave en el proceso de producción. Esto facilito una planificación de producción más precisa, una toma de decisiones informada y una reducción en los costos.
- **Recopilación de datos:** Esta fase se enfocó en reunir información detallada sobre los parámetros clave que afectan la calidad de los módulos plásticos de la banda modular, tales como las temperaturas de moldeo, el ciclo de producción, los materiales, entre otros. Esto resultó en un data set de 3,469 registros en total, el cual posteriormente fue procesado cuidadosamente para asegurar su consistencia de los datos.
- **Preparación de datos:** Se realizó un proceso exhaustivo para asegurar la calidad y consistencia de la data set de 3,469 registros. Primero, se limpiaron y estandarizaron los datos, eliminando valores atípicos y completando valores faltantes mediante imputación por media, dado que el tamaño de los datos a imputar era bajo y no afectaba significativamente. Posteriormente, se normalizaron las variables categóricas con el LabelEncoder para que pueda ser entendido por el modelo predictivo y facilitar su interpretación.

- **Modelado:** Se construyeron modelos de machine learning diferentes basados en los antecedentes de esta investigación y en las características que estos tenían para verificar su compatibilidad con el contexto de esta investigación. La implementación se realizó con el apoyo de las librerías Scikit-Learn y Tensor Flow, además se definió las métricas con las que fueron evaluadas tales como: error cuadrático medio (MSE), raíz de error cuadrático medio (RMSE), error absoluto (MAE), coeficiente de determinación (R2).
- **Evaluación:** Se evaluaron los modelos implementados anteriormente, donde las redes neuronales resultaron ser la opción más adecuada, debido a su capacidad para capturar la complejidad de los datos y adaptarse a las variables. A través de pruebas y ajustes, se logró alcanzar un R2 final del 94% asegurando la confiabilidad del modelo para este contexto.
- **Implementación:** Se integro el modelo con el sistema web de predicción de calidad de los módulos de plástico de la banda modular. El sistema web se diseñó con interfaces intuitivas que presenta los resultados de manera clara, permitiendo a los usuarios simular, predecir y analizar las variables para tomar decisiones informadas.

Resultados y discusión

En base a los objetivos el proyecto

La metodología de desarrollo para este proyecto es CRISP-DM, a continuación, se detallará el cumplimiento de los objetivos.

Objetivo específico 01: Elección de técnica de imputación para el balanceo de datos.

Para el cumplimiento de este objetivo se identificaron y analizaron tres técnicas siendo estas: imputación por media, imputación por valor y K-vecinos, entre las tres técnicas se eligió a imputación por media para el balanceo de datos, los indicadores que intervinieron para tal elección fueron: a) dispersión de los datos y b) tamaño del conjunto de datos imputados.

a) Dispersión de los datos:

· Imputación por media en algunas variables

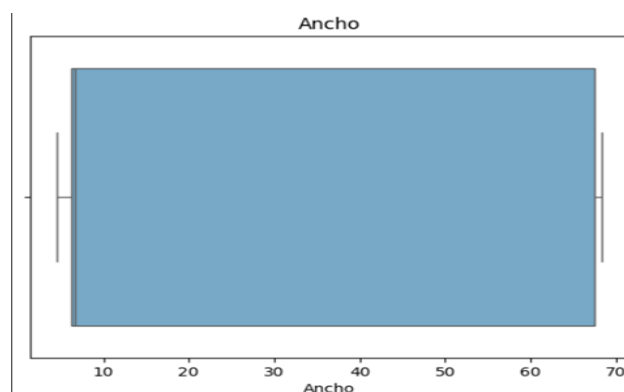


Ilustración 1 "Imputación por media en la variable ancho"

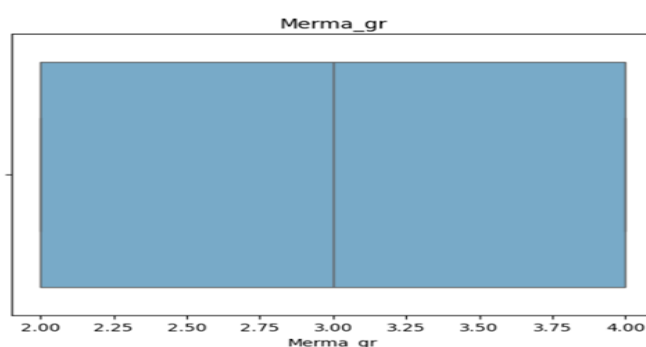


Ilustración 2 "Imputación por media de la variable merma_gr"

b) Tamaño del conjunto de los datos:

TABLA III
TABLA DE ELECCIÓN DE TÉCNICAS DE IMPUTACIÓN

Técnicas	Dispersión de los datos	Tamaño del conjunto de datos imputados
Media	Baja	55
Imputación por valor	Media	55
K- vecinos	Alta	55

La técnica de imputación que obtuvo mejores resultados fue la imputación por media, debido a que el tamaño de los datos a imputar es bajo y, al ser una técnica comúnmente utilizada en estos casos, garantiza la integridad del conjunto de datos. Por lo tanto, fue la elegida para el desarrollo de este proyecto.

Objetivo específico 02: Elección del algoritmo predictivo basado en minería de datos.

Para la realización de este objetivo se realizó una comparación de algunos modelos de machine learning de la librería Scikit-learn y Tensor Flow. Además, los modelos que se presentan a continuación se tuvieron en cuenta en base a los antecedentes de previos con realidades similares al de este proyecto, son los siguientes:

TABLA IV
CARACTERISTICAS COMPARATIVAS DE LOS ALGORITMOS EMPLEADOS

Modelo	Dificultad de implementación	Cantidad de registros	Tiempo de entrenamiento	Tipo de aprendizaje
Regresión Lineal (Scikit-learn)	Fácil	Media	Rápido	Supervisado
Regresión Lineal (Tensor Flow)	Media	Media	Rápido	Supervisado
Scikit-learn MLPRegressor	Media	Alto	Dependiente del tamaño	Supervisado
Tensor Flow Neural Network	Difícil	Alto	Dependiente del tamaño	Supervisado
SVM	Difícil	Alto	Dependiente del tamaño	Supervisado

Al poner a prueba cada uno de los modelos se obtuvieron los siguientes resultados

	Modelo	MSE	RMSE	R ²	MAE
0	Regresión Lineal (Scikit-learn)	51.018524	7.142725	0.357687	4.474680
1	Regresión Lineal (TensorFlow)	50.999604	7.141401	0.357925	4.481367
2	Scikit-learn MLPRegressor	30.632589	5.534672	0.614342	3.634109
3	TensorFlow Neural Network	13.688016	3.699732	0.827670	2.155203
4	Support Vector Machine	73.008858	8.544522	0.080832	5.062696

Ilustración 3 "Resultados de los modelos comparados"

Donde se puede apreciar que el modelo de inteligencia artificial que mejor se adapta a la realidad de esta investigación es el modelo de Tensor Flow Neural Network con un porcentaje R2 de 83. Por consiguiente, el modelo de Tensor Flow Neural Network fue seleccionado como modelo de inteligencia artificial para la realización de este proyecto.

Objetivo específico 03: Elaborar en base al algoritmo seleccionado un modelo predictivo para la calidad del producto.

Para este objetivo se procedió a la construcción del modelo en este caso el modelo de Tensor Flow Neural Network con algunos ajustes en sus parámetros obteniendo un porcentaje de R2 de 94, quedando finalmente como se puede apreciar en las siguientes ilustraciones:

```
# Define el modelo de red neuronal con TensorFlow
model = Sequential([
    Dense(100, activation='relu', input_shape=(X_train.shape[1],)),
    Dense(50, activation='relu'),
    Dense(1)
])
model.compile(optimizer='adam', loss='mean_squared_error')

# Entrena el modelo
model.fit(X_train_scaled, y_train, epochs=500, batch_size=42, verbose=0)
```

Ilustración 4 "Modelo de redes neuronales de Tensor Flow"

Este modelo está construido de la siguiente forma:

- Redes neuronales: El modelo se construye secuencialmente (Sequential), lo que significa que las capas se apilan una encima de la otra en secuencia. Este modelo tiene tres capas densas (Dense) que están completamente conectadas entre sí.
- Primera capa densa: Tiene 100 neuronas y utiliza la función de activación (relu), que introduce no linealidad en el modelo, lo que le permite aprender patrones más complejos en los datos de entrada.
- Segunda capa densa: Tiene 50 neuronas y también utiliza la función de activación (relu). Esta capa proporciona otra capa de abstracción sobre los datos de entrada, ayudando al modelo a aprender representaciones más profundas y abstractas de los datos.
- Capa de salida: Tiene una sola neurona sin función de activación específica. Esta capa produce la salida final del modelo, que en este caso es un valor numérico continuo.

	Modelo	MSE	RMSE	R ²	MAE
0	Regresión Lineal (Scikit-learn)	51.026855	7.143308	0.357582	4.474732
1	Regresión Lineal (TensorFlow)	51.612538	7.184187	0.350208	4.514937
2	Scikit-learn MLPRegressor	31.693813	5.629726	0.600981	3.712509
3	TensorFlow Neural Network	4.664943	2.159848	0.941269	1.339327
4	Support Vector Machine	73.008858	8.544522	0.080832	5.062696

Ilustración 5 "Resultado final del modelo seleccionado"

Además de estos indicadores se hicieron pruebas y comparativas entre los valores reales y los valores de predicción donde se pudo apreciar que el grado de error no es tan alto, como se puede observar en la siguiente Ilustración.

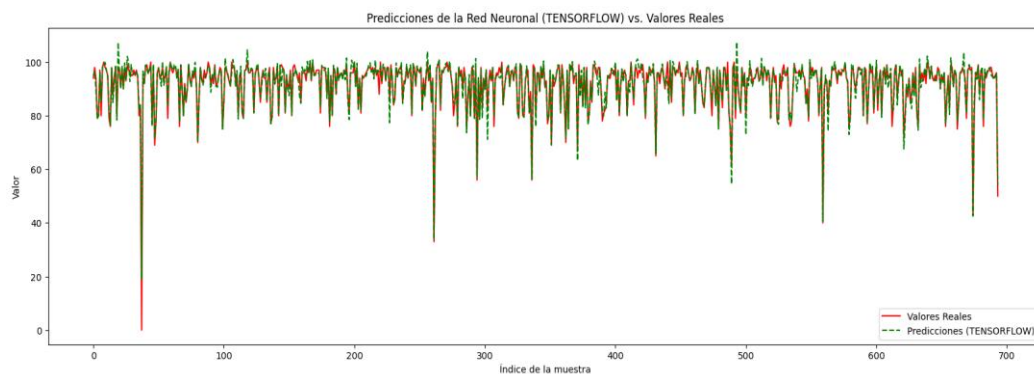


Ilustración 6 "Comparación de predicción y valores reales"

Por consiguiente, se puede concluir, que el modelo implementado es lo suficientemente confiable para la predicción de la calidad de la banda modular.

Objetivo específico 04: Validar el nivel de adecuación funcional del sistema de predicción con la norma ISO/IEC 25010

Se aplicó la ISO 25010 para validar la educación funcional del sistema se revisó el sistema con el administrador e ingeniero industrial la empresa, asegurando que los procesos y su funcionalidad cumplan con los requisitos establecidos por el empleado de calidad. Para ello se tomó el trabajo de [21], donde se detalla el proceso de evaluación de las sub características de la adecuación funcional que se puede observar en la tabla V. Como resultado se obtuvo un nivel de adecuación funcional de 91,67% clasificado como Alto.

TABLA V
MÉTRICAS DE EVALUACIÓN DE LAS SUB-CARACTERÍSTICAS PARA DETERMINAR EL NIVEL DE ADECUACIÓN FUNCIONAL

Sub-característica	Descripción	Fórmula	Parámetros
Complejidad funcional	Grado en que el conjunto de funcionalidades del producto cubre todas las tareas y objetivos de usuario especificados [21].	$x1 = 1 - ff/te$	<ul style="list-style-type: none"> - x1: Complejidad funcional - ff: Número de funciones que faltan o están incorrectamente implementadas - te: Número de funciones en la especificación de requisitos.
Corrección funcional	Capacidad del producto software para proporcionar funciones que faciliten la consecución de tareas y	$x2 = e/te$	<ul style="list-style-type: none"> - x2: Corrección funcional - e: Número de elementos de datos implementados - te: Número total de elementos de datos implementados.

objetivos específicos
[21].

Pertinencia funcional	El programa no realiza funciones innecesarias o excesivas [21].	$x_3 = fu/TFU$	<ul style="list-style-type: none"> - x3: Pertinencia funcional - fu: Número de funciones realmente útiles - TFU: Número de funciones implementadas para tareas específicas
-----------------------	---	----------------	--

El valor deseado debe estar entre el rango $0 \leq x_i \leq 1$, el valor más cercano al 1 es mejor. Además, la ponderación de la calidad es la equivalencia $1 \Rightarrow 100\%$ Rangos de calidad: bajo $< 75\%$, medio $\geq 75\%$ y Alto $\geq 85\%$ [21].

TABLA VI
ISO 25010 – ADECUACION FUNCIONAL

Sub característica	Propósito	Problema	Solución	Detalle	Cumplió
Compleitud Funcional	Permitir el ingreso de forma segura al sistema.	El sistema necesita evitar accesos no autorizados y proteger la información confidencial de los usuarios.	El sistema se encargará de validar los datos de acceso ingresados, Así mismo, solo mostrara las opciones a los cuales tienen permiso de ingresar.	Se ingresa al inicio de sesión del sistema que solicitara (Usuario y contraseña) verificando que solo las credenciales validas permitan ingresar y mostrar las opciones definidas para el usuario.	SI
	Permitir la navegación entre páginas del sistema y actualizar los ítems visibles según el tipo de usuario	Los usuarios requieren acceder a diferentes secciones del sistema y visualizar solo las opciones relevantes a sus roles.	Implementar una barra de navegación dinámica que permita cambiar entre páginas del sistema y actualice los elementos visibles en función del rol o permisos del usuario.	Al ingresar al sistema la barra de navegación asegura que cada usuario pueda acceder únicamente a las páginas permitidas según su rol (Administrador y Calidad) .	SI
	Permitir al usuario ingresar, modificar y filtrar datos según las vistas del sistema	Los usuarios necesitan gestionar datos de forma eficiente, incluyendo la capacidad de añadir, actualizar y filtrar información específica.	Implementar vistas dinámicas que proporcione formularios para el ingreso y modificación de datos, y herramientas de filtrado avanzadas para buscar información según criterios específicos.	Cada vista tiene sus funciones respectivas para de ingresar, modificar y filtrado de datos, asegurando que los datos se guarden, actualicen y filtren correctamente según las acciones del usuario.	SI
	Predecir la calidad de la banda modular	No se sabe con precisión si una banda modular cumple con los estándares de calidad	El sistema utilizará datos clave de la producción, para predecir la calidad de la banda modular, y asegurar que estén dentro de los rangos aceptables antes de que entren a producción.	Utilizando datos históricos de producción y aplicando el modelo de redes neuronales para predecir la calidad esperada de la banda modular, en el módulo de predicción del sistema	SI

	necesarios para el proceso de producción.		se verifico que las predicciones sean consistentes y estén alineadas con los estándares de calidad esperados.	
Recomendar ajustes en los parámetros para mejorar la predicción de calidad.	No tener una estimación precisa de que parámetros de producción se deben ajustar para mejorar la calidad de la banda modular.	El sistema mostrara ciertos parámetros clave de producción como (C_Molde_Macho, C_Molde_Hembra, C_Product, Zona_1, Zona_2, Zona_3, Zona_4 y Zona_5) para simular y evaluar el impacto en la calidad de la banda modular.	Se realiza de manera simulada iteraciones que ajustan los parámetros de producción mencionados sacando su promedio min y máximo y observando el impacto en la calidad de la banda modular, verificando que las recomendaciones proporcionen mejoras en la predicción de calidad.	SI
Sistema de almacenamiento de datos	de Almacenamiento de archivos de en documentos Excel, lo que provoca duplicidad, gestión ineficiente de datos (datos incompletos o inconsistentes) y demora en la ubicación de datos de fabricación.	El sistema de almacenamiento de datos utilizando es MySQL, eliminando la duplicación de archivos de datos y aumentando la velocidad de consulta.	Se creo una base de datos en MySQL y se usó Django como ORM para la gestión de este y creación de APIs, verificando la consistencia y accesibilidad de los datos, así como la rapidez en las consultas de información.	SI
Historial de predicción	de No se tiene un control sobre las predicciones anteriores, lo que dificulta el seguimiento de la evolución de la calidad.	Registrar las predicciones de calidad históricas permitiendo un mejor control y análisis de las tendencias en la calidad a lo largo del tiempo.	Crear gráficos que muestren las predicciones de calidad de cada banda modular. Verificando que se reflejen correctamente las tendencias y cambios en las predicciones.	SI

Resultados:

- Número de funciones que faltan o están incorrectamente implementadas: 0.
- Número de funciones establecidas en la especificación de requisitos: 8.
- Completitud funcional: $x1 = 1 - ff/ff \rightarrow x1 = 1 - 0/8 \rightarrow x1 = 1$

TABLA VII
ISO 25010 – CORRECCION FUNCIONAL

Sub característica	Propósito	Problema	Solución	Detalle	Cumple
Corrección Funcional	Predicción de la calidad de la banda modular.	¿Funciona correctamente la actividad sistema?	Al ingresar al módulo de predicción se debe digitar los datos de entrada para calcular el porcentaje de calidad de la banda modular.	Se tomado los datos de entrada para la fabricación de la banda modular y se utiliza para dar como resultado el porcentaje de calidad, basándonos en el historial con el que se entrenó la red neuronal. Además, los parámetros ingresados tienen que estar dentro de los límites establecidos.	SI
	Recomendación de ajustes en los parámetros para mejorar la predicción de calidad.	¿Funciona correctamente la actividad?	Al obtener el porcentaje de calidad esperado para alguna configuración de producción, se podrá tener la opción de recomendación que tomará los datos ingresados y procede iterar y evaluar que los parámetros.	Se toman los datos ingresados y se proceda a iterar y verificar que el sistema recomiende ajustes adecuados y que estos ajustes lleven a mejoras en la calidad de la banda modular. Evaluando que los resultados y asegurar que se mantengan dentro de los límites establecidos.	SI

Resultados

- Número de elementos de datos no implementados con el estándar específico de exactitud: 0.
- Número total de elementos de datos implementado: 2.
- Corrección funcional: $x2 = e/te \rightarrow x2 = 0/2 \rightarrow x2 = 1$

TABLA VIII
ISO 25010 – PERTINENCIA FUNCIONAL

Sub característica	Propósito	Problema	Solución	Detalle	Cumplimiento
Pertinencia Funcional	Validación de Actividades del Usuario	¿El sistema permite llevar a cabo actividades fuera de lo solicitado?	El usuario solo podrá acceder a actividades a las cuales tiene permiso.	Cada usuario tiene funcionalidades a las cuales solo ellos podrán visualizar. Además, son únicas con sus nombres y apellidos.	SI
	Consistencia en la Interfaz	Inconsistencias en la interfaz que pueden confundir al usuario.	Diseñar una interfaz de usuario coherente y consistente en todas las funcionalidades del sistema.	La interfaz sigue un diseño uniforme, con elementos y etiquetas consistentes a lo largo de todo el sistema. Además de estar separado del Backend se puede modificar sin temor a modificar del mismo.	SI
	Interacción con APIs	Las APIs utilizadas no proporcionan los datos correctos o no se integran adecuadamente.	Implementar validaciones y pruebas para garantizar que las APIs integradas proporcionen datos precisos, necesarios y que la integración sea correcta.	Verificar que las APIs estén correctamente integradas y que los datos recibidos sean correctos y los datos de respuesta sean solos los necesarios. Además, los únicos que pueden consumir las APIs son usuarios con Token de autenticación del sistema.	SI
	Documentación de APIs	La documentación de las APIs no es clara o está incompleta, lo que dificulta su uso.	Crear y mantener documentación clara y completa para todas las APIs utilizadas.	Revisar que la documentación de las APIs Web esté actualizada y sea fácil de entender, con ejemplos y descripciones claras.	SI

Resultados

- Número de funciones realmente útiles para realizar tareas específicas: 3
- Número de funciones implementadas para la consecución de tareas específicas: 4.
- Completitud funcional: $x_3 = f_u/t_f \rightarrow x_3 = 3/4 \rightarrow x_3 = 0.75$

Evaluación de la adecuación funcional:

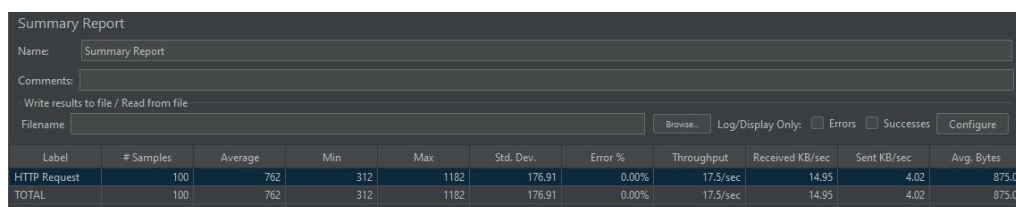
- Completitud funcional: 1.
- Corrección funcional: 1.
- Pertinencia funcional: 0.75.
- Adecuación funcional: $x = (1+1+0.75) / 3 \rightarrow x = 0.9167$
- Porcentaje de adecuación funcional: 91.67 %.
- Nivel de adecuación funcional: Alto

Se evaluó la adecuación funcional del sistema utilizando los criterios de completitud, corrección y pertinencia funcionales, obteniendo puntajes de 1, 1 y 0.75 respectivamente. Al promediar estos valores, se calculó un puntaje de 0.9167, lo que corresponde a un 91.67% de adecuación funcional. Este porcentaje clasifica el sistema en un nivel de adecuación funcional "Alto", demostrando que el sistema cumple de manera efectiva con las funciones requeridas y se ajusta adecuadamente a las necesidades operativas específicas.

PRUEBAS DE SOFTWARE

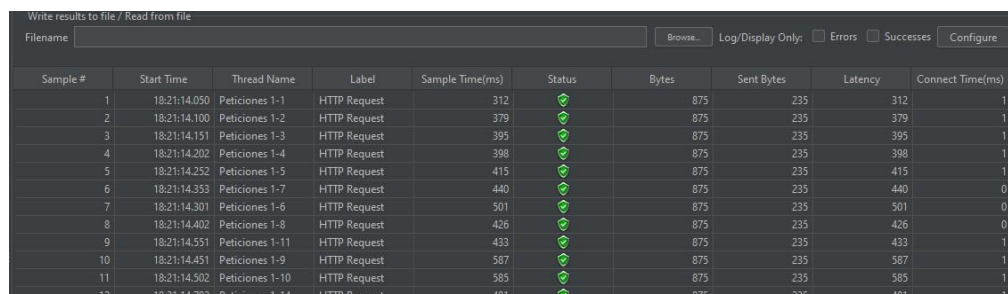
TABLA IX
PRUEBA DE CARGA – PETICIONES DE INGRESO AL SISTEMA

Acciones realizadas	Herramienta	Resultado
Enviar 100 peticiones de ingreso al sistema de manera simultánea, para saber cómo se comporta el sistema.	JMeter (5.6.3)	En la Ilustración 7 se muestra que el sistema respondió de manera óptima durante la prueba, con un error del 0.00% y una tasa de rendimiento de 17.5 peticiones por segundo. Además, la Ilustración 8 confirma que todas las solicitudes fueron atendidas satisfactoriamente. Sin embargo, como se observa en la Ilustración 9, el servidor experimentó cierta inestabilidad en sus primeras etapas de operación.



Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
HTTP Request	100	762	312	1182	176.91	0.00%	17.5/sec	14.95	4.02	875.0
TOTAL	100	762	312	1182	176.91	0.00%	17.5/sec	14.95	4.02	875.0

Ilustración 7 "Reporte de resumen de la prueba de carga"



Sample #	Start Time	Thread Name	Label	Sample Time(ms)	Status	Bytes	Sent Bytes	Latency	Connect Time(ms)
1	18:21:14.050	Peticiones 1-1	HTTP Request	312	✓	875	235	312	1
2	18:21:14.100	Peticiones 1-2	HTTP Request	379	✓	875	235	379	1
3	18:21:14.151	Peticiones 1-3	HTTP Request	395	✓	875	235	395	1
4	18:21:14.202	Peticiones 1-4	HTTP Request	398	✓	875	235	398	1
5	18:21:14.252	Peticiones 1-5	HTTP Request	415	✓	875	235	415	1
6	18:21:14.353	Peticiones 1-7	HTTP Request	440	✓	875	235	440	0
7	18:21:14.301	Peticiones 1-6	HTTP Request	501	✓	875	235	501	0
8	18:21:14.402	Peticiones 1-8	HTTP Request	426	✓	875	235	426	0
9	18:21:14.551	Peticiones 1-11	HTTP Request	433	✓	875	235	433	1
10	18:21:14.451	Peticiones 1-9	HTTP Request	587	✓	875	235	587	1
11	18:21:14.502	Peticiones 1-10	HTTP Request	585	✓	875	235	585	1
12	18:21:14.702	Peticiones 1-14	HTTP Request	481	✓	875	235	481	1

Ilustración 8 "Resultado de árbol de la prueba de carga"



Ilustración 9 "Grafico de la prueba de carga"

TABLA X
PRUEBA DE ESTRES

Acciones realizadas	Herramienta	Resultado
El sistema fue sometido a una prueba de límite al manejar 500 usuarios concurrentes, una carga significativamente mayor de lo esperado, para evaluar su comportamiento bajo estas condiciones extremas.	JMeter (5.6.3)	Como se muestra en la Ilustración 10, el sistema no pudo procesar una gran cantidad de solicitudes de acceso, con un porcentaje de error del 52.80%. Esta situación se evidencia en las Ilustración 11 y 12, donde se observa que muchas solicitudes terminaron en timeout debido a que el servidor no pudo responder a tiempo.

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
HTTP Request	500	402	0	1249	437.62	52.80%	57.8/sec	100.09	6.26	1774.2
TOTAL	500	402	0	1249	437.62	52.80%	57.8/sec	100.09	6.26	1774.2

Ilustración 10 "Reporte de resumen de la prueba de stress"

Sample #	Start Time	Thread Name	Label	Sample Time(ms)	Status	Bytes	Sent Bytes	Latency	Connect Time(ms)
1	19:09:36.892	Peticiones 1-1	HTTP Request	297	Success	875	235	297	1
2	19:09:37.307	Peticiones 1-27	HTTP Request	0	Failure	2578	0	0	0
3	19:09:36.909	Peticiones 1-2	HTTP Request	401	Success	875	235	401	0
4	19:09:37.323	Peticiones 1-28	HTTP Request	1	Failure	2578	0	0	1
5	19:09:36.973	Peticiones 1-6	HTTP Request	372	Success	875	235	372	0
6	19:09:36.959	Peticiones 1-5	HTTP Request	403	Success	875	235	403	0
7	19:09:37.371	Peticiones 1-31	HTTP Request	1	Failure	2578	0	0	1
8	19:09:37.387	Peticiones 1-32	HTTP Request	1	Failure	2578	0	0	1
9	19:09:37.435	Peticiones 1-35	HTTP Request	1	Failure	2578	0	0	1
10	19:09:37.051	Peticiones 1-11	HTTP Request	396	Success	875	235	396	0
11	19:09:37.067	Peticiones 1-12	HTTP Request	380	Success	875	235	380	1
12	19:09:37.452	Peticiones 1-36	HTTP Request	0	Failure	2578	0	0	0
13	19:09:37.466	Peticiones 1-37	HTTP Request	1	Failure	2578	0	0	1
14	19:09:37.482	Peticiones 1-38	HTTP Request	1	Failure	2578	0	0	1
15	19:09:37.500	Peticiones 1-39	HTTP Request	1	Failure	2578	0	0	1
16	19:09:36.989	Peticiones 1-7	HTTP Request	516	Success	875	235	516	1
17	19:09:36.927	Peticiones 1-3	HTTP Request	581	Success	875	235	581	0
18	19:09:37.083	Peticiones 1-13	HTTP Request	425	Success	875	235	425	1
19	19:09:37.006	Peticiones 1-8	HTTP Request	502	Success	875	235	502	0

Ilustración 11 "Resultado de árbol de la prueba de stress"

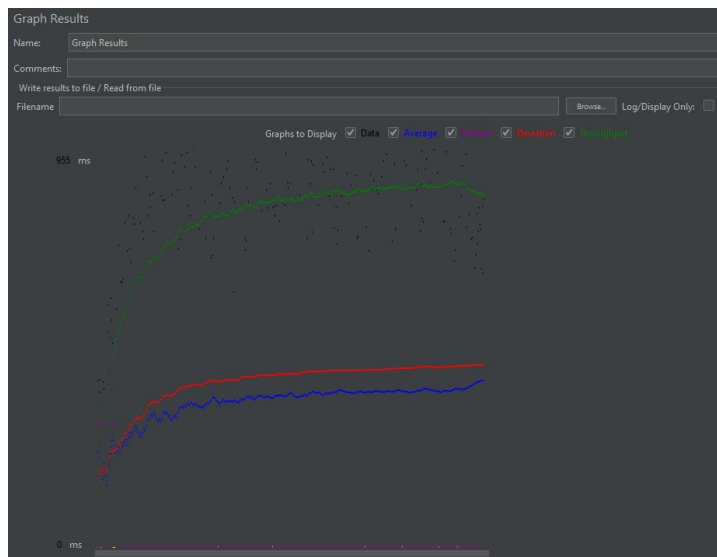


Ilustración 12 "Grafico de la prueba de stress"

TABLA XI
PRUEBA DE CALIDAD DE CODIGO BACKEND

Acciones realizadas	Herramienta	Resultado
El sistema fue sometido a una prueba de calidad de código, en la plataforma SonarCloud donde se evalúa la seguridad, fiabilidad y Mantenibilidad del código en base a buenas prácticas.	SonarCloud	Como se observa en la Ilustración 13, el código está calificado como aceptado, sin tener problemas de seguridad y fiabilidad. Sin embargo, si se obtuvieron 3 problemas de Mantenibilidad en el sistema, que al verlo a detalle se identificó que el problema se puede solucionar con renombramientos en las variables afectadas, lo cual no afecta negativamente al sistema como se puede observar en la Ilustración 14.

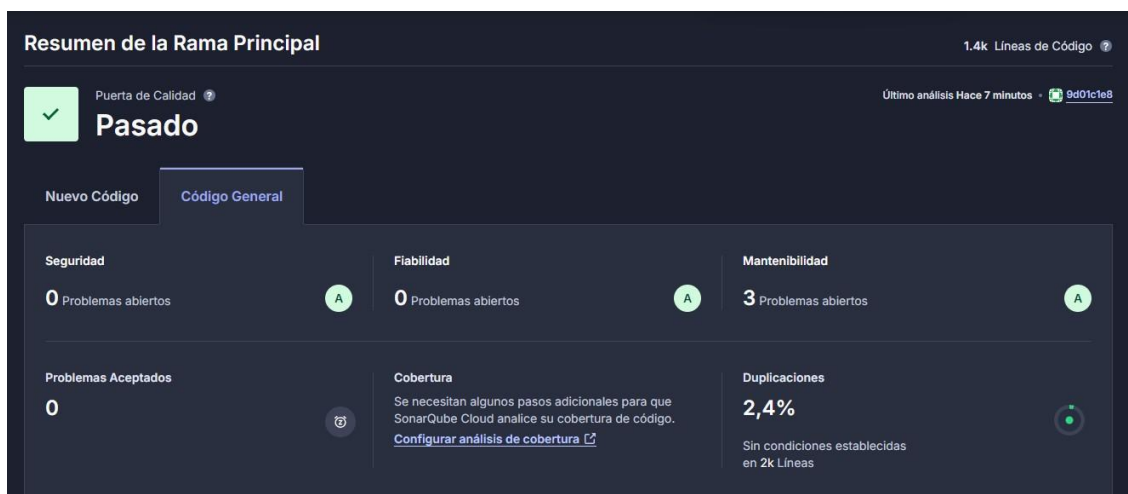


Ilustración 13 "Resumen de calidad de código backend"

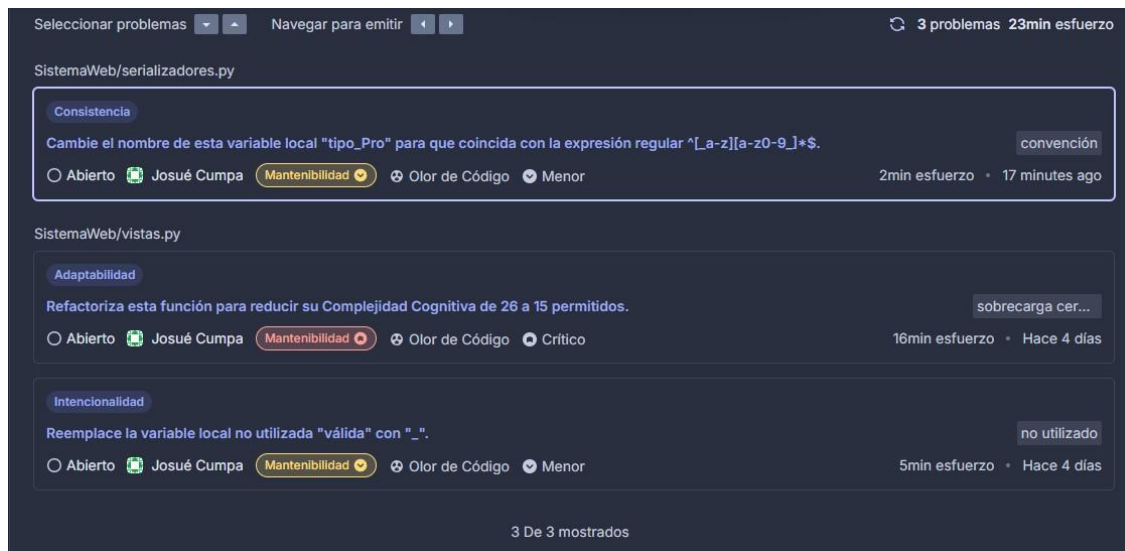


Ilustración 14 "Problemas de Mantenibilidad detectados en el backend"

Discusión

En la presente investigación, resulta crucial resaltar las decisiones metodológicas adoptadas, en contraste con estudios previos. El estudio realizado por [8] implemento de redes neuronales permitió lograr un modelo con un 95% de precisión, a diferencia de los otros modelos evaluados que estuvieron por debajo del 70%, lo que resalta el potencial de esta técnica para abordar problemas donde las relaciones entre las variables son complejas y no lineales. En esta investigación, el uso de redes neuronales también fue una elección clave debido a su capacidad para manejar grandes cantidades de datos y su flexibilidad en la captura de patrones complejos, se pudo conseguir una precisión del 94%.

La selección de un modelo adecuado para predecir la calidad de productos es crucial, como se evidencia en las investigaciones previas. En el estudio [9] desarrollaron un modelo predictivo utilizando minería de datos y aprendizaje automático, logrando una precisión del 95% en la predicción de la calidad de piezas de polímero para autos, centrándose en parámetros como el diseño del molde, espesor, temperatura y tiempo de ciclo, siendo este último el factor decisivo para la calidad. Por otro lado, en la investigación [10] en la fabricación de tableros de madera utilizaron redes neuronales y otros algoritmos, obteniendo una precisión del 69.95% al predecir la adherencia entre chapas, un proceso complejo debido a la variabilidad de la madera y las condiciones ambientales. Es por ello por lo que en esta investigación se realizó la comparación de modelos predictivos donde destaco las redes neuronales como el que mejor se adaptó para predecir y se identificó que los parámetros que más influyen en la calidad de los módulos de plásticos de la banda modular son las temperaturas y el ciclo de moldeo.

A diferencia de las metodologías implementadas en los otros estudios, se concuerda con el trabajo de [11] donde se resalta la metodología CRISP-DM se adapta mejor para la construcción del modelo predictivo el desarrollo de la aplicación web. Además, en la investigación [12] [13] resalta la importancia de validar el sistema de detección según la norma ISO 25010.

Se siguió un enfoque específico con la metodología CRISP-DM, al estar relacionado en investigaciones donde los modelos alcanzan una precisión estándar aceptable, junto con las redes neuronales. Además, que se validó todo el sistema web según la ISO 25010, respaldando el correcto funcionamiento del sistema, en base una evaluación de las funcionalidades principales.

Conclusiones

Se logró identificar la técnica de imputación adecuada para el balanceo de datos tras analizar el data set y determinar que la cantidad de datos a imputar era de 55, lo cual es relativamente baja garantizando la integridad de los datos; considerando la dispersión de los datos que cada técnica podía manejar, se concluyó que la imputación por media era la opción más adecuada, ya que mostró la menor dispersión en comparación con las otras técnicas evaluadas. Por lo que, se decidió utilizar esta técnica para el desarrollo del proyecto.

Luego del análisis exhaustivo de diversos modelos de inteligencia artificial utilizando las librerías Scikit-learn y Tensor Flow, se determinó que el modelo de Redes Neuronales de Tensor Flow es la mejor opción para esta investigación. Este modelo demostró un R^2 de 0.82, el valor más alto entre los modelos evaluados, y un MSE de 13.68, el más bajo, lo que indica una alta precisión y menor error en la predicción en comparación con otros métodos. Estos resultados evidencian la capacidad de las redes neuronales para capturar patrones complejos en los datos, siendo especialmente efectivas en el contexto de la fabricación de la banda modular para mejorar la calidad del producto.

Al haber aplicado la metodología CRISP-DM de manera exhaustiva, lo cual nos ayudó a una correcta preparación de la data e implementación de los modelos; se logró que el modelo seleccionado llegara a R^2 de 94% y un MSE de 4.66, lo cual nos indica una precisión aceptable del modelo. Este resultado refuerza la confianza en la elección de las redes neuronales como herramienta para abordar las necesidades del proyecto.

Se validó el sistema conforme a la norma ISO 25010, enfocándose en la subcategoría de adecuación funcional, para evaluar si el sistema cumple con los requisitos específicos y las expectativas del entorno de la fábrica de bandas modulares. El resultado obtenido fue un

91.67%, lo que clasifica el nivel de adecuación funcional como "Alto". Esto refleja que el sistema tiene una alta aceptación y un rendimiento consistente, garantizando que las funcionalidades implementadas se alinean con las necesidades y expectativas de los usuarios finales, asegurando así su eficacia y confiabilidad en el entorno industrial.

Recomendaciones

En esta investigación, se emplearon datos históricos para simular y comprender cómo estas afectan a la calidad del producto. No obstante, se sugiere que el sistema pueda evolucionar incorporando un módulo de monitoreo en tiempo real, el cual proporcionaría información actualizada al instante. Esta funcionalidad beneficiaría especialmente a la empresa, brindándole la capacidad de anticiparse a posibles variaciones en la calidad durante el proceso de fabricación.

Para enriquecer estudios futuros y robustecer la precisión del modelo, se recomienda llevar a cabo un análisis exhaustivo de aspectos que pueden influir en la calidad del producto, tales como el estado de la maquinaria o el desgaste de componentes críticos. En esta investigación no fue posible debido a la falta de algunos de estos datos en la empresa, pero su incorporación en el análisis permitiría construir una visión más amplia de los factores que intervienen en el proceso de producción, fortaleciendo así la capacidad del modelo para detectar los parámetros que más afectan la calidad del producto.

Además, se recomienda utilizar el algoritmo propuesto con un mayor volumen de datos de fabricación, lo que permitiría obtener resultados más precisos y una mejor identificación de los parámetros que influyen en la calidad del producto. Adicional a esto implementar un modelo para predecir la cantidad de piezas rechazadas, puesto que esta es muy variable y tiene una influencia grande en lo que es la calidad del producto final.

Finalmente, se sugiere adoptar la norma ISO 25010 para validar futuros cambios en el sistema, garantizando así una uniformidad en el desarrollo y evaluación de este, lo que aseguraría una mejora continua y estructurada.

Referencias

- [1] P. Agrawal, S. Navgotri y P. Nagesh, "Impact of emerging technologies on digital manufacturing: Insights from literature review", *Mater. Today: Proc.*, marzo de 2023. Accedido el 6 de abril de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.matpr.2023.03.187>
- [2] D. Zhang, M. Xie, H. Yan y Q. Liu, "Resilience dynamics modeling and control for a reconfigurable electronic assembly line under spatio-temporal disruptions", *J. Manuf. Syst.*, vol. 60, pp. 852–863, julio de 2021. Accedido el 6 de abril de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.jmsy.2021.01.011>
- [3] M. Sharma, R. D. Raut, R. Sehrawat y A. Ishizaka, "Digitalisation of manufacturing operations: The influential role of organisational, social, environmental, and technological impediments", *Expert Syst. with Appl.*, p. 118501, agosto de 2022. Accedido el 8 de abril de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.eswa.2022.118501>
- [4] CEPAL, "Tecnologías digitales para un nuevo futuro ", 2022. Accedido el 7 de abril de 2023. [En línea]. Disponible: https://repositorio.cepal.org/bitstream/handle/11362/46816/1/S2000961_es.pdf
- [5] "Innovación en la industria manufacturera y en las empresas de servicios intensivos en conocimiento". Ministerio de producción. 2018. Accedido el 8 de abril del 2023. [En línea]. Disponible: <https://ogeiee.produce.gob.pe/index.php/en/shortcode/oeedocumentos-publicaciones/publicaciones-anales/item/939-innovacion-en-laindustria-manufacturera-y-en-las-empresas-de-servicios-intensivos-enconocimiento>
- [6] J. Paredes Roldán, *Planificación y control de la producción*, IDIUC, Instituto de Investigaciones, Universidad de Cuenca, 2001.
- [7] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA: Morgan Kaufmann, 2011. [En línea]. Disponible: https://books.google.com.pe/books?id=bDtLM8CODsQC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- [8] P. H. Yi et al, "Incorporating Machine Learning in Computer-Aided Molecular Design for Fragrance Molecules," *Processes*, vol. 10, (9), pp. 1767, 2022.

- Accedido el 10 de abril de 2023. [En línea]. DOI: <https://doi-org.usat.lookproxy.com/10.3390/pr10091767>.
- [9] R. Kumari, K. Saini y A. Anand, "Predictive analytics to improve the quality of polymer component manufacturing", *Measurement: Sensors*, p. 100428, septiembre de 2022. Accedido el 2 de mayo de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.measen.2022.100428>
- [10] C. Urra y M. Ramos, "Un enfoque de machine learning para la predicción de la calidad de tableros contrachapados", *Maderas-Ciencia y Tecnología*, vol. 25, Abril de. 2023.
- [11] J. E. Bravo Gonzáles. "Sistema de soporte a las decisiones para apoyar la planificación de compras en una droguería de Chiclayo," tesis de licenciatura, Fac. de Ingeniería, Univ. USAT, Chiclayo, Perú, 2024. [En línea]. Disponible: <http://hdl.handle.net/20.500.12423/7555>
- [12] C. E. Delgado Ballena. "Sistema de predicción de comportamiento de clientes siguiendo su historial crediticio del Banco Azteca," tesis de licenciatura, Fac. de Ingeniería, Univ. USAT, Chiclayo, Perú, 2024. [En línea]. Disponible: <http://hdl.handle.net/20.500.12423/7369>
- [13] I. P. Lopez Sotelo. "Sistema inteligente para apoyar en la identificación de deficiencias nutricionales del cultivo de café," tesis de licenciatura, Fac. de Ingeniería, Univ. USAT, Chiclayo, Perú, 2023. [En línea]. Disponible: <http://hdl.handle.net/20.500.12423/6862>
- [14] C. Lorentz, *Machine Learning con MATLAB: Técnicas de Aprendizaje No Supervisado*. Independently Published, 2020.
- [15] R. Benítez, G. Escudero, S. Kanaan, and D. Masip Rodó, *Inteligencia Artificial Avanzada*. Barcelona, España: Editorial UOC, 2014.
- [16] C. Lopez Perez y D. Santin Gonzales, *Minería de datos. Técnicas y herramientas*, Ediciones Paraninfo, S.A.
- [17] J. Bobadilla, *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Bogotá, Colombia: Ediciones de la U, 2021.
- [18] OCDE, *Manual de Frascati: Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental*, 7ª ed. París, Francia: OCDE, 2015.

- [19] M.Marques, Minería de datos. La metodología CRISP-DM de ibm. [En línea]. Disponible:https://www.google.com.pe/books/edition/Mineria_de_Datos_La_Metodologia_Cris_DM/kMy5rQEACAAJ?hl=es-419
- [20] V. Galán Cortina, “Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario”, proyecto fin de carrera, Univ. Carlos III Madr. Esc. Politec. Super. Ing. En Inform., 2015.
- [21] J. V. Chávez Andrade, “Estandarización de los procesos de desarrollo de software utilizando buenas prácticas de programación y scrum como marco de trabajo ágil en departamentos de ti”, Maestría, Univ. Tec. Ambato, Ambato – Ecuador, 2019. Accedido el 9 de septiembre de 2024. [En línea]. Disponible: https://repositorio.uta.edu.ec/bitstream/123456789/29604/1/Tesis_t1561m_si.pdf

Anexos

**ANEXO N° 01. CARTA DE ACEPTACIÓN PARA LA EJECUCIÓN DEL
PRESENTE PROYECTO**

Representante Comercial
INDUSTRIA DE BANDAS MODULARES S.A.C.
R.U.C.: 20607792187



"Año de la Unidad, La Paz y el Desarrollo"

Chiclayo, 24 de junio del 2023

CARTA N°0001-2023

Señor
Ing. Hilder Mera Castro

Director de Escuela.
INGENIERIA DE SISTEMAS Y COMPUTACION

Por el presente medio hago constar que el estudiante Cumpa chancafe Josué Giancarlo identificado con DNI 74769120, de la escuela de INGENIERIA DE SISTEMAS Y COMPUTACION, ha sido aceptado la realización de su proyecto de tesis, que lleva como título. IMPLEMENTACIÓN DE UN SISTEMA DE PREDICCIÓN DE CALIDAD DEL PRODUCTO EN EL ÁREA DE PRODUCCIÓN DE UNA FÁBRICA DE BANDAS MODULARES DE LA REGIÓN LAMBAYEQUE.

Para la continuidad del proyecto del alumno en mención, le brindamos la información necesaria, las facilidades pertinentes, permisos y apoyo que amerite el caso.

Atentamente,

INDUSTRIA DE BANDAS MODULARES S.A.C.

Fernando J. Cortez Llontop
GERENTE

Industria De Bandas Modulares SAC
RUC: 20607792187

+51 + 074 + 759700 www.abraham.pe administracion@abraham.pe

Mz. 27 - Lte.: 4- Urb. Chosica de Norte / La Victoria / Chiclayo / Lambayeque

ANEXO N° 02. CONSTANCIA DE APROBACION DE PRODUCTO ACREDITABLE

Representante Comercial
INDUSTRIA DE BANDAS MODULARES S.A.C.
R.U.C.: 20607792187

**CONSTANCIA DE APROBACION DE PRODUCTO ACREDITABLE**

Por este medio de la presente certificamos la aprobación de la aplicación web presentada por el estudiante **Cumpa Chancafe Josue Giancarlo** identificado con el **DNI 74769120**, de la escuela de **INGENIERIA DE SISTEMAS Y COMPUTACION DE LA UNIVERSIDAD CATOLICA SANTO TORIBIO DE MOGROVEJO DE CHICLAYO**, como producto acreditable de su trabajo de investigación de fin de grado, cumpliendo los requisitos establecidos propuestos en el proyecto inicial.

Se expide la presente constancia a petición del interesado para los fines que crea correctamente.



Atentamente:


INDUSTRIA DE BANDAS MODULARES S.A.C.



Fernando J. Cortez Llontop
GERENTE

CORTEZ LLONTOP FERNANDO JAVIER
DNI N° 71029342
Gerente General

+51 + 074 + 759700  www.abraham.pe  administracion@abraham.pe

 Mz. 27 - Lte.: 4- Urb. Chosica de Norte / La Victoria / Chiclayo / Lambayeque

ANEXO N° 03 GUIA DE ENTREVISTA

ENCUESTADO: jefe de logística – jefe de producción – jefe de almacén

ENTREVISTADOR: Cumpa Chancafe Josue Giancarlos

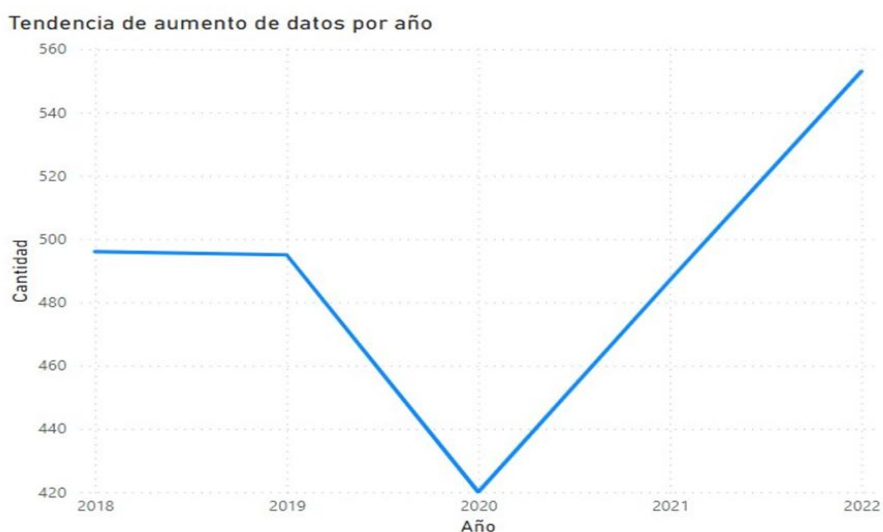
OBJETIVO: Conocer el contexto y las variables necesarias para determinar la calidad del producto.

FECHA: 03-04-2023

PREGUNTAS:

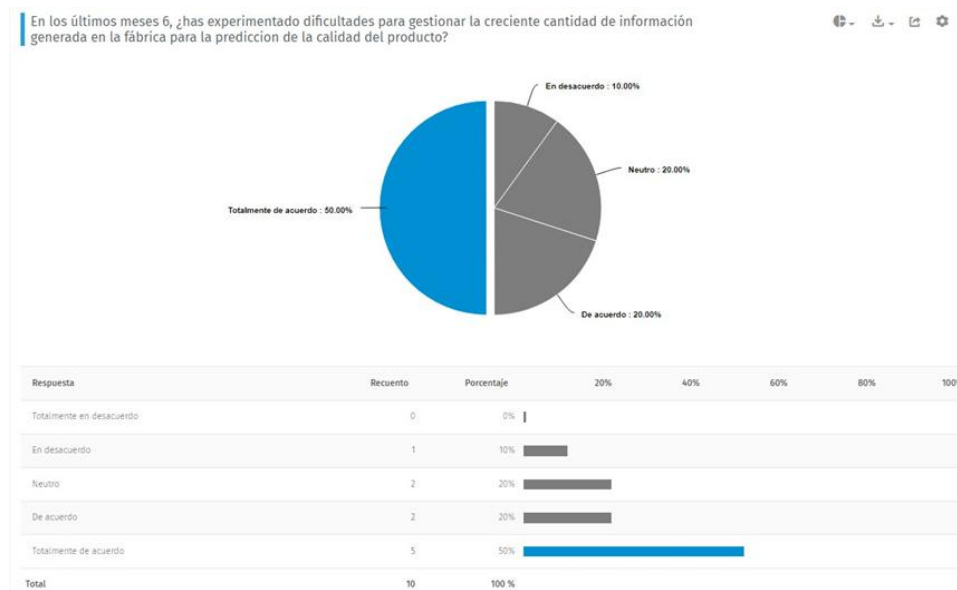
1. ¿Qué variables influyen en la calidad del producto?
2. ¿Cuál es el principal problema que se tiene en la fabricación del producto?
3. ¿En qué situaciones la producción o la calidad del producto se han visto afectadas?
¿Con que frecuencia?
4. ¿Existe una segmentación definida en los materiales y aditivos que se utilizan en la fabricación del producto?
6. ¿Cuáles son los materiales o aditivos que más utilizan en la producción del producto?
7. ¿Cuentan con data histórica de producciones realizadas?
8. ¿Manejan información completa en todas las variables de análisis de las producciones?
9. ¿Realizan el análisis de la data historia para determinar la calidad del producto? 10. ¿Qué software utilizan para analizar la calidad del producto e identificar los potenciales riesgos de fallas en la producción?
11. ¿Considera necesario conocer la probabilidad de calidad del producto?
12. ¿Cuánto tiempo toma evaluar los parámetros necesarios para la producción con los procesos actuales?

ANEXO N° 04. TENDENCIA DE AUMENTO DE DATOS POR AÑO



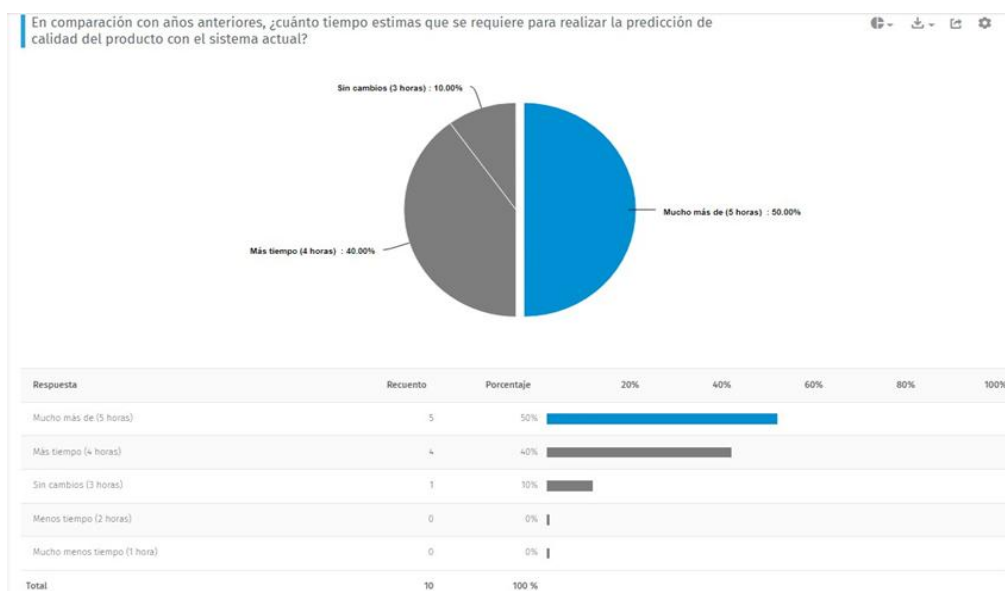
Fuente: Base de datos de la empresa

ANEXO N° 05. DIFICULTAD PARA GESTIONAR LA INFORMACIÓN EN LOS ÚLTIMOS 6 MESES



Fuente: Encuesta realiza a los empleados de logística de la empresa

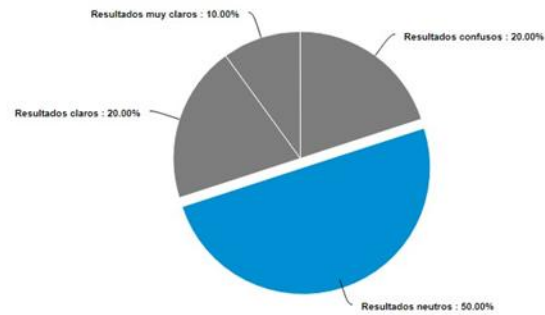
ANEXO N° 06. TIEMPO ESTIMADO QUE SE REQUIERE PARA REALIZAR LA PREDICCIÓN CON EL SISTEMA ACTUAL.



Fuente: Encuesta realiza a los empleados de logística de la empresa

ANEXO N° 07. RESULTADOS QUE PROPORCIONA EL SISTEMA ACTUAL.

¿El sistema actual de predicción de calidad del producto proporciona resultados claros y fáciles de interpretar?

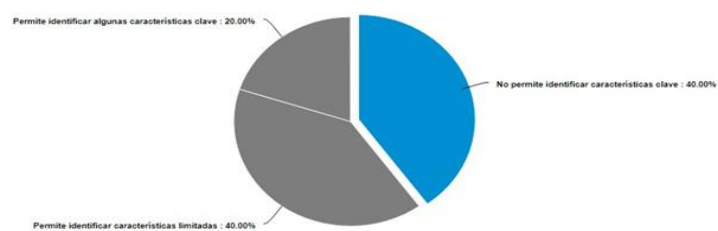


Respuesta	Recuento	Porcentaje	20%	40%	60%	80%	100%
Resultados muy confusos	0	0%					
Resultados confusos	2	20%	██████████				
Resultados neutros	5	50%	████████████████████				
Resultados claros	2	20%	██████████				
Resultados muy claros	1	10%	██████				
Total	10	100 %					

Fuente: Encuesta realiza a los empleados de logística de la empresa

ANEXO N° 08. ¿EL SISTEMA ACTUAL PERMITE IDENTIFICAR LAS CARACTERÍSTICAS QUE AFECTAN A LA CALIDAD DEL PRODUCTO?

¿El sistema actual permite identificar y medir de manera precisa las características clave que afectan la calidad del producto?

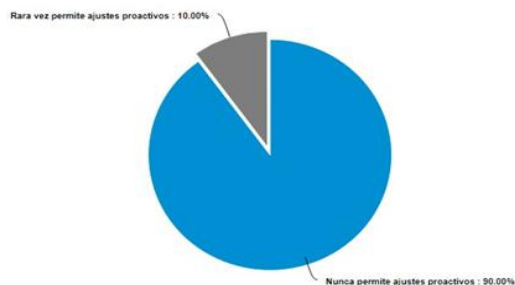


Respuesta	Recuento	Porcentaje	20%	40%	60%	80%	100%
No permite identificar características clave	4	40%		██████████			
Permite identificar características limitadas	4	40%		██████████			
Permite identificar algunas características clave	2	20%	██████				
Permite identificar características significativas	0	0%					
Permite identificar todas las características clave	0	0%					
Total	10	100 %					

Fuente: Encuesta realiza a los empleados de logística de la empresa

ANEXO N° 09. DISPONIBILIDAD DE AJUSTAR LOS PARÁMETROS EN LA FABRICACIÓN

¿El sistema actual de predicción de calidad del producto permite realizar ajustes proactivos en el proceso de fabricación para mejorar la calidad?

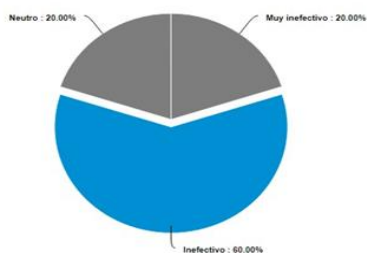


Respuesta	Recuento	Porcentaje	20%	40%	60%	80%	100%
Nunca permite ajustes proactivos	9	90%	<div style="width: 90%;"></div>				
Rara vez permite ajustes proactivos	1	10%	<div style="width: 10%;"></div>				
Ocasionalmente permite ajustes proactivos	0	0%	<div style="width: 0%;"></div>				
Frecuentemente permite ajustes proactivos	0	0%	<div style="width: 0%;"></div>				
Siempre permite ajustes proactivos	0	0%	<div style="width: 0%;"></div>				
Total	10	100%					

Fuente: Encuesta realiza a los empleados de logística de la empresa

ANEXO N° 10. VALORACIÓN DE EFECTIVIDAD DEL SISTEMA ACTUAL

¿En general, cómo calificarías la efectividad del sistema actual de predicción de calidad del producto?



Respuesta	Recuento	Porcentaje	20%	40%	60%	80%	100%
Muy inefectivo	2	20%	<div style="width: 20%;"></div>				
Inefectivo	6	60%	<div style="width: 60%;"></div>				
Neutro	2	20%	<div style="width: 20%;"></div>				
Efectivo	0	0%	<div style="width: 0%;"></div>				
Muy efectividad	0	0%	<div style="width: 0%;"></div>				
Total	10	100%					

Fuente: Encuesta realiza a los empleados de logística de la empresa