



**FACULTAD DE POSTGRADO
TRABAJO FINAL DE GRADUACIÓN**

**APLICACIÓN DE MODELOS DE APRENDIZAJE
AUTOMÁTICO PARA EL ANÁLISIS DE SERIES
TEMPORALES Y PRONÓSTICO DE ANOMALÍAS EN LA
NÓMINA DIARIA EN UNA PLANTA DE MANUFACTURA**

SUSTENTADO POR:

**GUSTAVO ERNESTO AVILA GRIFFIN
SEBASTIAN NUÑEZ FLORES**

PREVIA INVESTIDURA AL TÍTULO DE

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

SAN PEDRO SULA, CORTES, HONDURAS, C.A.

NOVIEMBRE, 2023

UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

UNITEC

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

RECTORA

ROSALPINA RODRÍGUEZ GUEVARA

PRORECTOR/ SECRETARIO GENERAL

ROGER MARTÍNEZ MIRALDA

VICERRECTOR ACADÉMICO

JAVIER ABRAHAM SALGADO LEZAMA

DIRECTORA NACIONAL DE POSTGRADO

ANA DEL CARMEN RETTALLY

**APLICACIÓN DE MODELOS DE APRENDIZAJE
AUTOMÁTICO PARA EL ANÁLISIS DE SERIES
TEMPORALES Y PRONÓSTICO DE ANOMALÍAS EN LA
NÓMINA DIARIA EN UNA PLANTA DE
MANUFACTURA**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE**

MÁSTER EN

ANALÍTICA DE NEGOCIOS

**ASESOR METODOLÓGICO
ALBA GABRIELA GARAY ROMERO**

**ASESOR TEMÁTICO
KEVIN EDUARDO FUNEZ**

MIEMBROS DE LA TERNA:

**MARIO ALBERTO GALLO SANDOVAL
ERICK LEONARDO IZAGUIRRE ARDÓN
HENRY DIOMEDES OSORTO NÚÑEZ**

DERECHOS DE AUTOR

© Copyright 2023
Gustavo Ernesto Avila Griffin
Sebastián Núñez Flores

Todos los derechos son reservados.

**AUTORIZACIÓN DEL AUTOR(ES) PARA LA CONSULTA, REPRODUCCIÓN PARCIAL O TOTAL Y
PUBLICACIÓN FÍSICA Y ELECTRÓNICA DEL TEXTO COMPLETO DEL TRABAJO FINAL DE GRADUACIÓN DE
UNITEC Y CESIÓN DE DERECHOS PATRIMONIALES**

Señores

**CENTRO DE RECURSOS PARA
EL APRENDIZAJE Y LA INVESTIGACION (CRAI)
UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA (UNITEC)**

Tegucigalpa

Estimados Señores:

Nosotros, Gustavo Ernesto Ávila Griffin y Sebastián Núñez Flores de San Pedro Sula, autor(es) del trabajo de pregrado titulado: APLICACIÓN DE MODELOS DE APRENDIZAJE AUTOMÁTICO PARA EL ANÁLISIS DE SERIES TEMPORALES Y PRONÓSTICO DE ANOMALÍAS EN LA NÓMINA DIARIA EN UNA PLANTA DE MANUFACTURA, presentado y aprobado en el noviembre/2023, como requisito previo para optar al título de máster en ANALÍTICA DE NEGOCIOS (en lo sucesivo, el “Trabajo Final de Graduación”) y reconociendo que la presentación del presente documento forma parte de los requerimientos establecidos del programa de maestrías de la Universidad Tecnológica Centroamericana (UNITEC) y del Centro Universitario Tecnológico (CEUTEC), por este medio **AUTORIZAMOS** a la Universidad Tecnológica Centroamericana (UNITEC) y el Centro Universitario Tecnológico (CEUTEC), para que:

- 1) A través de sus Centros Asociados y Bibliotecas de los “Centros de Recursos para el Aprendizaje y la Investigación (CRAI)”, para que, con fines académicos, puedan libremente registrar, copiar o utilizar la información contenida en él, con fines educativos, investigativos o sociales. Asimismo, para que exponga mi trabajo como medio didáctico en los Centros de Recursos para el Aprendizaje y la Investigación (CRAI o Biblioteca), y con fines académicos permita a los usuarios de dichos centros su consulta y acceso mediante catálogos electrónicos, repositorios académicos nacionales o internacionales, página web institucional, así como medios electrónicos en general, internet, intranet, DVD, u otro formato conocido o por conocer, así como integrados en programas de cooperación bibliotecaria académicos, que permitan mostrar al mundo la producción académica de la Universidad a través de la visibilidad de su contenido.
- 2) De conformidad con lo establecido en la Ley de Derechos de Autor y de los Derechos Conexos de la República de Honduras, se autoriza para que permita copiar, reproducir o transferir información del Proyecto de Graduación, conforme su uso educativo y debiendo citar en todo momento la fuente de información; esto permitirá ampliar los conocimientos a las personas que hagan uso del mismo, siempre y cuando resguarden la completa información textual o paráfrasis de esta.

Asimismo, en nuestra calidad de estudiante y/o autor del Trabajo Final de Graduación acepto que UNITEC/CEUTEC no se hace responsable del uso, reproducciones, venta y distribuciones de todo tipo de fotografías, imágenes, grabaciones, o cualquier otro tipo de presentación relacionado con el Trabajo Final

de Graduación que el mismo autor distribuya antes y después de la entrega del documento a la Universidad.

Finalmente, declaramos bajo fe de juramento, conociendo las consecuencias penales que conlleva el delito de perjurio: que somos autor(es) del presente Trabajo Final de Graduación, que el contenido de dicho trabajo es obra original los suscritos y de la veracidad de los datos incluidos en el documento. Eximo a UNITEC/CEUTEC; así como el Tutor y Lector que han revisado el presente, por las manifestaciones y/o apreciaciones personales incluidas en el mismo, de cualquier responsabilidad por su autoría o cualquier situación de perjuicio que se pudiera presentar.

De conformidad con lo establecido en los artículos 9.2, 18, 19, 35 y 62 de la Ley de Derechos de Autor y de los Derechos Conexos; los derechos morales pertenecen al autor y son personalísimos, irrenunciables, imprescriptibles e inalienables, asimismo, por tratarse de una obra colectiva, CEDEMOS de forma ilimitada y exclusiva a la UNITEC/CEUTEC la titularidad de los derechos patrimoniales que surjan o se deriven del Trabajo Final de Graduación. Es entendido que cualquier copia o reproducción del presente documento con fines de lucro no está permitida sin previa autorización por escrito de parte de UNITEC/CEUTEC.

En fe de lo cual, se suscribe el presente documento en la ciudad de San Pedro Sula a los 13 días del mes de noviembre del 2023.


Gustavo Ernesto Ávila Griffin
22213018


Sebastián Núñez Flores
22213065



FACULTAD DE POSTGRADO

APLICACIÓN DE MODELOS DE APRENDIZAJE AUTOMÁTICO PARA EL ANÁLISIS DE SERIES TEMPORALES Y PRONÓSTICO DE ANOMALÍAS EN LA NÓMINA DIARIA EN UNA PLANTA DE MANUFACTURA

**Gustavo Ernesto Avila Griffin
Sebastián Núñez Flores**

Resumen

En el competitivo ámbito de la industria manufacturera, alcanzar la excelencia operativa es esencial para garantizar eficiencia y rentabilidad. La optimización de procesos y recursos juega un papel clave en el éxito de una planta de manufactura, lo que hace que la detección y gestión de valores atípicos cobren relevancia como desafío estratégico. Dentro de este contexto, la revisión de nóminas adquiere un rol crucial al asegurar la precisión en los pagos y evaluar la eficiencia operativa de la planta. Este proyecto investigativo busca implementar un sistema de detección que identifique errores involuntarios e irregularidades fraudulentas en los datos de eficiencia de la revisión. Al corregir estos valores atípicos, se resguarda la integridad de los pagos y se generan ahorros con efectos tangibles. Dado que la compensación de los operarios depende de su eficiencia, identificar desviaciones tempranamente optimiza costos y recursos. La investigación explora en detalle las metodologías de detección de valores atípicos aplicables a los datos de eficiencia en la revisión de nóminas y sus procesos correctivos. Además, analiza los beneficios económicos y operativos, reafirmando la importancia de la precisión en la gestión de nóminas en un entorno industrial en constante evolución. En síntesis, esta tesis yace en la confluencia de la ingeniería industrial, la analítica de datos y la gestión de recursos humanos. Su propósito es asegurar la calidad de los datos de eficiencia en la revisión de nóminas y promover la eficacia operativa en la planta de manufactura, garantizando tanto la precisión financiera como la eficiencia operativa en las plantas de manufactura.

Palabras claves: (Analítica de datos, Manufactura, Nóminas, Valores Atípicos)



GRADUATE SCHOOL

APPLICATION OF MACHINE LEARNING MODELS FOR TIME SERIES ANALYSIS AND ANOMALY FORECASTING IN DAILY PAYROLL AT A MANUFACTURING PLANT

**Gustavo Ernesto Ávila Griffin
Sebastián Núñez Flores**

Abstract

In the competitive realm of the manufacturing industry, achieving operational excellence is essential to ensure efficiency and profitability. Process and resource optimization play a key role in a manufacturing plant's success, highlighting the significance of detecting and managing outliers as a strategic challenge. Within this context, payroll review takes on a crucial role in ensuring payment accuracy and assessing the operational efficiency of the plant. This research project aims to implement a detection system that identifies unintentional errors and fraudulent irregularities in the efficiency data of the review process. By correcting these outliers, the integrity of payments is safeguarded, and tangible savings are generated. Since operators' compensation depends on their efficiency, early identification of deviations optimizes costs and resources. This research explores in detail the outlier detection methodologies applicable to efficiency data in payroll review and their corrective processes. Additionally, it analyzes the economic and operational benefits, reaffirming the importance of accuracy in payroll management in an ever-evolving industrial environment. In summary, this thesis lies at the intersection of industrial engineering, data analytics, and human resource management. Its purpose is to ensure the quality of efficiency data in payroll review and promote operational efficiency in the manufacturing plant, ensuring both financial accuracy and operational efficiency in manufacturing plants.

Key Words: (Data analytics, Manufacturing, Outliers, Payroll Review)

DEDICATORIA

Dedico este logro a mi amada familia: a mi esposa Ela Rivera y a mis queridos hijos Ana, Camila, Gustavo y Ricardo. Su amor, apoyo y paciencia han sido mi mayor motivación a lo largo de este viaje. También dedico este trabajo a mi madre Leticia Griffin, quien siempre creyó en mí y me inspiró a alcanzar mis metas. Y a mi difunto padre Gustavo Avila Galindo, cuyo legado de trabajo duro y determinación continúa guiándome en cada paso que doy. Gracias por ser mi fuente de fortaleza y por compartir este éxito conmigo.

- Gustavo Ernesto Avila Griffin

Deseo dedicar este logro a mi madre Madelin Flores Peña. Ella ha estado desde un inicio en este viaje apoyándome y dándome las fuerzas necesarias para no rendirme y salir adelante. Este logro va dedicado a ella, ya que, ha sido un pilar muy importante en mi educación, ambición y deseo de estar constantemente preparándome. Como siempre me ha recalado e inculcado, no hay mejor inversión que la educación y la preparación que me puedo dar para conseguir más logros y cumplir más objetivos. Gracias por la confianza, el constante apoyo y por nunca dudar de mis capacidades para alcanzar este y muchos logros más que vendrán.

- Sebastián Núñez Flores

AGRADECIMIENTO

En este momento significativo, deseo expresar mi profundo agradecimiento a Dios, cuya gracia y guía han iluminado mi camino durante este desafiante viaje académico.

A mi amada familia, mi esposa Ela, mis queridos hijos Ana, Camila, Gustavo y Ricardo y mi madre Leticia, les estoy eternamente agradecido por su amor incondicional, apoyo constante y comprensión. Sin ustedes, este logro no habría sido posible.

Quiero extender mi gratitud a mis compañeros Sebastián, Ariel, Daphne y Carlo. Hemos compartido conocimientos, desafíos y logros a lo largo de este proceso, y su amistad y colaboración han enriquecido esta experiencia y sin duda alguna me han rejuvenecido.

Por último, también agradezco a los dedicados catedráticos que han compartido su sabiduría y experiencia conmigo. Sus enseñanzas y orientación han sido fundamentales para mi crecimiento académico y profesional.

Este logro no es solo mío, sino el resultado del apoyo y la contribución de muchas personas importantes en mi vida. Gracias a todos los que han formado parte de este viaje.

- Gustavo Ernesto Avila Griffin

Agradezco enormemente a Dios por un logro más. De igual forma a mi querida segunda madre, Catalina Portillo, por ser constante apoyo y amor incondicional en este camino académico de postgrado, por ser ese pilar que me motiva ser más cada día, por escucharme y apoyarme en todo momento. A mi hermana Madelin Calix por siempre apoyarme en todo.

Quiero de igual manera, agradecer la oportunidad que me dio la maestría de no solo crear compañeros sino amigos, muchas gracias por los conocimientos, experiencias, risas y sabidurías compartidas Gustavo Avila, Ariel Cabrera, Daphne Escobar y Carlo Merlo. A mis amigos que me han brindado su apoyo incondicional Nataly Portillo, Marvin Zelaya, Eduardo Corrales y Javier López, les agradezco su apoyo incondicional. Agradezco a mi compañero de tesis, Gustavo Ávila, por ser además de compañero en este proceso, ser un compañero y amigo durante todo el proceso del postgrado y por siempre compartirnos sus conocimientos sin egoísmo alguno.

- Sebastián Núñez Flores

ÍNDICE DE CONTENIDO

DEDICATORIA	9
AGRADECIMIENTO	10
ÍNDICE DE CONTENIDO	11
INDICE DE FIGURAS.....	13
INDICE DE TABLAS	14
CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN	15
1.1 INTRODUCCIÓN	15
1.2 ANTECEDENTES DEL PROBLEMA	16
1.3 DEFINICIÓN DEL PROBLEMA	20
1.4 OBJETIVOS DEL PROYECTO.....	21
1.4.1 OBJETIVO GENERAL	21
1.4.2 OBJETIVOS ESPECIFICOS	21
1.5 JUSTIFICACIÓN.....	22
CAPÍTULO II. MARCO TEÓRICO	23
2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL	23
2.2 CONCEPTUALIZACIÓN	24
2.3 TEORÍAS DE SUSTENTO	26
2.3.1 BASES TEÓRICAS	26
2.3.2 METODOLOGÍAS DESARROLLADAS	32
2.3.3 INSTRUMENTOS UTILIZADOS.....	37
2.4 MARCO LEGAL	39
CAPÍTULO III. METODOLOGÍA	41
3.1 CONGRUENCIA METODOLÓGICA.....	41
3.1.1 MATRIZ METODOLÓGICA.....	42
3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO.....	42
3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES	43
3.2 ENFOQUE Y MÉTODOS	43
3.3 DISEÑO DE LA INVESTIGACIÓN	44
3.3.1 POBLACIÓN	44
3.3.2 MUESTRA	44

3.3.3	TECNICAS DE MUESTREO.....	44
3.4	TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	45
3.5	FUENTES DE INFORMACIÓN.....	46
3.5.1	FUENTES PRIMARIAS	46
3.5.2	FUENTES SECUNDARIAS.....	47
CAPÍTULO IV. RESULTADOS Y ANÁLISIS		47
4.1	INFORME DE PROCESO DE RECOLECCIÓN DE DATOS	47
4.2	RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS	49
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....		55
5.1	CONCLUSIONES	55
5.2	RECOMENDACIONES	57
CAPÍTULO VI. APLICABILIDAD.....		59
6.1	NOMBRE DE LA PROPUESTA	59
6.2	JUSTIFICACIÓN DE LA PROPUESTA	59
6.3	ALCANCE DE LA PROPUESTA	59
6.4	DESCRIPCIÓN Y DESARROLLO	61
6.4.1	DESCRIPCIÓN	61
6.4.2	DESARROLLO.....	62
6.5	MEDIDAS DE CONTROL	65
6.6	CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO	67
6.7	CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA	70
REFERENCIA BIBLIOGRAFICA		71
ANEXOS		74
Anexo 1 Código Utilizado.....		74
Anexo 2 Cotización AWS por servicio de almacenaje de data.....		78
Anexo 3 Cotización adquisición servicio Microsoft Office.....		79
Anexo 4 Cotización equipos para análisis.....		79
Anexo 5 Carta de Autorización de Empresa		80

INDICE DE FIGURAS

FIGURA 1. EVENTOS REGISTRADOS EFICIENCIA DIARIA >140%	17
FIGURA 2. DASHBOARD EFICIENCIAS > 140%	17
FIGURA 3. FASES PROYECTO CONTROL DE EFICIENCIAS	18
FIGURA 4. DETECCIÓN Y CORRECCIÓN DE EFICIENCIAS >140%	19
FIGURA 5. UTILIZACIÓN DE LIMITES DE CONTROL PARA DETECCIÓN DE ANOMALÍAS	38
FIGURA 6. ANOMALÍA DETECTADA. EFICIENCIA OPERADOR NO REPRESENTA LA CAPACIDAD REAL DEL ASOCIADO.	40
FIGURA 7. COMPORTAMIENTO ESPERADO DE LA RELACIÓN EFICIENCIA OPERADOR Y PLANTA	40
FIGURA 8. OPERACIONALIZACIÓN DE VARIABLES	43
FIGURA 9. ENFOQUE Y MÉTODOS	43
FIGURA 10. BASE DE DATOS DE EFICIENCIAS DIARIAS POR EMPLEADO	47
FIGURA 11. LIMPIEZA Y TRANSFORMACIÓN DE DATOS	48
FIGURA 12. VISUALIZACIÓN DE DATOS LIMPIOS	48
FIGURA 13. TRANSFORMACIÓN DE DATOS. TABLA CONVERTIDA EN FORMATO ANIDADADO (NESTED TIBBLE), AGRUPADO POR EMPLEADO Y CON LA SEPARACIÓN DE DATOS DE ENTRENAMIENTO Y PRUEBA.	49
FIGURA 14. SERIE MUESTRA	50
FIGURA 15. RESULTADOS SERIE MUESTRA	50
FIGURA 16. REPORTE DE EXACTITUD DE MODELOS ENTRENADOS	51
FIGURA 17. SELECCIÓN AUTOMÁTICA DE MEJOR MODELO POR SERIE	51
FIGURA 18. CÁLCULO DE AHORRO POTENCIAL	53
FIGURA 19. REENTRENAMIENTO DE MODELOS CON TOTAL DE DATOS HISTÓRICOS (REFIT)	53
FIGURA 20. EXPORTACIÓN DE PREDICCIONES EN CSV	53
FIGURA 21. DATA FORECAST	54
FIGURA 22. ADICIÓN DE DETECCIÓN Y CORRECCIÓN DE ANOMALÍAS	54
FIGURA 23. CRONOGRAMA DE IMPLEMENTACIÓN	67

INDICE DE TABLAS

TABLA 1. PROCESO METODOLÓGICO	41
TABLA 2. MATRIZ METODOLÓGICA.....	42
TABLA 3. PRESUPUESTO ESTIMADO	69
TABLA 4. CONCORDANCIA DE SEGMENTOS DE TESIS.....	70

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

En el ámbito competitivo de la industria manufacturera, alcanzar la excelencia operativa es esencial para garantizar eficiencia y rentabilidad. La optimización de procesos y recursos juega un papel clave en el éxito de una planta de manufactura, lo que hace que la detección y gestión de valores atípicos cobren relevancia como desafío estratégico.

Dentro de este panorama, la revisión de nóminas adquiere un rol crucial al asegurar la precisión en los pagos y evaluar la eficiencia operativa de la planta. En este contexto, la investigación se enfoca en una faceta esencial: la detección de valores atípicos en las eficiencias diarias de la revisión de nóminas en un entorno manufacturero.

Llevada a cabo por supervisores de unidad, la revisión implica verificar información crítica como horas trabajadas, tiempo no estándar, producción y datos afines. El propósito central es garantizar la exactitud de estos datos, asegurando una compensación justa para los trabajadores. Sin embargo, errores humanos y actos fraudulentos pueden distorsionar estas eficiencias, generando impactos financieros para empleados y la empresa.

El proyecto investigativo busca implementar un sistema de detección que identifique errores involuntarios e irregularidades fraudulentas en los datos de eficiencia de la revisión. Al corregir estos valores atípicos, se resguarda la integridad de los pagos y se generan ahorros con efectos tangibles. Dado que la compensación de los operarios depende de su eficiencia, identificar desviaciones tempranamente optimiza costos y recursos.

Dicha investigación explora en detalle las metodologías de detección de valores atípicos aplicables a los datos de eficiencia en la revisión de nóminas y sus procesos correctivos. Además, analiza los beneficios económicos y operativos, reafirmando la importancia de la precisión en la gestión de nóminas en un entorno industrial en constante evolución.

En síntesis, la tesis yace en la confluencia de la ingeniería industrial, la analítica de datos y la gestión de recursos humanos. El propósito es asegurar la calidad de los datos de eficiencia en la revisión de nóminas y promover la eficacia operativa en la planta de manufactura.

1.2 ANTECEDENTES DEL PROBLEMA

La detección de valores atípicos se ha convertido en un campo de interés para muchos investigadores y profesionales, y ahora es una de las principales tareas en la minería de datos de series de tiempo. La detección de valores atípicos se ha estudiado en una variedad de dominios de aplicación, como la detección de fraudes en tarjetas de crédito, la detección de intrusiones en ciberseguridad o el diagnóstico de fallas en la industria.

Se han derivado varias definiciones del término anomalía (outlier) y varios métodos de detección se han propuesto. Una definición comúnmente utilizada para anomalía es proporcionada por Hawkins, (1980): "Una observación que se desvía tanto de otras observaciones como para despertar sospechas de que fue generada por un mecanismo diferente."

Por lo que las anomalías pueden ser vistas como observaciones que no siguen un comportamiento esperado.

En el 2018, a raíz de la detección de promedios salariales elevados, se inició un proyecto de investigación que llevo a encontrar una oportunidad en las eficiencias de las planillas diarias. Por política interna y basado en un estudio realizado por el departamento de Ingeniería, la eficiencia máxima posible y permitida para pago en la compañía bajo el sistema de medición MTM o Medida del Tiempo de los Métodos (Method Time Measurement, en inglés), que representa una de las técnicas de medición de trabajo que permiten calcular tiempos teóricos de ejecución de actividades totalmente influenciados por el hombre, es de **140%**. Para ello, Ingeniería monitoreaba las eficiencias de pago de forma semanal y suponía controlado ese indicador de medición.

La oportunidad que se detectó en el desarrollo del proyecto fue que la medición que hacía ingeniería era a nivel semanal, es decir, eficiencias semanales no diarias que son las consideradas para el pago. A nivel semanal, Ingeniería reportaba menos de 10 casos cada semana, pero en realidad a diario se detectaban muchos más, la cifra se elevaba a los miles.

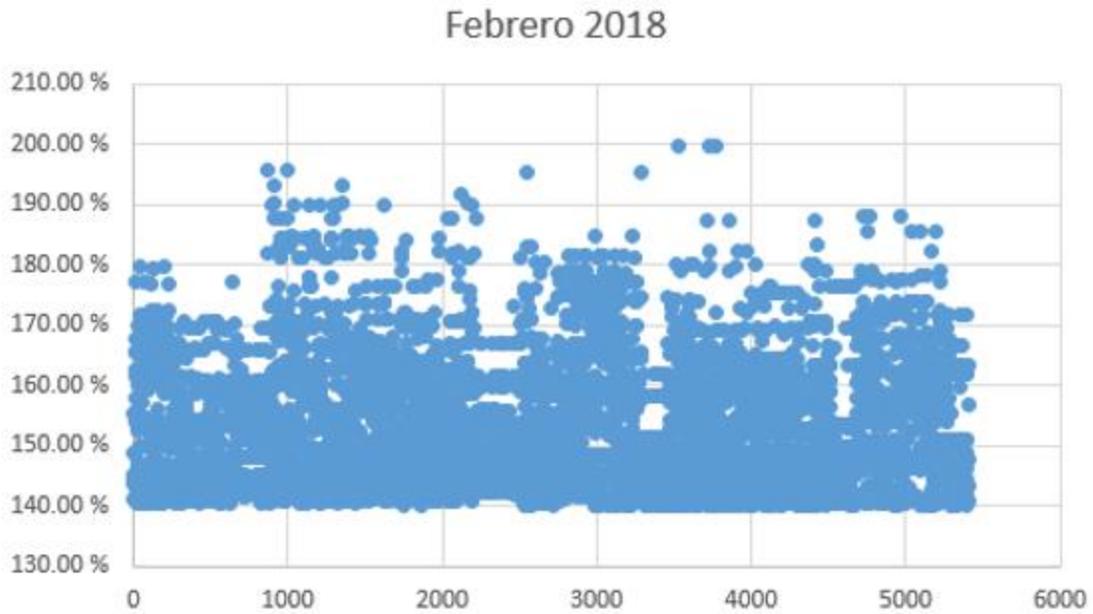


Figura 1. Eventos Registrados Eficiencia Diaria >140%

Se cuantifico la oportunidad de ahorro en el proyecto basado en el total pagado durante el 2017 incluyendo pagos especiales (decimos, vacaciones, feriados, etc.) y se determinó que excedía los **\$300,000 anuales**. Por lo que de inmediato se comenzó a monitorear los eventos, clasificar las causas y generar planes de remediación.

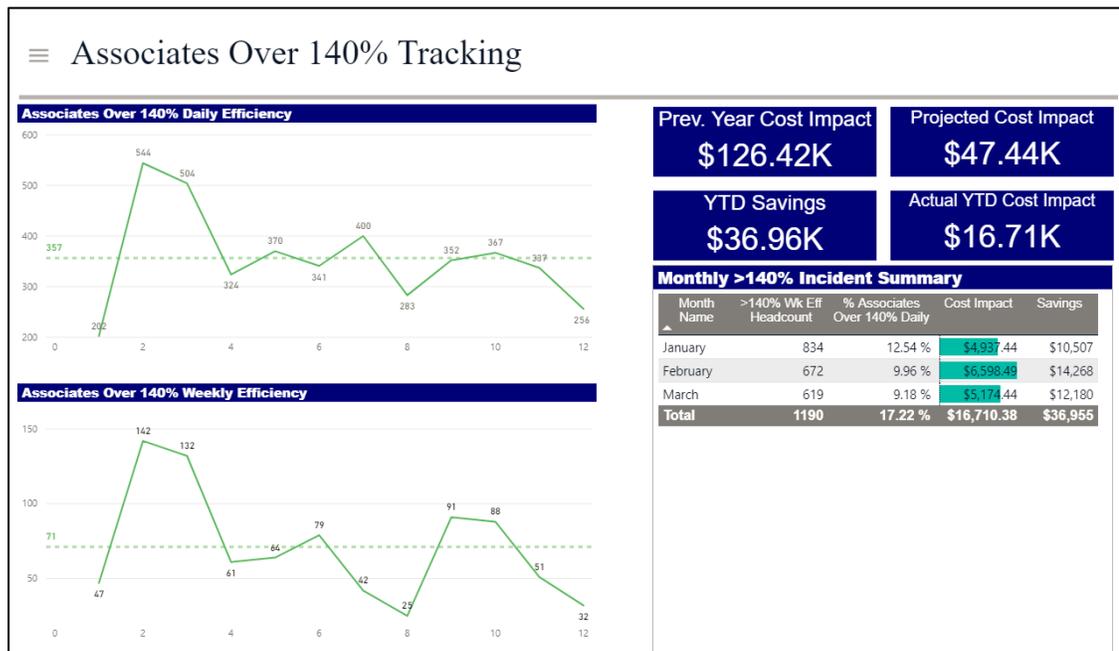


Figura 2. Dashboard Eficiencias > 140%

Gradualmente se fue controlando el indicador, estableciendo controles y reduciendo el “techo” en el sistema de planillas hasta llevarlo a 140% a mediados del 2019. Actualmente, el sistema de planilla de producción bloquea el cálculo cuando se detectan eficiencias mayores a 140% y no permite el avance hasta que se corrijan.

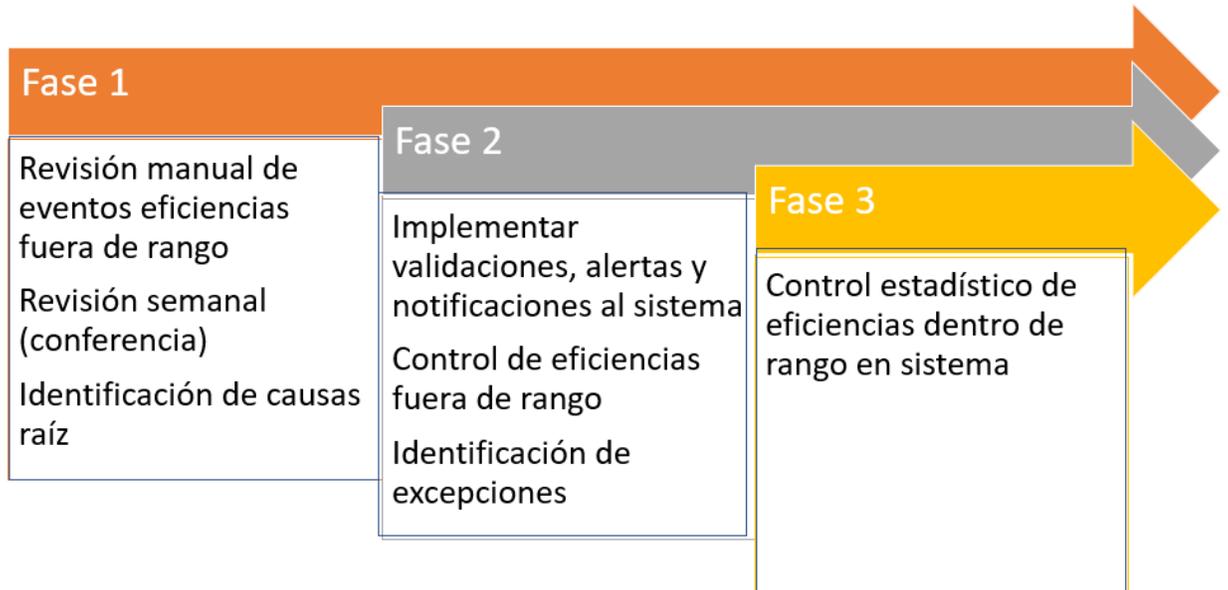


Figura 3. Fases Proyecto Control de Eficiencias

La Fase III del proyecto es lo que se pretende ejecutar en este estudio. La detección de anomalías en el rango de 0 a 140% que hasta el momento se denomina como “correcto” porque esta debajo del techo permitido, aplicando un sistema de control estadístico.

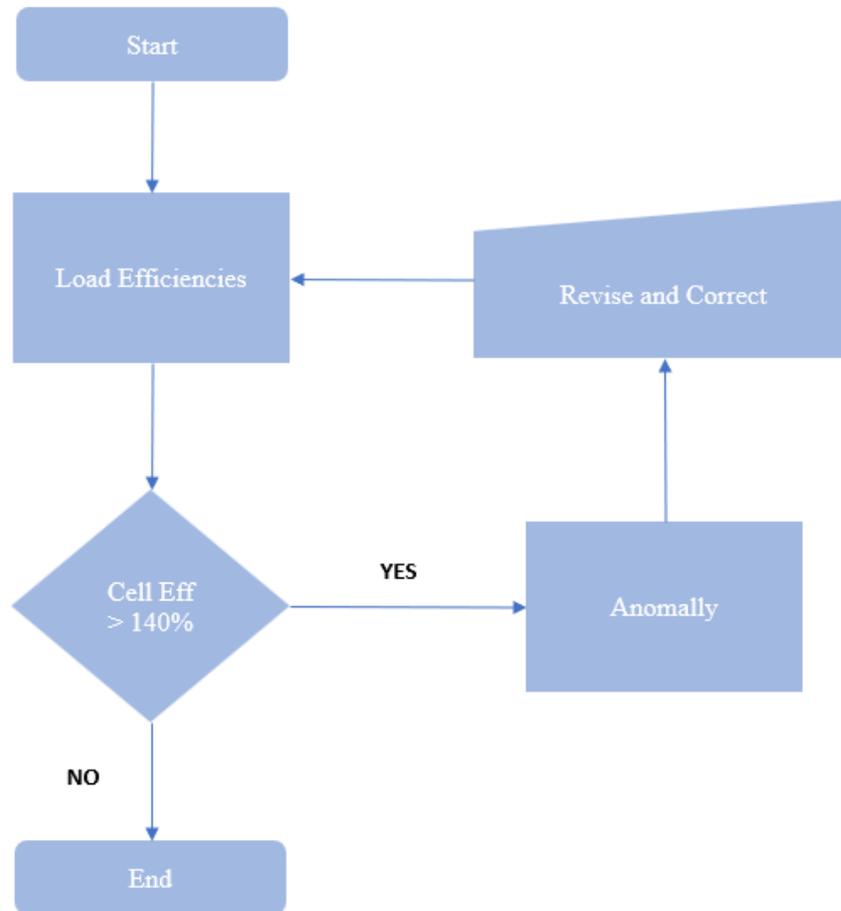


Figura 4. Detección y Corrección de Eficiencias >140%

Uno de los principales retos que se enfrentaron en la Fase I y II fue la detección y corrección oportuna de las anomalías, ya que debe ser antes del pago. Por este motivo debe ser un proceso automático y masivo, para que la corrección se ejecute en la revisión de la planilla diaria, tal como se muestra en la Fig. 4. Esta validación se integró al sistema de planillas y al detectar una eficiencia mayor a 140% bloquea el proceso y no permite seguir con la revisión ni cálculos hasta que se haya hecho la corrección.

El mismo reto se enfrenta en el estudio actual sumado a la dificultad de la detección correcta de la anomalía ya que no es tan evidente como lo era en las primeras fases. Por ello se busca el enfoque estadístico y la aplicación de modelos de aprendizaje automático que permitan la detección de forma masiva y lo más automatizado posible.

1.3 DEFINICIÓN DEL PROBLEMA

En un contexto industrial altamente competitivo, la eficiencia y precisión en la revisión de nóminas se erigen como pilares esenciales para asegurar la remuneración justa de los trabajadores y optimizar los recursos de la empresa. En este escenario, surge la necesidad de abordar una problemática clave: la detección y gestión de valores atípicos en las eficiencias diarias de la revisión de nóminas en una planta de manufactura.

El problema a resolver radica en la identificación temprana y precisa de anomalías en los datos de eficiencia de la revisión de nóminas. Estas anomalías pueden originarse por errores involuntarios en la captura de datos, omisiones en la validación de registros o incluso actividades fraudulentas. La falta de un sistema efectivo de detección de valores atípicos conlleva riesgos financieros tanto para los trabajadores, al no recibir una compensación acorde a su desempeño real, como para la empresa, al sufrir pérdidas innecesarias debido a errores o manipulaciones en los datos.

Flores & Ortiz, (2018) definen: “La detección de los valores atípicos es una tarea extremadamente importante en una amplia variedad de dominios de aplicación. Con frecuencia estos valores son eliminados para mejorar la precisión de la información, pero a veces la presencia de un valor atípico tiene un cierto sentido o explicación que se puede perder si se elimina, puesto que su identificación puede conducir al descubrimiento de un conocimiento inesperado en diversas áreas.”

La formulación específica del problema se resume en la siguiente interrogante: ¿Cómo diseñar un sistema de detección de valores atípicos en las eficiencias diarias de la revisión de nóminas en una planta de manufactura para asegurar la precisión en la remuneración de los trabajadores y generar ahorros tangibles para la empresa?

Este problema presenta una serie de limitaciones y desafíos. En primer lugar, la detección de valores atípicos en el ámbito de la revisión de nóminas se enfrenta a la complejidad inherente de los datos industriales, donde la variabilidad es alta y los patrones anómalos pueden no ser evidentes. Además, la falta de un marco de referencia previo en la detección de valores atípicos específicamente en este contexto limita la aplicación de enfoques convencionales.

Para abordar este problema, se plantean las siguientes preguntas de investigación:

¿Cómo diseñar un sistema de detección de valores atípicos en las eficiencias diarias de la revisión de nóminas que garantice la precisión en la remuneración de los trabajadores y genere ahorros para la empresa?

¿Cuáles son las metodologías y técnicas más apropiadas para detectar valores atípicos en los datos de eficiencia de la revisión de nóminas?

¿Cómo puede diseñarse un sistema de detección de valores atípicos que sea capaz de adaptarse a las particularidades de los procesos y variables involucrados en la revisión de nóminas?

¿Cuál es el impacto potencial de la implementación de un sistema de detección de valores atípicos en términos de ahorros para la empresa?

La exploración y resolución de estas preguntas de investigación permitirá abordar de manera integral la problemática de detección de valores atípicos en la revisión de nóminas, generando “insights” valiosos para la mejora continua de los procesos y la toma de decisiones en el entorno manufacturero.

1.4 OBJETIVOS DEL PROYECTO

1.4.1 OBJETIVO GENERAL

Diseñar un sistema de detección de valores atípicos en las eficiencias diarias de la revisión de nóminas en una planta de manufactura, empleando herramientas de machine learning, con el propósito de garantizar la precisión en la remuneración de los trabajadores y generar ahorros tangibles para la empresa.

1.4.2 OBJETIVOS ESPECIFICOS

- Evaluar y seleccionar las metodologías y técnicas más adecuadas para la detección de valores atípicos en los datos de eficiencia de la revisión de nóminas empleando técnicas de aprendizaje automático.
- Diseñar un sistema de detección de valores atípicos que se adapte de manera precisa a las particularidades de los procesos y variables involucradas en la revisión de nóminas.
- Evaluar el impacto potencial de la implementación del sistema de detección en términos de la generación de ahorros para la empresa.

1.5 JUSTIFICACIÓN

En el sector maquilador y por la naturaleza de su operación, una planta de costura-empaque como la que se está analizando en este estudio, el principal gasto lo compone la mano de obra directa. Por este motivo, la revisión precisa de nóminas es esencial para la compensación justa y la medición eficaz de la eficiencia en la manufactura. La detección de valores atípicos en estas eficiencias se convierte en un reto crítico. Esta investigación aborda este desafío con la implementación de un sistema de detección de valores atípicos.

La importancia radica en su potencial para optimizar costos y recursos, mejorar la toma de decisiones y crear un entorno laboral más justo y motivador. Desde una perspectiva económica, la precisión en la remuneración y la reducción de pérdidas financieras son beneficios tangibles. Operativamente, la información confiable respalda la mejora continua y la equidad en la compensación fomenta la productividad.

Desde un punto de vista social, la equidad en la revisión de nóminas y la detección de fraudes refuerzan una cultura de confianza y compromiso. En resumen, la investigación no solo resuelve un desafío, sino que también conlleva mejoras económicas, operativas y sociales. La justificación reside en la transformación de problemas en oportunidades para el crecimiento y el éxito sostenible de la empresa.

CAPÍTULO II. MARCO TEÓRICO

2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL.

En el contexto actual de la planta de producción, se ha implementado un enfoque de producción modular o en celda, donde los empleados operan como equipos interdependientes. La retribución de los empleados se basa en un sistema que valora directamente su eficiencia en el trabajo. La eficiencia se calcula mediante la división de los tiempos estándar (SAMs) necesarios para completar las tareas entre el tiempo real laborado (horas hombre). Esta relación resultante determina la eficiencia del equipo y, a su vez, dictamina la compensación por hora que se otorga.

En el sistema, el supervisor de la unidad tiene la responsabilidad de llevar a cabo una revisión exhaustiva de los indicadores clave de producción a diario. Esto incluye la producción total, los tiempos estándar, las horas trabajadas por el equipo y, lo más crucial, la eficiencia obtenida.

Sin embargo, se han identificado desafíos importantes en este proceso. Dada la naturaleza variada de las operaciones y la producción, la eficiencia puede elevarse de manera incorrecta en algunas instancias llegando al punto de brindarse como una posible anomalía. Esto puede deberse a una variedad de factores, como registros de producción erróneos, inexactitudes en las horas laboradas o incluso fluctuaciones temporales en los tiempos estándar debido a cambios en los procesos. Lo que es más preocupante es que, en muchas ocasiones, estas discrepancias pasan desapercibidas debido a la revisión manual llevada a cabo por el supervisor.

Alcalde (2018) afirma que una anomalía es un dato muy distinto del resto. Esto puede deberse a fallos en mediciones, o a la propia naturaleza del dato. Por ejemplo, una intrusión a un sistema informático puede considerarse una anomalía, ya que por norma general el resto de las actividades en dicho sistema serán legítimas.

Este escenario plantea un desafío significativo en cuanto a la precisión de los pagos de nómina. Las eficiencias infladas o incorrectamente calculadas pueden llevar a una asignación inadecuada de compensación por hora, lo que afecta tanto a los empleados como a la empresa en su conjunto. Además, la detección manual de estas anomalías se ha vuelto cada vez más compleja a medida que la producción ha aumentado en escala.

En resumen, el enfoque actual en la planta de producción involucra una medición crítica

de la eficiencia de los equipos basada en los tiempos estándar y el tiempo laborado. Aunque este sistema tiene el potencial de ser altamente efectivo, la revisión manual de los datos de producción y eficiencia ha demostrado ser susceptible a errores y desafíos en la detección de anomalías. En las siguientes secciones, exploraremos cómo la aplicación de análisis de series temporales puede mitigar estos problemas y mejorar la precisión en la evaluación de eficiencias en la nómina.

2.2 CONCEPTUALIZACIÓN

Celda: Es el grupo de asociados que trabajan en un mismo equipo. El pago de cada uno depende del trabajo de todos.

Sistema de producción Modular o en Celda: Es el método a través del cual el flujo en el proceso de producción corre de pieza en pieza dentro de cada celda con el fin de obtener una eficiencia común.

Asociado Directo: Son asociados que contribuyen directamente al proceso productivo de la prenda, ya sea de costura, empaque o corte, estos son: operarias(os), comodines, limpieza de prenda y limpieza de planta, ayudantes de producción y manejo de materiales, cuya función agrega valor directo al producto.

Comodines: Son considerados aquellos Asociados Directos que conocen más de una operación. Las(os) Comodines se encuentran asignadas(os) a la celda # 1 de cada unidad. El pago de estas horas trabajadas dentro de la celda dependerá de la eficiencia de esta.

Standard Allowed Minutes (SAM): Los Standard Allowed Minutes o SAMs son los tiempos que se ha definido para el cálculo de las horas producidas de cada asociado o celda. Estos minutos están definidos por estilo, talla y operación, de esta forma serán procesados en el sistema y se deberá asociar para cada asociado su operación respectiva.

Horas en Estándar (Horas ON): Es la cantidad de horas que los asociados han permanecido en la Planta, donde no se ha experimentado ningún paro de producción.

Horas Fuera de Estándar (Horas OFF): Son horas durante las cuales la producción se ve interrumpida por diferentes causas.

Protección por Reentrenamiento: Es considerado el estatus de protección para celdas y/o asociados directos para salvaguardar el pago durante un periodo de tiempo determinado debido a cambio de operación. El salario del asociado es protegido al promedio antes del cambio siempre y

cuando cumpla con su curva de crecimiento.

Aprendizaje automático: Técnicas que le dan a las computadoras la capacidad de mejorar en una tarea sin ser programadas explícitamente (y el campo de estudio que cubre esas técnicas). Para muchas aplicaciones, el proceso está guiado por datos de entrenamiento, por lo que decimos que la máquina "aprende de los datos" para mejorar en la tarea. Específicamente, genera un modelo predictivo a partir de los datos; el modelo es lo que se "aprende". Para las aplicaciones empresariales de aprendizaje automático, los datos a partir de los cuales aprende generalmente consisten en una lista de casos previos o conocidos (es decir, datos etiquetados). Esta lista de casos equivale a una codificación de "experiencia", por lo que la computadora "aprende de la experiencia".

Aprendizaje automático supervisado: Aprendizaje automático que está guiado por datos etiquetados. Las etiquetas guían o "supervisan" el proceso de aprendizaje y también sirven como base para evaluar un modelo predictivo. El aprendizaje automático supervisado es la forma más común de aprendizaje automático y es el enfoque de toda esta serie de tres cursos, por lo que generalmente nos referiremos a él simplemente como aprendizaje automático.

Aprendizaje automático NO supervisado: En el algoritmo de Aprendizaje No Supervisado, no tenemos ninguna variable objetivo o de resultado para predecir/estimar. El objetivo del aprendizaje no supervisado es modelar la estructura subyacente o la distribución en los datos para aprender más sobre ellos o segmentarlos en diferentes grupos basados en sus atributos. Ejemplos de algoritmos de Aprendizaje No Supervisado: el algoritmo Apriori, K-means.

R Lang: R es un lenguaje de programación de código abierto y un entorno de software para estadísticas, aprendizaje automático y visualización de datos.

Características de R:

Es independiente de la plataforma, por lo que es compatible con múltiples sistemas operativos. R cuenta con un soporte en línea muy fuerte y consistente por parte de la comunidad. Las capacidades gráficas de R son impresionantes. Hay una gran cantidad de literatura para aprender R. R es un lenguaje de programación y un entorno de software ampliamente utilizado en estadísticas y análisis de datos. Destaca por su capacidad para realizar análisis estadísticos avanzados, visualización de datos y manipulación de información de manera eficiente. Es una

herramienta esencial para científicos de datos, estadísticos y profesionales que trabajan con datos, ya que proporciona una amplia gama de paquetes y funciones especializadas para abordar una variedad de tareas analíticas.

2.3 TEORÍAS DE SUSTENTO

2.3.1 BASES TEÓRICAS

Para abordar el problema en estudio se debe analizar y comprender las tendencias y patrones de los datos de eficiencia a lo largo del tiempo. Para ello se debe utilizar la **Teoría de Series Temporales** que implica el estudio y análisis de datos a lo largo del tiempo. Dentro de esta teoría existen métodos específicos como ARIMA (AutoRegressive Integrated Moving Average) o modelos más avanzados como SARIMA (Seasonal ARIMA), para comprender y predecir patrones en tus datos.

ALGOTIVE (2022) indica que Machine Learning (ML) o aprendizaje automático es una rama de la IA y la inteligencia artificial autónoma que permite que las máquinas aprendan a partir de experiencias con grandes cantidades de datos, sin tener que ser programadas para ello. Sintetizan e interpretan la información para el entendimiento humano, de acuerdo a parámetros preestablecidos, ayudando ahorrar tiempo, disminuir errores, crear acciones preventivas y automatizar procesos en grandes operaciones y empresas.

El Machine Learning (ML) o aprendizaje automático es una destacada rama de la Inteligencia Artificial (IA) que desempeña un papel crucial en la automatización de procesos y la mejora de la toma de decisiones en una variedad de campos. Esta disciplina permite que las máquinas adquieran la capacidad de aprender y tomar decisiones a partir de experiencias previas, sin necesidad de una programación explícita para cada tarea. En lugar de depender de algoritmos e instrucciones precisas, el ML se basa en la exposición a grandes volúmenes de datos y el procesamiento de estos para identificar patrones y tendencias relevantes. Al sintetizar y interpretar esta información de manera eficiente, el ML tiene como objetivo facilitar la comprensión humana y optimizar procesos, todo ello en función de parámetros preestablecidos. Esta característica del ML resulta invaluable en el ahorro de tiempo y recursos, al mismo tiempo que contribuye a la reducción significativa de errores en tareas complejas. A través de la automatización y el análisis de datos, el Machine Learning se convierte en una herramienta esencial en la mejora de la eficiencia operativa y la toma de decisiones informadas en diversos sectores, incluyendo la gestión

de nóminas en la industria de manufactura, como se explorará en profundidad en el presente estudio.

Ulloa et al. (2022) afirma que una ARIMA una serie de tiempo estacionaria es aquella cuyas propiedades estadísticas se distinguen por ser constantes con respecto al tiempo.

Una serie es estacionaria cuando llega a ser estable a lo largo del tiempo, cuando su media y varianza son constantes en el tiempo. Este modelo ARIMA es aplicable en el análisis a desarrollar, ya que, las medias de productividad pueden llegar a ser constantes. Es por ello, que se busca detectar posibles anomalías cuando exista un comportamiento atípico.

José Alberto Mauricio (2007) menciona que una serie temporal (o simplemente una serie) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos.

Una serie temporal, a menudo denominada simplemente "serie", constituye un elemento fundamental en la investigación y el análisis de datos en diversos campos. Se define como una secuencia ordenada y equidistante cronológicamente de N observaciones o datos. Estas observaciones se registran a lo largo del tiempo y se centran en una característica específica, lo que se conoce como serie univariante o escalar, o bien, pueden abordar múltiples características, en cuyo caso se clasifican como series multivariantes o vectoriales. La unidad observable en estudio puede ser una variable única o una colección de variables interrelacionadas, y estas observaciones se recopilan en momentos diferentes a lo largo de un período determinado. En esencia, una serie temporal se convierte en una ventana hacia el cambio y la evolución de las características o variables a lo largo del tiempo, lo que permite el análisis de tendencias, patrones cíclicos, fluctuaciones estacionales, y la detección de eventos anómalos. Este concepto y su aplicación son esenciales para comprender la dinámica temporal de una amplia gama de fenómenos en la investigación científica y empresarial, y desempeñan un papel fundamental en el análisis de series temporales y pronóstico de anomalías, como se explorará en detalle en esta tesis.

Chinlli (2021) indica que los modelos SARIMA se basan en la aplicación de los modelos ARIMA a una serie temporal transformada donde el comportamiento estacional y no estacional han sido eliminados.

En el campo del análisis de series temporales, los modelos SARIMA (Seasonal Autoregressive Integrated Moving Average) desempeñan un papel fundamental. Como señala Chinlli (2021), estos modelos se basan en la aplicación de los modelos ARIMA (Autoregressive Integrated Moving Average) a una serie temporal previamente transformada, de modo que se eliminan tanto el comportamiento estacional como el no estacional. Esto permite un análisis más preciso y eficiente de los datos temporales, lo que resulta esencial en la identificación de patrones y tendencias.

El libro "Time Series Analysis and Its Applications: With R Examples," de Robert H. Shumway y David S. Stoffer, se presenta como una herramienta integral para el análisis de series temporales. Combina la teoría con aplicaciones prácticas, utilizando R como una herramienta clave. Desde conceptos fundamentales hasta aplicaciones prácticas, este libro destaca la utilidad de R en el análisis de datos temporales.

Shumway (2017) afirma: "Una introducción a la regresión con series temporales, que implica combinar el análisis de series temporales con el análisis de regresión."

Además, como afirma Shumway (2017), el enfoque en la regresión con series temporales implica la combinación del análisis de series temporales con el análisis de regresión. Esta integración es valiosa para comprender las relaciones entre variables en un contexto temporal, lo que puede proporcionar información crítica en áreas como el pronóstico y la toma de decisiones.

El libro "Forecasting: Principles and Practice," de Rob J Hyndman y George Athanasopoulos, aborda de manera completa y accesible los fundamentos y aplicaciones del pronóstico de datos. Se enfoca en principios clave, técnicas de pronóstico y su implementación práctica utilizando el software R. A través de ejemplos y estudios de casos, ofrece una visión integral de cómo el pronóstico de datos es relevante y aplicable en diversos campos.

El libro "Forecasting: Principles and Practice," de Rob J Hyndman y George Athanasopoulos,

ofrece una visión completa y accesible del pronóstico de datos. Se centra en los principios fundamentales y las técnicas de pronóstico, utilizando R como plataforma para su implementación práctica. A través de ejemplos y estudios de casos, el libro ilustra cómo el pronóstico de datos es relevante y aplicable en diversos campos, destacando la importancia de la precisión y la planificación futura en la toma de decisiones empresariales y estratégicas.

(Hyndman & Athanasopoulos, 2014) afirma: “Calcular las predicciones utilizando el Suavizado Exponencial Simple implica tomar un promedio ponderado de los valores anteriores de la serie temporal, donde los pesos se determinan de manera exponencial. Los valores más recientes tienen un peso mayor en la predicción, lo que refleja la creencia de que los datos recientes son más informativos para el pronóstico.”

Posteriormente, como complemento y finalidad del proyecto, debemos estudiar **Teorías de Detección de Anomalías** que explora métodos estadísticos y algoritmos de aprendizaje automático utilizados para identificar puntos de datos inusuales en un conjunto de datos.

Avila et al. (2019) mencionan que la detección de anomalías es una técnica de Minería de Datos que permite el reconocimiento de nuevos patrones con comportamiento inusual, los cuales pueden ser traducidos como acciones no válidas o anómalas sobre los datos. La detección de anomalías ha permitido la identificación y prevención de actividades maliciosas como fraude e intrusiones, entre otros. El uso de técnicas tradicionales para la detección de anomalías ha reportado muy buenos resultados.

La detección de anomalías emerge como una técnica esencial en el campo de la Minería de Datos, con el propósito de identificar patrones inusuales o comportamientos atípicos dentro de un conjunto de datos. Estos patrones inusuales pueden interpretarse como acciones no válidas o anómalas en el contexto de los datos analizados. La detección de anomalías se ha establecido como una herramienta valiosa en la identificación y prevención de una variedad de actividades maliciosas, tales como fraudes y intrusiones, entre otros eventos indeseados.

Esta técnica, que se apoya en la exploración y análisis de datos, es particularmente útil en situaciones donde la detección temprana de comportamientos atípicos es crucial para la toma de decisiones informadas y la mitigación de riesgos. Tradicionalmente, las técnicas empleadas en la detección de anomalías han demostrado ser eficaces en la identificación de eventos inusuales, reportando resultados notables en la protección de sistemas y la seguridad de datos.

El uso de enfoques tradicionales en la detección de anomalías ha sido ampliamente aceptado y probado en una variedad de dominios. Estas técnicas forman una base sólida para la aplicación de detección de anomalías en el análisis de series temporales y su capacidad para pronosticar y prevenir anomalías en contextos diversos, incluyendo la gestión de la nómina en entornos de manufactura, como se abordará en detalle en esta tesis.

Así mismo como base teórica se analiza la Industria 4.0 representa una evolución significativa en la historia de las revoluciones industriales. A diferencia de sus predecesoras, que se enfocaron en la eficiencia de la producción, la cuarta Revolución Industrial se centra en la interconexión de personas, información y procesos. Este estudio investiga cómo y por qué se desarrolló la Industria 4.0 y analiza sus implicaciones en la infraestructura de red.

Patel (2023) menciona que la industria 4.0 y las tecnologías que impulsan la industria moderna brindan soluciones orientadas a procesos hasta los análisis de datos, requiriendo de una sólida conectividad de red por cable e inalámbrica.

El Blog Ruralvía (2023) afirma que la Cuarta Revolución Industrial se define por avances tecnológicos como el Internet de las Cosas (IoT), la inteligencia artificial (IA), la analítica de datos, la robótica avanzada y la realidad virtual. Su objetivo es transformar la forma en que las empresas operan, producen y brindan servicios, a través de la automatización, el análisis de datos en tiempo real y la toma de decisiones basada en información precisa.

La industria 4.0 representa una revolución tecnológica que está transformando radicalmente la forma en que se conciben, producen y entregan bienes y servicios en todo el mundo. Este paradigma industrial se caracteriza por la convergencia de tecnologías avanzadas como la Internet de las cosas (IoT), la inteligencia artificial, la robótica, la realidad virtual y

aumentada, y la computación en la nube.

Redacción APD (2021) afirma: “La ciber industria o 4ª Revolución Industrial trae consigo múltiples beneficios a las empresas. La automatización de los procesos ayuda a minimizar los tiempos, los recursos y el esfuerzo necesarios para aumentar la productividad en una fábrica. Sin embargo, deben ser tratadas con cautela las amenazas que trae consigo esta transformación.”

A través de la interconexión y la automatización de procesos, la industria 4.0 promete aumentar la eficiencia, la calidad y la personalización de la producción, al tiempo que plantea desafíos y oportunidades en términos de ciberseguridad, privacidad de datos y cambios en la fuerza laboral. Esta tesis se adentra en el análisis y la comprensión de los impactos multidimensionales de la industria 4.0 en la economía global y las empresas, explorando su influencia en la competitividad, la sostenibilidad y la innovación de las organizaciones en el siglo XXI.

Sanchez (2023) asegura: “En la Industria 4.0, las máquinas se vuelven más inteligentes y conectadas, permitiendo una mayor eficiencia y productividad, así como una mayor personalización de los productos para satisfacer las necesidades específicas de los consumidores. En resumen, la Industria 4.0 es el futuro de la producción, donde la innovación y la tecnología avanzada se unen para crear un mundo más eficiente, inteligente y conectado.”

La visión de la Industria 4.0, tal como lo sostiene Sánchez (2023), trae consigo una transformación radical en la fabricación y la producción industrial. En este contexto, las máquinas adquieren un nivel de inteligencia y conectividad sin precedentes, lo que da lugar a mejoras significativas en la eficiencia y productividad de las operaciones. Además, la Industria 4.0 facilita una mayor personalización de los productos, permitiendo satisfacer las necesidades específicas de los consumidores de una manera más efectiva.

La convergencia de la innovación y la tecnología avanzada en la Industria 4.0 representa el futuro de la producción. Este enfoque se caracteriza por la creación de un entorno más eficiente, inteligente y conectado, que ofrece soluciones adaptadas a las demandas del mercado actual.

En el sector maquilero, la industria 4.0 ha generado notables avances, como la introducción de sistemas de fabricación avanzada con robots colaborativos y máquinas conectadas, lo que ha permitido una producción más eficiente y flexible. Además, la implementación de sensores y análisis de datos en tiempo real ha mejorado la calidad del producto y reducido los desperdicios.

La trazabilidad y la gestión de la cadena de suministro también se han beneficiado, proporcionando mayor visibilidad y toma de decisiones informada, lo que impulsa la eficiencia y la competitividad en un mercado global cada vez más competitivo.

En el sector maquilero, la influencia de la Industria 4.0 es innegable, y ha dado lugar a notables avances. La introducción de sistemas de fabricación avanzada, que incluyen robots colaborativos y máquinas interconectadas, ha impulsado la producción hacia niveles de eficiencia y flexibilidad sin precedentes. La integración de sensores y el análisis de datos en tiempo real han llevado a mejoras sustanciales en la calidad del producto y a la reducción de desperdicios.

Además, la trazabilidad y la gestión de la cadena de suministro han experimentado mejoras significativas gracias a la Industria 4.0, lo que se traduce en una mayor visibilidad y en la capacidad de tomar decisiones informadas de manera más efectiva. Estos avances en la gestión y control operativo impulsan la eficiencia y la competitividad en un mercado global que se torna cada vez más desafiante y competitivo.

EDSRobotics (2020) define como desventaja de la Industria 4.0: “La industria 4.0 presenta desafíos significativos, como la preocupación por la pérdida de empleos debido a la automatización, riesgos de ciberseguridad debido a la alta interconexión, la complejidad y costos iniciales de implementación, obsolescencia tecnológica rápida y preocupaciones éticas y de privacidad relacionadas con la gestión de grandes cantidades de datos.”

La Industria 4.0 ofrece ventajas notables, pero conlleva desafíos importantes. Entre estos desafíos se encuentra la preocupación por la pérdida de empleos debido a la automatización, riesgos de ciberseguridad debido a la alta interconexión, costos iniciales y complejidad en la implementación, obsolescencia tecnológica rápida y preocupaciones éticas y de privacidad relacionadas con la gestión de datos masivos. Abordar estos desafíos se vuelve esencial para aprovechar al máximo las oportunidades que brinda la Industria 4.0.

2.3.2 METODOLOGÍAS DESARROLLADAS

Lee et al. (2023) afirma: “Los enfoques de detección de anomalías desarrollados durante las últimas décadas se pueden clasificar en dos categorías: enfoques de aprendizaje estadístico y enfoques de aprendizaje automático. La primera categoría de enfoques ajusta un modelo estadístico a un conjunto de datos normales dado y luego utiliza el modelo para determinar si un

nuevo punto de datos se ajusta o no a este modelo. Si el punto de datos tiene una baja probabilidad de haber sido generado por el modelo, se considera que el punto de datos representa una anomalía."

Según la afirmación de Lee et al. (2023), el campo de la detección de anomalías ha evolucionado a lo largo de las últimas décadas, y los enfoques utilizados pueden dividirse en dos categorías fundamentales: enfoques de aprendizaje estadístico y enfoques de aprendizaje automático. Estas dos categorías desempeñan un papel esencial en la identificación de comportamientos inusuales dentro de conjuntos de datos.

La primera categoría, los enfoques de aprendizaje estadístico, se caracteriza por el ajuste de un modelo estadístico a un conjunto de datos normales predefinidos. Una vez que se ha establecido el modelo, se utiliza para evaluar si un nuevo punto de datos se ajusta a este modelo o si se desvía significativamente de él. Si el punto de datos en cuestión muestra una baja probabilidad de haber sido generado por el modelo estadístico, se clasifica como una anomalía. Este enfoque se basa en la premisa de que los datos normales siguen un patrón estadístico particular, y cualquier desviación significativa de este patrón puede indicar la presencia de una anomalía.

Lee et al. (2023) utiliza el algoritmo RePAD (Real-time Proactive Anomaly Detection for Time Series) para la detección de anomalías en tiempo real. Su énfasis en el estudio fue comprobar la eficacia del algoritmo comparado con otros utilizados anteriormente.

Según la investigación de Lee et al. (2023), se ha llevado a cabo un estudio centrado en la aplicación del algoritmo denominado RePAD (Real-time Proactive Anomaly Detection for Time Series) para la detección de anomalías en tiempo real en contextos específicos. El objetivo fundamental de este estudio fue evaluar la eficacia de dicho algoritmo en comparación con otros métodos previamente utilizados en la detección de anomalías en series temporales.

El algoritmo RePAD se distingue por su capacidad para detectar anomalías en tiempo real, lo que lo convierte en una herramienta de gran utilidad en situaciones en las que la detección inmediata de comportamientos inusuales es crucial. El énfasis de la investigación se centra en someter a prueba y comparar el rendimiento de RePAD con otros algoritmos previamente empleados en el mismo contexto.

La evaluación comparativa de RePAD con otros enfoques es esencial para determinar si este algoritmo representa una mejora significativa en la detección de anomalías en series

temporales. Esta investigación busca proporcionar una base sólida y respaldada empíricamente para la selección del enfoque más eficiente y efectivo en la identificación de comportamientos inusuales, lo que es fundamental en una variedad de aplicaciones, como la seguridad cibernética, el mantenimiento predictivo y la detección de anomalías en procesos industriales.

Schmidl et al. (2022) indica que los algoritmos se originan de diferentes áreas de investigación y pertenecen a diferentes familias. En su evaluación, se examinaron 158 publicaciones que describe cada una un método único para detección de anomalías en series de tiempo.

De acuerdo con la investigación realizada por Schmidl et al. (2022), se destaca que los algoritmos utilizados en la detección de anomalías en series de tiempo provienen de diversas áreas de investigación y se pueden agrupar en diferentes familias. Esta diversidad de orígenes y enfoques refleja la complejidad y la amplitud del campo de estudio.

En su evaluación exhaustiva, se examinaron un total de 158 publicaciones, cada una de las cuales describe un método único para la detección de anomalías en series de tiempo. Cada uno de estos métodos aporta enfoques y técnicas específicas para abordar la tarea de identificar comportamientos inusuales dentro de datos temporales. La variedad de enfoques permite una comprensión más profunda de las estrategias y técnicas disponibles para la detección de anomalías en este contexto.

La revisión de estas publicaciones y la evaluación de los métodos propuestos constituyen un recurso esencial para los investigadores y profesionales que trabajan en el campo de la detección de anomalías en series de tiempo. La diversidad de enfoques y la riqueza de las técnicas presentadas proporcionan una visión global de las posibilidades y desafíos en este dominio, lo que contribuye a un mayor conocimiento y avance en la detección efectiva de comportamientos anómalos en datos temporales.

Iglesias (2022) afirma que las técnicas no supervisadas se emplean normalmente en una situación en la que no se conoce ningún conocimiento previo del conjunto de datos, por lo tanto no se presenta información de la etiqueta. En este caso, existen varias técnicas alternativas como: aproximaciones gráficas, aproximaciones paramétricas, aproximaciones basadas en vecinos cercanos y aproximaciones basadas en clustering.

En su investigación, Iglesias (2022) subraya la importancia de las técnicas no supervisadas en el contexto de la detección de anomalías, particularmente en situaciones donde se carece de conocimiento previo sobre el conjunto de datos y, por lo tanto, no se dispone de información de etiquetas que indique qué observaciones son anómalas y cuáles son normales.

Las técnicas no supervisadas se destacan por su capacidad para identificar patrones y comportamientos inusuales en datos sin requerir información previa sobre las anomalías. Esto las convierte en un enfoque valioso cuando se trabaja con datos en los que las anomalías pueden ser variadas y difíciles de prever.

Iglesias señala que, en el contexto de técnicas no supervisadas, existen varias aproximaciones alternativas. Estas incluyen aproximaciones gráficas, que se basan en la visualización de datos para detectar patrones inusuales; aproximaciones paramétricas, que utilizan modelos estadísticos para caracterizar la distribución de datos; aproximaciones basadas en vecinos cercanos, que se fundamentan en la proximidad entre observaciones para identificar anomalías; y aproximaciones basadas en clustering, que agrupan datos en clústeres y buscan observaciones que se desvíen significativamente de sus clústeres.

La elección de la técnica no supervisada más adecuada dependerá de la naturaleza de los datos y los objetivos específicos del análisis. Al considerar estas alternativas, los investigadores y profesionales pueden abordar eficazmente la detección de anomalías en situaciones donde la información de etiquetas no está disponible y se requiere un enfoque más flexible y exploratorio.

Lee et al. (s.f.) afirma que los enfoques requieren la comprensión de patrones de datos y a menudo necesitan un largo período fuera de línea para construir un modelo o una red para los datos específicos. Proporcionar detección de anomalías en tiempo real y de manera proactiva para series de tiempo en streaming sin intervención humana y sin necesidad de conocimientos especializados es altamente valioso, ya que reduce en gran medida el esfuerzo humano y permite tomar medidas apropiadas antes de que ocurra un daño, una falla u otro evento perjudicial.

Como señala Lee et al. (s.f.) en su investigación, muchos de los enfoques tradicionales para la detección de anomalías requieren una comprensión profunda de los patrones de datos y, con frecuencia, un extenso período fuera de línea para construir modelos o redes específicas para los datos en cuestión. Esta necesidad de conocimientos previos y tiempo de modelado representa un desafío significativo en la detección de anomalías, ya que puede limitar la capacidad de respuesta

y la eficacia de los sistemas de detección, especialmente en entornos en constante cambio.

En respuesta a esta limitación, la propuesta de proporcionar detección de anomalías en tiempo real y de manera proactiva para series de tiempo en streaming, sin requerir intervención humana ni conocimientos especializados, emerge como un enfoque altamente valioso. Este enfoque tiene el potencial de reducir considerablemente la carga de trabajo humana y permite la adopción de medidas apropiadas antes de que ocurran daños, fallos u otros eventos perjudiciales.

La detección proactiva de anomalías en tiempo real en datos en streaming ofrece la ventaja de actuar de manera preventiva, en lugar de reactiva, lo que puede ser esencial en situaciones donde la anticipación y la acción temprana son críticas. Esta aproximación no solo reduce el riesgo de consecuencias adversas, sino que también libera recursos humanos para tareas más estratégicas y de toma de decisiones.

Schmidl et al. (2022b) afirman que detectar subsecuencias anómalas en datos de series temporales es una tarea importante en áreas que abarcan desde procesos de fabricación hasta aplicaciones financieras y monitoreo de la salud. Una anomalía puede indicar eventos importantes, como fallos en la producción, cuellos de botella en la entrega, defectos en el sistema o latidos cardíacos irregulares, y, por lo tanto, es de interés central. Debido a que las series temporales suelen ser extensas y presentan patrones complejos, los científicos de datos han desarrollado diversos algoritmos especializados para la detección automática de estos patrones anómalos. El número y la variedad de algoritmos de detección de anomalías han crecido significativamente en el pasado, y debido a que muchas de estas soluciones se han desarrollado de forma independiente y por diferentes comunidades de investigación, no existe un estudio exhaustivo que evalúe y compare sistemáticamente los diferentes enfoques. Por esta razón, elegir la mejor técnica de detección para una tarea de detección de anomalías específica es un desafío difícil.

Según lo expresado por Schmidl et al. (2022b) en su investigación, la detección de subsecuencias anómalas en datos de series temporales representa una tarea de gran relevancia en diversos campos, que abarcan desde procesos de fabricación hasta aplicaciones financieras y el monitoreo de la salud. La presencia de una anomalía en una serie temporal puede indicar eventos cruciales, como fallos en la producción, congestión en la entrega de productos, defectos en sistemas críticos o incluso irregularidades en el ritmo cardíaco. Esta importancia radica en que la detección de anomalías se centra en la identificación temprana de situaciones que requieren

atención inmediata, lo que puede prevenir consecuencias graves y costosas.

Dado que las series temporales suelen ser extensas y caracterizadas por patrones complejos, los científicos de datos han desarrollado una variedad de algoritmos especializados para llevar a cabo la detección automática de patrones anómalos. Sin embargo, el crecimiento significativo en el número y la diversidad de estos algoritmos ha sido notable en los últimos años. Esto se debe a que muchas de estas soluciones han sido desarrolladas de forma independiente por diversas comunidades de investigación, lo que ha resultado en la falta de un estudio exhaustivo que evalúe y compare de manera sistemática los diferentes enfoques disponibles.

Por lo tanto, la elección de la mejor técnica de detección para una tarea de identificación de anomalías específica se convierte en un desafío complicado. La falta de una evaluación integral y comparativa de los enfoques disponibles hace que la selección de la técnica adecuada sea una decisión crítica y a menudo difícil de tomar. La investigación de Schmidl et al. destaca esta necesidad de una mayor comprensión y comparación de los algoritmos de detección de anomalías en series temporales, con el objetivo de proporcionar a los profesionales y científicos de datos una base sólida para abordar esta tarea esencial de manera más efectiva.

2.3.3 INSTRUMENTOS UTILIZADOS

En esta sección, describiremos los instrumentos y enfoques que utilizaremos para detectar anomalías en los datos de eficiencia de la nómina, basados en las teorías de series temporales y detección de anomalías previamente discutidas.

XGBoost (eXtreme Gradient Boosting) y THIEF (Temporary HIERarquical Forecasting): Utilizaremos dos modelos para identificar patrones estacionales y tendencias en los datos de eficiencia diaria. Estos nos permitirán ajustar un modelo estadístico a los datos y predecir valores futuros basados en el historial de eficiencias.

Mendoza (2020) menciona: “La base de los modelos XGBoost es generar múltiples modelos secuenciales de predicción “sencilla”, y cada nuevo modelo toma el resultado del modelo anterior, haciendo que los resultados obtenidos en cada secuencia sean cada vez más “robustos”, y, por tanto, obteniendo en cada iteración resultados más exactos.”

Athanasopoulos et al. (n.d.) define THIEF como: “Definimos una jerarquía temporal como la conexión estructural a través de los niveles de agregación. Cada nivel destaca diferentes

características de la serie temporal, lo que resulta en pronósticos independientes que contienen información diferente.”

Método de Detección de Anomalías Basado en Desviación Estándar: Aplicaremos un enfoque simple pero efectivo utilizando desviaciones estándar para identificar eficiencias que se desvíen significativamente de la media histórica. Si una eficiencia se encuentra a más de X desviaciones estándar de la media, se marcará como una posible anomalía.

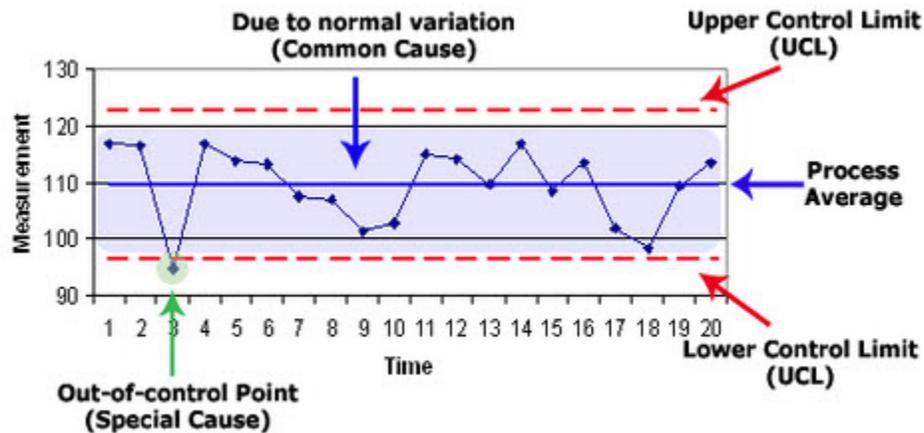


Figura 5. Utilización de Límites de Control para Detección de Anomalías

TIBCO (n.d.) hace referencia a las anomalías basadas en desviaciones estándar y mencionan: “Existen dos tipos de valores anómalos, multivariados y univariados. Los valores anómalos univariados son un punto de datos que resulta extremo para una variable. Un valor anómalo multivariado es una combinación de puntos de datos inusuales, que incluyen al menos dos puntos de datos.”

Uso de Técnicas de Visualización: Emplearemos gráficos y visualizaciones interactivas para representar las series temporales de eficiencia y resaltar visualmente los puntos de datos que podrían ser considerados anomalías. Esto nos permitirá comprender mejor la distribución de los datos y facilitará la identificación de patrones inusuales.

Validación Cruzada y Evaluación de Resultados: Utilizaremos técnicas de validación cruzada para evaluar la efectividad de los métodos de detección de anomalías. Esto nos ayudará a determinar la precisión y la confiabilidad de las detecciones realizadas y ajustar los enfoques en función de los resultados obtenidos.

Amazon Web Services (n.d.) define: “La validación cruzada es una técnica para evaluar modelos mediante el entrenamiento de varios modelos en subconjuntos de los datos de entrada disponibles y evaluarlos con el subconjunto complementario de los datos. Utilice la validación cruzada para detectar el sobreajuste, es decir, en aquellos casos en los que no se logre generalizar un patrón.”

2.4 MARCO LEGAL

Haciendo una revisión de los manuales y procedimientos de planillas, producción y recursos humanos, no se encuentra un apartado que describa lo que se debe considerar una anomalía o un detalle los tipos de anomalías que se presentan en las planillas.

En su lugar, lo que detallan los procedimientos son las practicas que debe ejecutar el personal involucrado para el correcto procesamiento de las planillas. Las anomalías se dan por la omisión de estas instrucciones.

Aunque no se dispone de parámetros preestablecidos para definir una anomalía, se manejan de forma empírica, escenarios que se pueden catalogar como errores que se deben corregir en base a observación histórica de la eficiencia de una celda o asociados individuales. Por ello, el enfoque de la estrategia de detección se basa en control estadístico del comportamiento de la eficiencia.

RAE (n.d.) indica: “Se conoce anomalía como desviación o discrepancia de una regla o de un uso. Defecto de forma o de funcionamiento.”

Entre los escenarios que se pueden reconocer, está el incremento de eficiencia inusual que se da cuando se marca un fuera de estándar de forma incorrecta. Como se observa en la Figura 6, cuando no se marca fuera de estándar, la eficiencia de la celda sufre una caída. Esto es indicativo que la eficiencia operador con la que se le ha estado pagando no representa la capacidad real del equipo y por ende se catalogaría como un error.

El comportamiento esperado se observa en la Figura 7, donde claramente se puede apreciar que la eficiencia operador si representa la capacidad del empleado o celda ya que al no existir un paro (fuera de estándar), la eficiencia planta incrementa y se iguala a la eficiencia operador.



Figura 6. Anomalía Detectada. Eficiencia Operador no representa la capacidad real del asociado.



Figura 7. Comportamiento esperado de la relación Eficiencia Operador y Planta

Otros casos menos evidentes y por ello, más difíciles de detectar, son los casos de registro de producción o tiempo laborado incorrecto.

Al analizar las incidencias de detección de anomalías, se clasificarán y se registrarán para documentación interna en los procedimientos. De igual forma, se propondrán mejoras al algoritmo para poder capturar de forma eficiente todas las incidencias posibles.

CAPÍTULO III. METODOLOGÍA

3.1 CONGRUENCIA METODOLÓGICA

Esta tesis se centra en establecer un sólido camino metodológico para el análisis de series temporales con el propósito de detectar anomalías en la nómina diaria de una planta de manufactura. La elección de la metodología está en línea con los objetivos definidos, abordando de manera completa la tarea de identificar patrones inusuales en los registros de nómina, que podrían señalar errores, fraudes o ineficiencias en la gestión de los recursos humanos y las operaciones.

Tabla 1. Proceso Metodológico

Etapas	Actividades	Herramientas/Métodos
Recopilación y Preprocesamiento de Datos	- Obtención y recopilación de los registros diarios de nómina.	Base de datos de la planta de manufactura.
	- Validación de la calidad y confiabilidad de los datos.	Métodos de validación de datos.
	- Aplicación de técnicas de preprocesamiento para garantizar la integridad de los datos.	Limpieza de datos, normalización.
Exploración y Visualización de Series Temporales	- Visualización y análisis inicial de las series temporales de nómina.	Gráficos de series temporales, herramientas de visualización.
	- Identificación de patrones estacionales, tendencias y posibles anomalías visuales.	Métodos de análisis visual.
Selección y Ajuste de Modelos	- Evaluación de técnicas de análisis de series temporales para detección de anomalías.	Métodos estadísticos, aprendizaje automático.
	- Selección de modelos más adecuados para la detección de anomalías.	Evaluación comparativa de modelos.
	- Ajuste de parámetros de los modelos.	Técnicas de ajuste de parámetros.
Implementación de Técnicas de Detección de Anomalías	- Implementación de las técnicas de detección de anomalías seleccionadas.	Programación, herramientas de análisis de datos.
	- Aplicación de técnicas a las series temporales de nómina.	Programación, procesamiento de datos.
	- Identificación de valores atípicos y patrones inusuales.	Métodos de detección de anomalías.
Evaluación y Validación de Resultados	- Evaluación cuantitativa y cualitativa de los resultados obtenidos.	Métricas de desempeño, análisis cualitativo.
	- Comparación de resultados con situaciones conocidas.	Análisis comparativo.
	- Análisis de las implicaciones para la gestión de la planta de manufactura.	Interpretación de resultados.

3.1.1 MATRIZ METODOLÓGICA

Tabla 2. Matriz Metodológica

Preguntas de Investigación	Objetivos	Hipótesis	Metodologías	Instrumentos e Indicadores	Variables	
					Independientes	Dependientes
¿Cómo se aplican técnicas de análisis de series temporales para detectar patrones anómalos en los datos de nómina diaria de una planta de manufactura, con el fin de optimizar la gestión de recursos humanos y minimizar errores en el proceso de pago?	Implementar técnicas de análisis de series temporales en la nómina diaria de una planta de manufactura para identificar patrones anómalos y mejorar la gestión de recursos humanos, reduciendo errores en el proceso de pago.	N/A considerando el alcance descriptivo del trabajo	Investigación no experimental	Instrumentos de diagnósticos y de análisis	Fechas	Eficiencia
¿Cuáles son los métodos y herramientas óptimos para detectar anomalías en tiempo real en los registros de nómina, en el contexto de una planta de manufactura, y cómo pueden ser implementados eficazmente para mejorar la toma de decisiones y la eficiencia operativa?	Evaluar métodos y herramientas para detectar anomalías en tiempo real en la nómina de una planta de manufactura y diseñar su implementación para mejorar la eficiencia operativa y la toma de decisiones	N/A considerando el alcance descriptivo del trabajo	Análisis estadístico de data	Instrumentos comparativos de monitoreo	Implementación de Series de Tiempo para detección de anomalías	Precisión de las Predicciones

3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO

El conjunto de datos proporcionados y sometidos a análisis por parte de la planta de manufactura consta aproximadamente de 292,000 registros, cada uno asociado a la eficiencia obtenida por un empleado en una ubicación – fecha específica.

Se dispone de información sobre las planillas del 2022 de toda la planta que consta de alrededor de 1600 empleados. Cada registro detalla el periodo de planilla, fecha, ubicación (unidad-celda), código de empleado, tipo de empleado, tiempo ON, tiempo extra, tiempo Off, SAMs y finalmente eficiencia.

3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES

En el contexto de este estudio, se presenta un enfoque basado en análisis de series temporales para la detección de anomalías en la nómina diaria de una planta de manufactura. Para llevar a cabo dicho análisis se extrajo de la base de datos de planillas de producción los registros de las eficiencias diarias por empleado.

La data resultante contiene 3 variables, Fecha, id Empleado y la eficiencia operador que se operacionalizan en la Fig. 8.

Variable	Tipo de Variable	Operacionalizacion	Dimensiones	Definicion	Indicador	Nivel de Medicion	Unidad de Medida
Fecha	Catógica	Registro de la fecha	Temporal	Fecha en que se realiza la medición	-	Nominal	-
idEmpleado	Catógica	Identificación del empleado	Identificación	Código único que identifica a cada empleado	-	Nominal	-
Eficiencia Operador	Númerica	Medida de eficiencia	Unidimensional	Nivel de eficiencia operativa del operador	Porcentaje (%)	Ratio	Porcentaje (%)

Figura 8. Operacionalización de Variables

3.2 ENFOQUE Y MÉTODOS

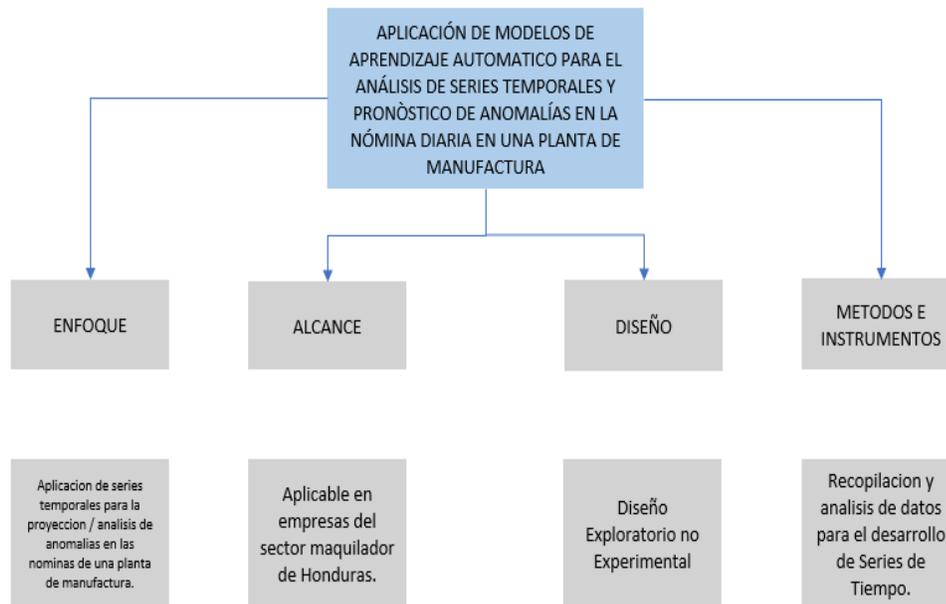


Figura 9. Enfoque y Métodos

3.3 DISEÑO DE LA INVESTIGACIÓN

En el presente capítulo, se detalla el diseño metodológico que guiará la investigación titulada "Análisis de Series Temporales para la Detección de Anomalías en la Nómina Diaria en una Planta de Manufactura". Este diseño está estructurado de manera coherente y sistemática, constituye un diseño no experimental con el propósito de lograr los objetivos planteados y responder a la pregunta de investigación.

3.3.1 POBLACIÓN

Actualmente la Corporación Fruit of the Loom HN está conformada por 3 plantas de costura – empaque, situadas en El Progreso, Yoro, Villanueva y Choloma, Cortes. La población operativa o directa de las 3 plantas oscila entre 1,500 y 1,600 empleados por planta.

Para efectos de la investigación planteada se tomará la población operativa de la planta El Provenir Manufacturing – Tela de la corporación situada en Zip El Porvenir en El Progreso, Yoro. Debido a la naturaleza del análisis, se debe incluir únicamente información del personal operativo de costura y empaque que devenga por producción y que consta de alrededor de 1500 empleados.

3.3.2 MUESTRA

Se considera una muestra no probabilística. El análisis de series temporales se hará sobre toda la población activa a la fecha y considerando la data de planilla de producción de la planta en mención del año 2023.

Para elaborar pruebas, se segmentará la data por periodos de planilla para tomar históricos, hacer predicciones y luego comparar contra la data real y correr la detección de anomalías.

Esta muestra es representativa respecto al comportamiento de las demás plantas ya que a nivel operativo, la revisión de planillas y los aspectos que influyen en la incidencia de anomalías es igual para todas las plantas.

3.3.3 TECNICAS DE MUESTREO

El análisis de series se hará sobre toda la población activa, considerando las eficiencias registradas en el 2023 hasta la última planilla cerrada. Se filtrarán los datos de la última planilla para efectos de prueba y se entrenara el modelo con el restante de semanas cerradas.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

En el marco de la investigación sobre "Análisis de Series Temporales para la Detección de Anomalías en la Nómina Diaria en una Planta de Manufactura", se implementarán diversas técnicas, utilizarán instrumentos específicos y seguirán procedimientos meticulosos para alcanzar los objetivos establecidos. A continuación, se describen las principales técnicas, instrumentos y procedimientos que se aplicarán:

Técnicas:

- **Análisis de Series Temporales:** Se aplicará un análisis profundo de las series temporales de la nómina diaria para identificar patrones estacionales, tendencias y comportamientos anómalos.
- **Modelos de Pronóstico:** Se emplearán dos modelos de pronóstico; XGBoost (eXtreme Gradient Boosting) y THIEF (Temporary HIERarquical Forecasting). Se correrán dentro del paquete "Modeltime" que permite escalabilidad para realizar análisis de series temporales en volúmenes altos de series.
- **Detección de Anomalías:** Se utilizarán técnicas estadísticas y de aprendizaje automático para detectar valores atípicos y patrones inusuales en las series temporales, indicando posibles anomalías.

Instrumentos:

- **Software de Análisis:** Se utilizará SQL Server para la extracción de datos de la BD de planillas y posteriormente se utilizará R y RStudio para análisis de datos y la creación de modelos.
- **Herramientas de Visualización:** Librería de visualización ggplot2 permitirá crear gráficos y visualizar patrones en las series temporales.

Procedimientos:

- **Recopilación y Preprocesamiento de Datos:** Los registros diarios de nómina se recopilarán de la base de datos de la planta de manufactura. Se realizará una revisión minuciosa para garantizar su calidad y confiabilidad. Se llevará a cabo un proceso de limpieza y transformación de datos para asegurar la coherencia y

completitud.

- Exploración y Visualización Inicial: Se realizará un análisis exploratorio inicial de las series temporales mediante gráficos de líneas y diagramas de caja para identificar patrones y posibles anomalías visuales.
- Modelado de Series Temporales: Se ajustarán modelos de pronóstico XGBoost (eXtreme Gradient Boosting) y THIEF (Temporary HIERarquical Forecasting) para prever tendencias y patrones futuros.
- Detección de Anomalías: Se aplicarán técnicas de detección de anomalías, como el método de desviación estándar o el uso de modelos de detección basados en aprendizaje automático, para identificar valores inusuales en las series temporales.
- Evaluación de Resultados: Se compararán los patrones detectados con situaciones conocidas para validar la efectividad de las técnicas aplicadas. Se analizarán las implicaciones de las anomalías detectadas en la gestión de la planta de manufactura.
- Documentación y Reporte: Se documentarán detalladamente los procedimientos, resultados y conclusiones. Se elaborará un informe final que presente los hallazgos y recomendaciones.

Estas técnicas, instrumentos y procedimientos permitirán realizar un análisis riguroso y efectivo de las series temporales de la nómina diaria en la planta de manufactura, identificando patrones y anomalías que podrían ser cruciales para la optimización de los procesos y la gestión de recursos humanos.

3.5 FUENTES DE INFORMACIÓN

3.5.1 FUENTES PRIMARIAS

El Porvenir Manufacturing es una planta de costura y empaque con una población de aproximadamente 1,500 empleados directos. La planta elabora un total de 5 productos o familias de productos; Camiseta, Sudadera, Pantalón, Short y Panty.

Por la variedad de estilos que elabora, los equipos están constantemente siendo convertidos de un producto a otro, lo que hace que sea la planta más flexible de la compañía en Honduras.

La base de datos a utilizar se extrajo de los registros de planillas de producción en una base

de datos SQL. Se utilizó una consulta en SQL Server para hacer la consulta a la BD y extraer la información de todos los registros del 2023 filtrando empleados activos a la fecha para planta El Porvenir.

3.5.2 FUENTES SECUNDARIAS

No aplica para la investigación actual.

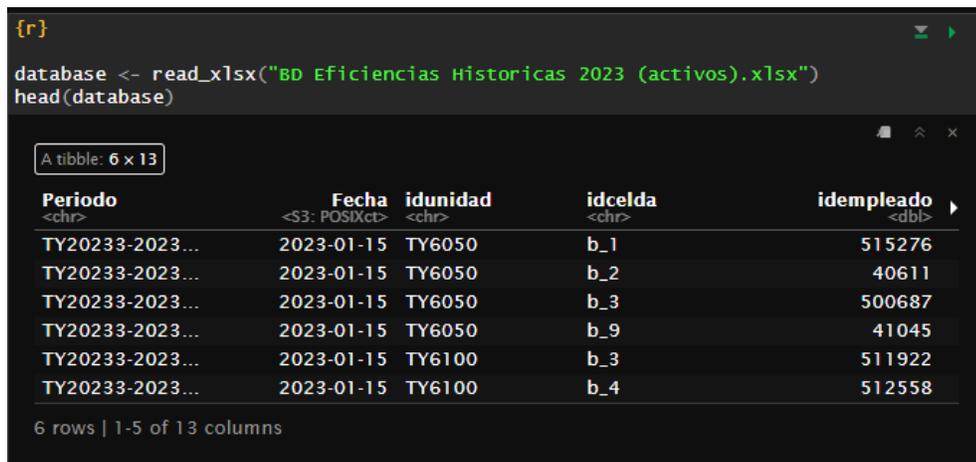
CAPÍTULO IV. RESULTADOS Y ANÁLISIS

4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS

La recolección de datos se hizo utilizando SQL Server. Se extrajo los datos de eficiencia diaria por empleado de enero al 2 de septiembre del año en curso que corresponde a la fecha de cierre de la última planilla pagada.

Muñoz (2020) menciona que para determinar si los datos de una serie son estacionarios, deberán identificarse los periodos de repetición. Este tipo de series son, en ocasiones, periódicas o cuasi periódicas. Puede que no todos los datos de la serie cumplan las características, pero sí un subconjunto de éstos.

Los datos se guardaron en un archivo de Excel y luego fueron cargados a RStudio para su análisis, tal como se muestra en la Fig. 10, los datos se guardaron en un dataframe bajo el nombre “database” utilizando la función “read_xlsx”. La tabla cargada contiene más *de 153,000 registros y 13 columnas*.



```
{r}
database <- read_xlsx("BD Eficiencias Historicas 2023 (activos).xlsx")
head(database)
```

A tibble: 6 x 13

Periodo <chr>	Fecha <S3: POSIXct>	idunidad <chr>	idcelda <chr>	idempleado <dbl>
TY20233-2023...	2023-01-15	TY6050	b_1	515276
TY20233-2023...	2023-01-15	TY6050	b_2	40611
TY20233-2023...	2023-01-15	TY6050	b_3	500687
TY20233-2023...	2023-01-15	TY6050	b_9	41045
TY20233-2023...	2023-01-15	TY6100	b_3	511922
TY20233-2023...	2023-01-15	TY6100	b_4	512558

6 rows | 1-5 of 13 columns

Figura 10. Base de Datos de Eficiencias Diarias por Empleado

En la Fig. 11 se presenta el código con el que se ejecutó la limpieza de los datos. Primero se filtró la data para excluir unidades de Maternidades e Inspectores que son empleados que no generan eficiencia. Luego se procedió a eliminar las columnas innecesarias y se conservaron únicamente 3 columnas, idempleado, fecha y eficiencia operador. Por último, se filtraron los empleados que no tuvieran un mínimo de 2 semanas laboradas (9 puntos de eficiencia). Los dataframe con los datos “limpios” se guardaron en la variable “sample_data_individual”.

```
{r}
sample_data_individual <- database %>%
  filter(!(idunidad %in% c("TY MAT", "TY-INSPECT"))) %>%
  select(-c(Periodo, exp_nombres_apellidos, idunidad, idcelda, idtipoemp,
  tiempo_on_normal, Textra25, `Tiempo off`, SAMS, `SAMS en off`)) %>%
  group_by(idempleado) %>%
  filter(!n() < 9) %>%
  ungroup()

head(sample_data_individual)
```

Fecha	idempleado	Eficiencia Oper...
<S3: POSIXct>	<dbl>	<dbl>
2023-01-15	515276	0.3298995
2023-01-15	40611	0.5597501
2023-01-15	500687	0.5446248
2023-01-15	41045	0.5624477
2023-01-15	511922	1.0093100
2023-01-15	512558	1.0021900

6 rows

Figura 11. Limpieza y Transformación de Datos

En la Fig. 12 se visualizan los datos de eficiencia diaria de 4 empleados. Aunque se observan varios puntos que se podrían considerar como anomalías, se conservaron los datos históricos tal cual se dieron para el análisis preliminar.

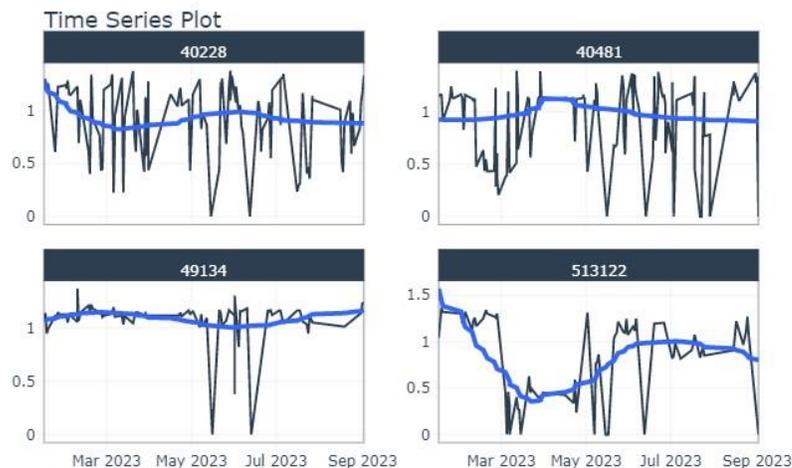


Figura 12. Visualización de Datos Limpios

4.2 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

Para el procesamiento de el alto volumen de series que se deben analizar, se seleccionó el paquete “modeltime” de R que funciona bajo el ecosistema de “Tidymodels”. Este paquete permite el entrenamiento y análisis de múltiples series de tiempo de forma simultánea. Adicionalmente, modeltime permite el entrenamiento de varios modelos por cada serie, sin necesidad de la utilización de **loops**.

Siguiendo la documentación del paquete modeltime, como primer paso, se modificó el formato del dataframe sample_data_individual un formato anidado o “nested tibble” bajo la variable “serie_muestra” necesario para correr los modelos a escala. La tabla ahora contiene un formato pivoteado que detalla un registro por empleado con sus datos y al mismo tiempo la partición de los datos de entrenamiento y prueba.

Para datos de prueba se dejaron 4 puntos o días que es lo que se proyectara semanalmente para detectar anomalías.

```
# A tibble: 1,653 × 4
  idempleado .actual_data      .future_data      .splits
  <dbl> <list>                <list>            <list>
1   515276 <tibble [96 × 3]> <tibble [4 × 3]> <split [92|4]>
2   40611 <tibble [102 × 3]> <tibble [4 × 3]> <split [98|4]>
3   500687 <tibble [97 × 3]> <tibble [4 × 3]> <split [93|4]>
4   41045 <tibble [98 × 3]> <tibble [4 × 3]> <split [94|4]>
5   511922 <tibble [92 × 3]> <tibble [4 × 3]> <split [88|4]>
6   512558 <tibble [68 × 3]> <tibble [4 × 3]> <split [64|4]>
7   512831 <tibble [88 × 3]> <tibble [4 × 3]> <split [84|4]>
8   512544 <tibble [84 × 3]> <tibble [4 × 3]> <split [80|4]>
9   512547 <tibble [88 × 3]> <tibble [4 × 3]> <split [84|4]>
10  515246 <tibble [103 × 3]> <tibble [4 × 3]> <split [99|4]>
```

Figura 13. Transformación de Datos. Tabla convertida en formato anidado (nested tibble), agrupado por empleado y con la separación de datos de entrenamiento y prueba.

Se crearon 3 modelos con los que se hará el entrenamiento:

- XGBoost: con un learn rate de 0.35
- XGBoost: con un learn rate de 0.50
- THIEF

En la Fig. 14 se puede observar el código con el que se corrió una prueba con un solo empleado para determinar si los 3 modelos corrían de forma esperada. Posteriormente se corrió el mismo código sobre los 1,600 empleados.

```
{r}
serie_muestra <- data_individual_agrupada %>%
  slice(1) %>%
  modeltime_nested_fit(
    model_list = list(
      wflow_xgb_1,
      wflow_xgb_2,
      wflow_thief
    ),
    control = control_nested_fit(
      verbose = TRUE,
      allow_par = FALSE
    )
  )
serie_muestra %>%
  extract_nested_test_forecast() %>%
  group_by(idempleado) %>%
  plot_modeltime_forecast()
```

Figura 14. Serie Muestra

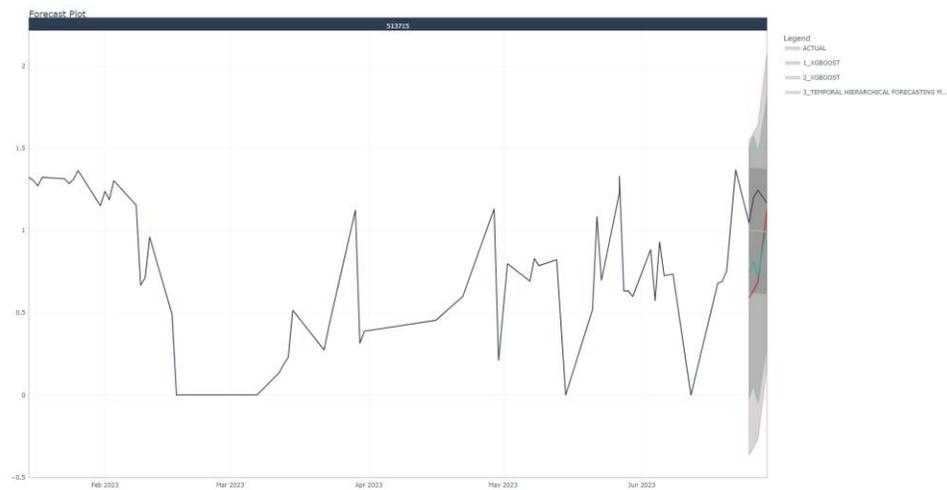


Figura 15. Resultados Serie Muestra

En la Fig. 15 se puede observar una muestra grafica de la proyección de los 4 puntos futuros para un solo empleado.

En la Fig. 16 se puede observar el reporte de exactitud para todos los modelos entrenados generado posterior al entrenamiento de los 3 modelos sobre toda la población de más de 1,600 empleados.

idempleado	.model_id	.model_desc	.type	mae	mape	mase	smape	rmse	rsq
1	513715	1 XGBOOST	Test	0.40	34.65	4.35	44.48	0.46	0.03
2	513715	2 XGBOOST	Test	0.33	28.44	3.60	34.20	0.37	0.00
3	513715	3 TEMPORAL HIERARCHICAL FORECASTING MODEL	Test	0.17	13.92	1.79	15.16	0.18	0.00
4	352602	1 XGBOOST	Test	0.02	1.92	0.67	1.95	0.03	0.00
5	352602	2 XGBOOST	Test	0.03	2.62	0.92	2.67	0.04	0.06
6	352602	3 TEMPORAL HIERARCHICAL FORECASTING MODEL	Test	0.06	4.82	1.69	4.95	0.06	NA
7	350650	1 XGBOOST	Test	0.21	21.02	2.88	18.53	0.23	0.15
8	350650	2 XGBOOST	Test	0.10	9.79	1.35	9.25	0.10	0.68
9	350650	3 TEMPORAL HIERARCHICAL FORECASTING MODEL	Test	0.18	18.46	2.54	16.65	0.19	0.02
10	515430	1 XGBOOST	Test	0.32	44.25	0.87	41.30	0.39	0.13
11	515430	2 XGBOOST	Test	0.38	46.42	1.05	51.80	0.44	0.13
12	515430	3 TEMPORAL HIERARCHICAL FORECASTING MODEL	Test	0.22	38.32	0.60	27.00	0.29	0.00
13	40131	1 XGBOOST	Test	1.15	97.63	13.88	190.74	1.16	NA
14	40131	2 XGBOOST	Test	1.16	98.07	13.94	192.45	1.16	0.33
15	40131	3 TEMPORAL HIERARCHICAL FORECASTING MODEL	Test	0.29	24.78	3.52	30.83	0.37	0.04

Figura 16. Reporte de Exactitud de Modelos Entrenados

Debido al alto volumen de series y modelos a revisar para seleccionar el mejor por serie, se utiliza una opción de forma masiva para esta selección. En la Fig. 17 se puede observar el código con el que se hace esta selección.

De forma predeterminada la función “`modelftime_nested_select_best`” utiliza la métrica RMSE (root mean square error) para seleccionar el mejor modelo de forma individual.

```
{r}
mejores <- serie_escalada_completa %>%
  modelftime_nested_select_best()

extract_nested_best_model_report(mejores)
```

idempleado	.model_id	.model_desc	.type
515276	3	TEMPORAL HIERARC...	Test
40611	3	TEMPORAL HIERARC...	Test
500687	3	TEMPORAL HIERARC...	Test
41045	3	TEMPORAL HIERARC...	Test
511922	2	XGBOOST	Test
512558	2	XGBOOST	Test
512831	2	XGBOOST	Test
512544	2	XGBOOST	Test
512547	1	XGBOOST	Test
515246	2	XGBOOST	Test

Figura 17. Selección Automática de Mejor Modelo por Serie

Hasta este momento, se dispone de un modelo entrenado para cada serie. Este modelo hace un pronóstico de la eficiencia por empleado para los próximos 4 días que representa su jornada semanal inmediata. Ya que el propósito es encontrar anomalías, se debe evaluar la capacidad del pronóstico de identificar las anomalías del periodo proyectado. Para ello, solo se utilizará el límite superior de los intervalos de confianza.

Al correr la comparación entre los datos reales y el límite superior del intervalo de confianza, se observó una detección bastante baja de anomalías. De 4,976 registros de eficiencia solo se denominan anomalías 46 registros. Y dada la naturaleza modular de producción de la planta, para determinar que en efecto es anomalía, debe serlo para todos los miembros de la celda o grupo de trabajo. En este caso no se detectó ninguno.

Debido a que los intervalos de confianza hacen referencia al grado de confianza sobre los parámetros, en este caso la proyección. No es correcto utilizar este límite para identificar anomalías. En su lugar, se debe calcular y utilizar la desviación estándar ya que esta indica la dispersión de los datos.

Se tomo la misma proyección, y se calcularon los límites de tolerancia utilizando la desviación estándar y se detectaron 1,129 posibles anomalías individuales. Al analizar a nivel modular o celda, resulto en 164 celdas con anomalías.

Para determinar el ahorro potencial, se tomó un impacto promedio de 3 puntos de eficiencia por incidencia en base a lo detectado en la muestra. Para definir las horas trabajadas afectadas, se consideró el potencial que tiene cada incidencia en convertirse en una anomalía comprobada y que genere una corrección. Se determino un rango de 472 horas como base, que es lo que se catalogó como una corrección definitiva, y 6,018 horas como techo, asumiendo que todas las incidencias son anomalías.

Por último, se considera el valor por hora pagado de cada una de las 3 plantas que tiene la compañía en el país y se asume un impacto similar para cada instalación debido a las similitudes de proceso de revisión de planillas.

En la Fig. 18 se muestra el cálculo, **el ahorro potencial anual oscila entre \$8,200 a \$105,000.**

Descripcion	Horas Trabajadas	Rate / Hr (por Planta)	Total Correccion Semanal (HNL)	Total Correccion Anual (HNL)	Total Correccion Anual (USD)
Rango Inferior	471.50	1.4398	678.86	32,585.46	1,324.61
		6.0476	2,851.42	136,868.16	5,563.75
		1.4608	688.75	33,059.92	1,343.90
					8,232.26
Rango Superior	6,018.00	1.4398	8,664.69	415,905.23	16,906.72
		6.0476	36,394.16	1,746,919.56	71,012.99
		1.4608	8,790.85	421,960.98	17,152.89
					105,072.59

Figura 18. Cálculo de Ahorro Potencial

Posterior al proceso de selección del mejor modelo para cada serie, se procede al reentrenamiento de los modelos con la data completa, es decir, sin la separación de datos de entrenamiento y prueba. A este proceso se le denomina “refit models” tal como se detalla en la Fig. 19.

```

Reajustar Modelos (Refit)
{r}
parallel_start(6)
modelos_reajustados <- mejores %>%
  modeltime_nested_refit(
    control = control_refit(
      verbose = TRUE,
      allow_par = TRUE
    )
  )
parallel_stop()

```

Figura 19. Reentrenamiento de Modelos con Total de Datos Históricos (refit)

Por último, como se muestra en la Fig. 20, se exporta el forecast por empleado a un archivo formato CSV para ser luego subido a SQL Server con un ETL.

```

Exportar Forecast a CSV
{r}
write.csv(modelos_reajustados %>%
  extract_nested_future_forecast() %>%
  filter(.key == "prediction"), "Predicciones.csv")

```

Figura 20. Exportación de Predicciones en CSV

	idempleado	.model_id	.model_desc	.key	.index	.value	.conf_lo	.conf_hi
1	515276	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/3/2023	1.054612948	0.942496091	1.166729805
2	515276	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/4/2023	1.061323902	0.949207045	1.173440759
3	515276	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/5/2023	1.07037602	0.958259163	1.182492877
4	515276	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/6/2023	1.077086974	0.964970117	1.189203831
5	40611	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/3/2023	0.989812854	0.878394666	1.101231042
6	40611	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/4/2023	1.015719889	0.904301701	1.127138076
7	40611	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/5/2023	1.020554344	0.909136157	1.131972532
8	40611	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/6/2023	1.003631341	0.892213153	1.115049528
9	500687	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/3/2023	0.986809209	0.87938961	1.094228807
10	500687	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/4/2023	0.996406879	0.88898728	1.103826477
11	500687	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/5/2023	0.994045444	0.886625846	1.101465043
12	500687	3	TEMPORAL HIERARCHICAL FORECASTING MODEL	prediction	9/6/2023	0.994045444	0.886625846	1.101465043

Figura 21. Data Forecast

Lo que procede posteriormente es una validación en el sistema de planillas local comparando la eficiencia real procesada vrs los límites de tolerancia proyectados con el análisis de series de tiempo. Esta validación se agregará a la validación ya existente de eficiencias >140% como se muestra en la Fig. 22.

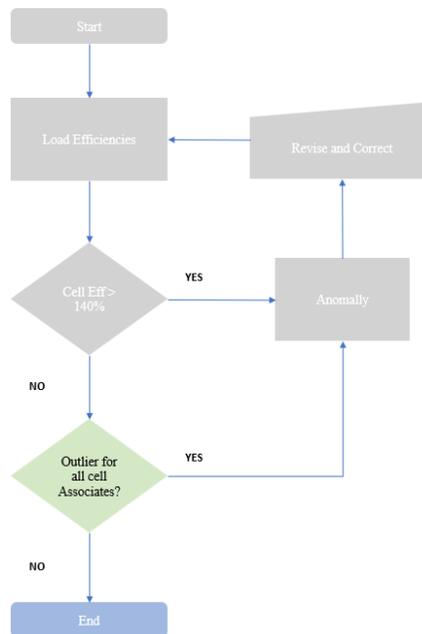


Figura 22. Adición de Detección y Corrección de Anomalías

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.

5.1 CONCLUSIONES

Las conclusiones de esta investigación sobre la aplicación de modelos de aprendizaje automático en el análisis de series temporales y el pronóstico de anomalías en la nómina diaria en una planta de manufactura representan un paso significativo hacia la optimización y eficiencia en la gestión laboral. A lo largo de este estudio, se han explorado y evaluado minuciosamente los resultados obtenidos, considerando tanto los logros alcanzados como las áreas de mejora identificadas. Estas conclusiones no solo reflejan el impacto potencial en la industria manufacturera, sino también el impulso hacia una gestión de nómina más precisa y eficaz. A continuación, se presentan de manera detallada las principales conclusiones derivadas de esta investigación.

1. Esta investigación confirma que la aplicación de modelos de aprendizaje automático en la revisión diaria de la nómina en una planta de manufactura ayuda a identificar posibles anomalías en los pagos y con ello generaría un ahorro para la empresa.
2. En conclusión, esta investigación ha identificado que las metodologías y técnicas más apropiadas para detectar valores atípicos en los datos de eficiencia de la revisión de nóminas en una planta de manufactura son los modelos de aprendizaje automático, como el análisis de series de tiempo. La preparación cuidadosa de los datos es esencial para mejorar la precisión en la detección de valores atípicos. Estos modelos tienen la capacidad de aprender patrones complejos y se adaptan a diferentes contextos, lo que los hace efectivos y versátiles para abordar este desafío en la gestión de nómina.
3. En resumen, esta investigación ha demostrado que es posible diseñar un sistema de detección de valores atípicos adaptable a las particularidades de los procesos y variables en la revisión de nóminas de una planta de manufactura. Los modelos de aprendizaje automático muestran flexibilidad para ajustarse a diferentes contextos. La personalización en la preparación de datos es esencial, al considerar características específicas y necesidades del entorno laboral. Además, se destaca la

importancia de una evaluación continua y el aprendizaje para mantener la efectividad del sistema en un entorno laboral en constante cambio. Esto proporciona una base sólida para mejorar la gestión de nómina en plantas de manufactura.

4. Por último, esta investigación ha demostrado que la implementación de un sistema de detección de valores atípicos en la revisión de nóminas en una planta de manufactura tiene un impacto potencial significativo en términos de ahorros para la empresa. Esto se logra mediante la reducción de errores en la gestión de nómina, mayor transparencia, eficiencia operativa y una mejora en la competitividad de la empresa. Estos resultados validan la utilidad de los modelos de aprendizaje automático en la gestión laboral en el contexto de la manufactura.

5.2 RECOMENDACIONES

En la era de la automatización y la toma de decisiones basada en datos, el análisis de series temporales y el pronóstico de anomalías desempeñan un papel fundamental en diversas industrias. En particular, en el contexto de una planta de manufactura, la gestión eficiente de la nómina diaria es esencial para garantizar la productividad y la optimización de recursos. Esta tesis se sumerge en el emocionante mundo de la inteligencia artificial y se centra en la aplicación de modelos de aprendizaje automático para abordar los desafíos específicos relacionados con la nómina en una planta de manufactura. En este trabajo exploramos cómo estos modelos revolucionan el análisis de las series temporales y se pronostican anomalías en el contexto de la gestión laboral, ofreciendo valiosas recomendaciones para asegurar la medición correcta de la eficiencia y la toma de decisiones en esta industria crítica.

Para el desarrollo y mejor aplicación del modelo de modelos de aprendizaje automático para el análisis de series temporales y pronóstico de anomalías en la nómina diaria en una planta de manufactura son los siguientes:

- Se recomienda analizar los datos históricos, detectar e imputar las anomalías de la data de entrenamiento para ayudar a reducir los intervalos de tolerancia.
- Ya que la librería “Modeltime” permite correr múltiples modelos de forma simultánea a cada serie, se recomienda agregar modelos al entrenamiento (i.e. ARIMA, Exponential Smoothing, Prophet, Linear Regresión, MARS) para así tener más opciones al momento de filtrar el mejor modelo por cada serie. Esto es especialmente útil cuando se dan errores en ciertos modelos.
- Se recomienda explorar y experimentar con más variables para agregar al proceso validaciones adicionales que ayuden a identificar los posibles eventos atípicos tal como la existencia de fuera de estándar, la comparación de eficiencia operador vrs eficiencia planta, etc.
- Ya habiendo explorado la capacidad de detección de anomalías utilizando los intervalos de confianza vrs intervalos de tolerancia calculados con la desviación estándar, se recomienda la utilización de los intervalos de tolerancia experimentando con 2 o 3 sigmas de tolerancia.

- Se recomienda implementar un proceso integral para clasificar y documentar las diferentes fuentes de anomalías identificadas y confirmadas, así como los tipos de anomalías en la nómina diaria. Este proceso permitirá una detección precisa de anomalías, acciones correctivas efectivas y mejora continua.

CAPÍTULO VI. APLICABILIDAD.

La aplicación de modelos de aprendizaje automático en el análisis de series temporales y el pronóstico de anomalías en la nómina diaria es un tema de investigación de tesis que abre nuevas perspectivas en la gestión eficiente de recursos humanos en plantas de manufactura. En un entorno empresarial donde la precisión y la anticipación son fundamentales, la utilización de técnicas avanzadas de aprendizaje automático puede proporcionar una herramienta poderosa para identificar irregularidades en la nómina, mejorar la toma de decisiones y optimizar la gestión de personal en tiempo real.

6.1 NOMBRE DE LA PROPUESTA

Aplicación de modelos de aprendizaje automático para el análisis de series temporales y pronóstico de anomalías en la nómina diaria en una planta de manufactura.

6.2 JUSTIFICACIÓN DE LA PROPUESTA

La aplicación de aprendizaje automático en la gestión de nómina en una planta de manufactura mejora la eficiencia, reduce errores y permite una planificación de personal más efectiva. Esto resulta en ahorros de costos y una operación más eficiente. En resumen, el aprendizaje automático beneficia la gestión de recursos humanos y la toma de decisiones en la planta de manufactura.

6.3 ALCANCE DE LA PROPUESTA

La propuesta implica aplicar aprendizaje automático en una planta de manufactura para:

- **Analizar nóminas y detectar patrones y anomalías.** La aplicación de aprendizaje automático permitirá realizar un análisis detallado de las nóminas diarias en la planta de manufactura. Los modelos de aprendizaje automático serán capaces de identificar patrones recurrentes, estacionales y tendencias en los datos, lo que facilitará una comprensión más profunda de la dinámica de la nómina. Además, se enfocarán en la detección de anomalías que puedan indicar irregularidades en la asistencia, los salarios o el comportamiento de los empleados. Esta capacidad es fundamental para mantener la integridad y la precisión de los registros de nómina.
- **Prever posibles problemas en la nómina.** La propuesta tiene como objetivo

prever posibles problemas en la nómina antes de que se conviertan en desafíos significativos para la planta de manufactura. Los modelos de aprendizaje automático analizarán las tendencias históricas y los datos en tiempo real para identificar desviaciones y anomalías que puedan sugerir problemas potenciales. Esta anticipación permitirá a la empresa tomar medidas preventivas de manera proactiva, evitando costosos errores y interrupciones en la operación.

- **Optimizar la asignación de personal.** Una de las aplicaciones clave de la propuesta es la optimización de la asignación de personal en la planta de manufactura. Los modelos de aprendizaje automático podrán proporcionar recomendaciones sobre la asignación óptima de trabajadores en función de la demanda de producción, las horas de trabajo y otras variables relevantes. Esta optimización no solo aumentará la eficiencia operativa, sino que también ayudará a reducir los costos laborales innecesarios.
- **Reducir errores y costos.** La aplicación de aprendizaje automático tiene el potencial de reducir significativamente los errores y los costos asociados a la gestión de la nómina y los recursos humanos. Al detectar anomalías y anticipar problemas, se minimizarán las correcciones manuales y los costos derivados de errores en la nómina. Además, la optimización de la asignación de personal contribuirá a una utilización más eficiente de los recursos humanos, lo que resultará en una disminución de los costos laborales.
- **Mejorar la planificación de personal.** La propuesta también busca mejorar la planificación de personal en la planta de manufactura. Los modelos de aprendizaje automático proporcionarán información valiosa para planificar la dotación de personal de manera más precisa, teniendo en cuenta las necesidades de producción, las fluctuaciones estacionales y otros factores influyentes. Esto permitirá una programación más eficiente y evitará la subutilización o sobreutilización de recursos humanos.
- **Capacitar al personal y ajustar continuamente los modelos.** La implementación exitosa de esta propuesta requerirá la capacitación del personal en el manejo de la solución basada en aprendizaje automático. El personal responsable de la nómina y

la gestión de recursos humanos deberá adquirir las habilidades necesarias para utilizar eficazmente los modelos y comprender sus resultados. Además, se establecerá un proceso de ajuste continuo de los modelos para adaptarlos a los cambios en las operaciones de la planta y los patrones emergentes.

6.4 DESCRIPCIÓN Y DESARROLLO

La propuesta implica usar el aprendizaje automático para mejorar la gestión de recursos humanos en una planta de manufactura, desde el análisis de datos de nómina hasta la toma de decisiones más eficientes. Esto se logra mediante la recopilación y análisis de datos, el desarrollo de modelos de aprendizaje automático, la identificación de patrones y anomalías, y la optimización de la asignación de personal. Se asegura un proceso de evaluación constante y la capacitación del personal para mantener y ajustar los modelos. En resumen, la propuesta busca eficiencia y toma de decisiones basadas en datos en la gestión de recursos humanos.

6.4.1 DESCRIPCIÓN

Qué Haremos: Implementar modelos de aprendizaje automático para mejorar la gestión de la nómina diaria en una planta de manufactura.

Cómo lo Haremos:

- **Recopilación y preparación de Datos Históricos de Nómina.** El primer paso en la implementación de esta metodología consistirá en la recopilación y preparación de datos históricos de nómina. Esta fase será esencial para garantizar que los datos utilizados para el entrenamiento de los modelos sean precisos y confiables. Se realizará una exhaustiva recolección de registros que abarquen un período significativo de tiempo. Los datos recopilados incluirán información detallada sobre las nóminas diarias, como horas de trabajo, salarios, datos de asistencia, y otros datos relacionados con los empleados. La preparación de datos implica la limpieza, la estandarización y la transformación de los datos en un formato adecuado para su análisis.
- **Desarrollo de Modelos de Aprendizaje Automático para el Análisis de Series Temporales y Detección de Anomalías.** Una vez que los datos históricos estén disponibles y preparados, se procederá al desarrollo de modelos de aprendizaje

automáticos diseñados específicamente para el análisis de series temporales y la detección de anomalías en las nóminas diarias. Estos modelos serán seleccionados en función de su capacidad para abordar las características de los datos y los objetivos específicos de la planta de manufactura. Se considerarán modelos como ARIMA, redes neuronales recurrentes (RNN) y modelos de detección de anomalías basados en autoencoders.

- **Entrenamiento de los Modelos con Datos Históricos.** Una vez que se hayan desarrollado los modelos, se procederá al entrenamiento utilizando los datos históricos preparados. Esta etapa es crucial para que los modelos aprendan los patrones y las tendencias presentes en los datos de nómina. Se realizará un proceso de ajuste de hiperparámetros y validación cruzada para garantizar la robustez y el rendimiento de los modelos.
- **Análisis de la Nómina en Tiempo Real y Pronóstico de Anomalías.** Con los modelos entrenados en funcionamiento, se analizará la nómina diaria en tiempo real. Los modelos serán capaces de pronosticar anomalías y detectar desviaciones significativas en comparación con los patrones históricos. Cualquier anomalía identificada se informará al personal encargado para que se puedan tomar medidas preventivas o correctivas de manera oportuna.
- **Evaluaciones continuas y Capacitación del Personal.** La metodología incluirá evaluaciones continuas del rendimiento de los modelos y la identificación de posibles áreas de mejora. El personal encargado de la nómina y la gestión de recursos humanos recibirá capacitación para utilizar efectivamente la solución basada en aprendizaje automático y comprender sus resultados. Además, se establecerá un proceso de ajuste continuo de los modelos para adaptarlos a los cambios en las operaciones de la planta y los patrones emergentes.

En resumen, se aplicará el aprendizaje automático para analizar la nómina, detectar problemas y optimizar la asignación de personal en la planta de manufactura.

6.4.2 DESARROLLO

El proceso de implementación de modelos de aprendizaje automático en una planta de

manufactura implica una serie de componentes y herramientas esenciales para garantizar el éxito de la solución. A continuación, se describen en detalle estos elementos clave, proporcionando una guía completa para la empresa:

- **Herramientas de Programación: R y RStudio como Entorno de Desarrollo.**

El proceso de implementación se basará en el uso de las herramientas de programación R y RStudio como el entorno de desarrollo principal. R es un lenguaje de programación ampliamente utilizado en la comunidad de ciencia de datos y análisis estadístico, y RStudio proporciona una interfaz amigable para el desarrollo de código en R. Estas herramientas son esenciales para la programación, la implementación de modelos y la visualización de resultados.

- **Bibliotecas de Aprendizaje Automático: Tidyverse, XGBoost, THIEF y Forecast.**

Se utilizarán diversas bibliotecas de aprendizaje automático para la implementación de los modelos. Esto incluirá el uso de Tidyverse para la manipulación y limpieza de datos, XGBoost para la construcción de modelos de clasificación y regresión, THIEF para el análisis de series temporales y Forecast para el pronóstico de datos. Estas bibliotecas son componentes fundamentales para el desarrollo y el entrenamiento de modelos de aprendizaje automático.

- **Recopilación de Datos de la Nómina Diaria.**

La recopilación de datos de la nómina diaria será el punto de partida del proceso. Se requerirá un sistema eficiente para la adquisición de datos en tiempo real o la extracción de datos históricos de sistemas de registro existentes. La calidad de los datos es esencial para el éxito del proyecto, y se deben establecer procesos para garantizar la integridad de los datos recopilados.

- **Limpieza y Preparación de Datos.**

La limpieza y preparación de datos son pasos críticos en el proceso de implementación. Esto implica la eliminación de duplicados, valores atípicos y datos faltantes, así como la realización de transformaciones necesarias en los datos. Un flujo de trabajo eficiente y reglas predefinidas garantizarán la consistencia y la calidad de los datos antes de su uso en la construcción y entrenamiento de modelos.

- **Entrenamiento de Modelos Utilizando Datos Históricos.**

Los modelos de aprendizaje automático se entrenarán utilizando datos históricos de la nómina. Estos datos servirán como base para que los modelos aprendan patrones y tendencias en los registros de nómina. El proceso de entrenamiento incluirá la selección de algoritmos adecuados, la partición de datos en conjuntos de entrenamiento y validación, y el ajuste de hiperparámetros para optimizar el rendimiento de los modelos.

- **Análisis en Tiempo Real y Pronóstico de Anomalías.**

Una vez que los modelos estén entrenados, se configurará un flujo de datos para permitir el análisis en tiempo real de la nómina. Los modelos serán capaces de pronosticar y detectar anomalías en tiempo real, lo que permitirá una respuesta inmediata a problemas potenciales en la gestión de recursos humanos.

- **Optimización de Recursos Humanos Utilizando los Resultados de los Modelos.**

La información generada por los modelos se utilizará para optimizar la asignación de personal de manera eficiente. Las recomendaciones de los modelos sobre la asignación de trabajadores se incorporarán en la toma de decisiones de gestión de recursos humanos, lo que contribuirá a una utilización más efectiva de los recursos humanos y ahorros en costos laborales.

- **Evaluación Continua para Medir la Precisión de los Modelos y Realizar Ajustes.**

La evaluación continua de los modelos será una parte integral del proceso de implementación. Se llevarán a cabo pruebas periódicas para medir la precisión de los modelos en la detección de anomalías y el pronóstico de la nómina. Los resultados se utilizarán para realizar ajustes y mejoras en los modelos a medida que se requieran.

- **Capacitación del Personal para Utilizar y Mantener los Modelos.**

La capacitación del personal es fundamental para el éxito continuo de la solución. Los empleados encargados de la nómina y la gestión de recursos humanos recibirán capacitación en el uso de la solución basada en aprendizaje automático. Esto incluirá la interpretación de resultados, la toma de decisiones basada en los datos generados por los modelos y la capacidad de realizar ajustes menores en el sistema.

- **Documentación y Reportes para Comunicar Resultados.**

Se generará documentación detallada sobre la implementación de la solución y los resultados obtenidos. Esto permitirá una comunicación efectiva de los hallazgos y recomendaciones tanto dentro de la empresa como con otras partes interesadas. Los reportes generados facilitarán la toma de decisiones informadas y la retroalimentación constante.

- **Seguridad de Datos para Proteger la Privacidad y la Integridad de los Datos.**

La seguridad de los datos es un aspecto crítico de la implementación. Se establecerán protocolos de seguridad para proteger la privacidad y la integridad de los datos de la nómina. Esto incluirá medidas de seguridad cibernética, políticas de acceso y el cumplimiento de regulaciones de privacidad de datos aplicables.

- **Monitoreo y Mantenimiento para Asegurar un Funcionamiento Continuo.**

El monitoreo constante será esencial para garantizar el funcionamiento continuo de la solución. Se establecerán alertas y sistemas de monitoreo que permitirán identificar y abordar de manera proactiva cualquier problema que pueda surgir. El mantenimiento regular y las actualizaciones de los modelos y del sistema se realizarán para mantener el rendimiento óptimo.

- **Escalabilidad para Adaptarse a Cambios en la Planta.**

La solución se diseñará con la escalabilidad en mente, lo que permitirá adaptarse a cambios en la planta de manufactura. A medida que la operación evolucione y crezca, la solución podrá ajustarse y expandirse para satisfacer las necesidades cambiantes.

- **Comunicación con los Interesados para Cumplir con los Objetivos del Proyecto.**

La comunicación efectiva con los interesados será fundamental para cumplir con los objetivos del proyecto. Se establecerán canales de comunicación claros y se mantendrá una colaboración constante con los departamentos de recursos humanos y operaciones para garantizar que la solución se alinee con los objetivos estratégicos de la empresa.

6.5 MEDIDAS DE CONTROL

La implementación de modelos de aprendizaje automático en la gestión de la nómina diaria en una planta de manufactura requerirá un conjunto sólido de medidas de control para garantizar que el proyecto cumpla con sus objetivos y se alinee de manera efectiva con los objetivos

estratégicos de la empresa. La comunicación efectiva y la colaboración constante con los interesados son elementos clave de estas medidas de control.

- **Establecimiento de Canales de Comunicación Claros**

Para garantizar una implementación exitosa, se establecerán canales de comunicación claros y efectivos entre los equipos de proyecto, los departamentos de recursos humanos, operaciones y otros interesados relevantes. La comunicación será bidireccional, permitiendo la retroalimentación constante y la resolución de problemas de manera oportuna. Los canales de comunicación incluirán reuniones regulares, informes periódicos y sistemas de seguimiento de problemas.

- **Colaboración Constante con Departamentos Clave**

Se mantendrá una colaboración constante con los departamentos de recursos humanos y operaciones, que son los principales interesados en la aplicación de modelos de aprendizaje automático. Esta colaboración involucrará la alineación de los objetivos del proyecto con los objetivos estratégicos de la empresa. Los equipos de proyecto trabajarán en estrecha colaboración con los departamentos de recursos humanos y operaciones para garantizar que la solución se adapte de manera efectiva a las necesidades de la planta de manufactura.

- **Aseguramiento de la Comprensión y Aceptación de la Solución**

Se realizarán esfuerzos significativos para asegurar que los interesados comprendan y acepten la solución basada en aprendizaje automático. Esto implicará la capacitación del personal relevante en el uso de la solución y la explicación clara de cómo los modelos de aprendizaje automático pueden mejorar la gestión de la nómina y la asignación de personal. La aceptación y el compromiso de los interesados son críticos para el éxito del proyecto.

- **Gestión de Expectativas y Resolución de Conflictos**

La gestión de expectativas será una parte integral de las medidas de control. Se abordarán las expectativas realistas y los posibles desafíos que puedan surgir durante la implementación. Además, se establecerán procedimientos para la resolución de conflictos en caso de desacuerdos o problemas no resueltos. La resolución eficaz de conflictos garantizará que el proyecto siga avanzando sin obstáculos significativos.

- **Monitoreo y Evaluación Continua**

Se implementará un sistema de monitoreo y evaluación continua para seguir de cerca el progreso del proyecto. Los indicadores clave de rendimiento (KPI) se establecerán para medir el éxito y el impacto de la solución en la gestión de la nómina y la asignación de personal. Cualquier desviación de los objetivos establecidos se abordará de manera proactiva.

- **Retroalimentación y Evaluación Continua**

La retroalimentación constante será alentada y valorada como parte de las medidas de control. Los comentarios de los interesados, el personal y otros involucrados se utilizarán para realizar mejoras continuas en la solución y en el proceso de implementación. Se establecerá un ciclo de mejora continua para garantizar que la solución evolucione con las necesidades cambiantes de la empresa y la planta de manufactura.

6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

Cronograma de Implementación:

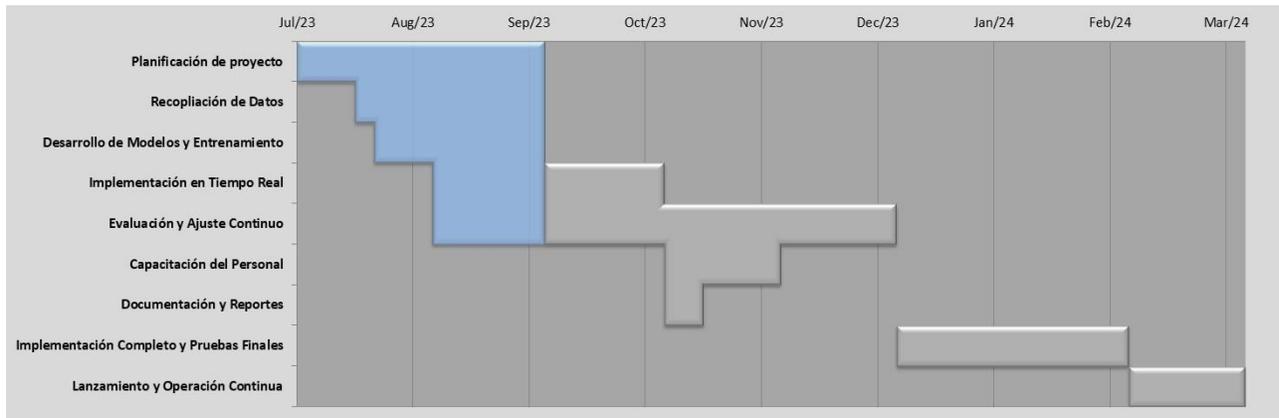


Figura 23. Cronograma de Implementación

El presupuesto para el proyecto de investigación "Aplicación de Modelos de Aprendizaje Automático para el Análisis de Series Temporales y Pronóstico de Anomalías en la Nómina Diaria en una Planta de Manufactura" se compone de diversos elementos esenciales que garantizarán el éxito de la implementación de la solución de aprendizaje automático en la gestión de la nómina. Cada uno de estos elementos se presenta a continuación en formato de tesis y en tercera persona:

Presupuesto Estimado:

- **Personal**

El presupuesto incluye los costos asociados con el personal necesario para el proyecto. Esto abarca tanto el equipo de proyecto interno como cualquier personal adicional que pueda ser requerido para la implementación y el mantenimiento continuo de la solución. Los costos de personal incluyen salarios, beneficios y posibles bonificaciones por el éxito del proyecto. Dicho costo incurre en un valor de L 15,000 para las reuniones de capacitación de personal que se agenden durante el periodo de implementación del proyecto.

- **Hardware y Software**

Se asignarán recursos para la adquisición de hardware y software necesarios para la implementación de la solución. Esto incluye la compra de servidores, estaciones de trabajo, licencias de software, y otros recursos tecnológicos esenciales para ejecutar los modelos de aprendizaje automático y almacenar los datos. El costo de almacenaje de información en servidor de Amazon Web Service es de L 9,977.76 (\$405.60), tres licencias de software empresarial de Microsoft que tienen un costo total de L 541.2 para un total de 2 maquinas serian L 1,082.4 (\$22 c/u). Y por ultimo el costo de dos equipos que estarán constantemente destinado a procesar tema de anomalías con costo de L 25,000 (L 12,500 c/u).

- **Datos**

La adquisición y preparación de datos requerirá recursos financieros para garantizar la calidad y la disponibilidad de los datos de nómina necesarios para el proyecto. Esto puede incluir costos asociados con la extracción de datos históricos, la integración de sistemas y la adquisición de fuentes de datos adicionales si es necesario. El monto total de inversión para la adquisición y preparación de datos, que comprende la garantía de calidad y disponibilidad de los datos de nómina esencial para el proyecto, abarcará costos relacionados con la extracción de datos históricos, la integración de sistemas y la adquisición de fuentes de datos adicionales, si es necesario, asciende a L 3,000.

- **Capacitación**

La capacitación del personal en el manejo de la solución de aprendizaje automático es

esencial. Los recursos financieros se destinarán a la capacitación tanto del personal de recursos humanos como de operaciones para garantizar un entendimiento completo de la solución y la capacidad de utilizarla de manera efectiva. El costo asociado a las reuniones de capacitación de personal programadas durante la fase de implementación del proyecto asciende a L 5,000.

- **Gastos Operativos**

Los gastos operativos incluyen costos continuos necesarios para el funcionamiento de la solución, como el consumo de energía, el mantenimiento de equipos, la infraestructura de TI y otros gastos operativos asociados con la gestión de la solución. Los gastos operativos incluyen costos continuos necesarios para el funcionamiento de la solución, como el consumo de energía, el mantenimiento de equipos, la infraestructura de TI y otros gastos operativos asociados con la gestión de la solución, ascendiendo a L 4,500.

- **Consultoría Externa**

En algunos casos, la consultoría externa puede ser requerida para proporcionar experiencia adicional en la implementación de modelos de aprendizaje automático. Los costos de consultoría se incluirán en el presupuesto si se determina que son necesarios. En algunos casos, la consultoría externa puede ser requerida para proporcionar experiencia adicional en la implementación de modelos de aprendizaje automático. Los costos de consultoría se incluirán en el presupuesto si se determina que son necesarios, ascendiendo a L 8,500.

Tabla 3. Presupuesto Estimado

Resumen de Presupuesto Estimado	
Motivo	Monto
Personal	L 15,000
Hardware y Software	L 36,060.16
Datos	L 8,000
Gastos Operativos	L 4,500
Consultoría Externa	L 8,500
TOTAL	L 72,060.16

6.7 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

Tabla 4. Concordancia de Segmentos de Tesis

Título de investigación	Capítulo I		Capítulo II		
	Objetivo general	Objetivo específico	Teorías/Metodologías de sustento		
Aplicación de modelos de aprendizaje automático para el análisis de series temporales y pronóstico de anomalías en la nómina diaria en una planta de manufactura.	Diseñar un sistema de detección de valores atípicos en las eficiencias diarias de la revisión de nóminas en una planta de manufactura, con el propósito de garantizar la precisión en la remuneración de los trabajadores y generar ahorros tangibles para la empresa.	Evaluar y seleccionar las metodologías y técnicas más adecuadas para la detección de valores atípicos en los datos de eficiencia de la revisión de nóminas.	Teoría de Series Temporales		
		Diseñar un sistema de detección de valores atípicos que se adapte de manera precisa a las particularidades de los procesos y variables involucradas en la revisión de nóminas. Evaluar el impacto potencial de la implementación del sistema de detección en términos de la generación de ahorros para la empresa.	Teorías de Detección de Anomalías Detección de Anomalías		
	Capítulo III		Capítulo V	Capítulo VI	
Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la propuesta-objetivo de la propuesta	
Periodo de planilla Fechas Tipo de empleado, ubicación (unidad- celda)	1600 empleados de la planta El porvenir Manufacturing	Análisis de series temporales Modelos de pronóstico Detección de anomalías	La aplicación de aprendizaje automático en la detección de anomalías en nóminas de manufactura mejora la precisión de pagos y ahorra costos y optimiza la gestión de nómina. Esta investigación respalda el uso de aprendizaje automático, como el análisis de series de tiempo, para detectar valores atípicos en nóminas de manufactura, destacando la importancia de la preparación de datos y su eficacia en la gestión de nómina.	Aplicación de modelos de aprendizaje automático para el análisis de series temporales y pronóstico de anomalías en la nómina diaria en una planta de manufactura.	Implementar modelos de aprendizaje automático para mejorar la gestión de la nómina diaria en una planta de manufactura.

REFERENCIA BIBLIOGRAFICA

- Alcalde, A. (2018, March 4). *Aprendizaje no Supervisado y Detección de Anomalías: ¿Qué es una Anomalía?* <https://elbauldelprogramador.com/aprendizaje-nosupervisado-anomalias/>
- ALGOTIVE. (2022, March 18). *Machine Learning: ¿Qué es el aprendizaje automático y cómo funciona?* [https://www.algotive.ai/es-mx/blog/machine-learning-que-es-el-aprendizaje-autom%C3%A1tico-y-c%C3%B3mo-funciona#:~:text=El%20Machine%20Learning%20\(ML\)%20o,que%20ser%20programadas%20para%20ello.](https://www.algotive.ai/es-mx/blog/machine-learning-que-es-el-aprendizaje-autom%C3%A1tico-y-c%C3%B3mo-funciona#:~:text=El%20Machine%20Learning%20(ML)%20o,que%20ser%20programadas%20para%20ello.)
- Amazon Web Services. (n.d.). *Validación cruzada*. https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/cross-validation.html
- Athanasopoulos, G., Hyndman, R., Kourentzes, N., & Petropoulos, F. (n.d.). *Forecasting with Temporal Hierarchies*. robjhyndman.com/temporalhierarchies.pdf
- Avila, L., Mendoza, N., & Alonso, A. (2019, July). *Detección de anomalías basada en aprendizaje profundo: Revisión*. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992019000300107#:~:text=La%20detecci%C3%B3n%20de%20anomal%C3%ADas%20es,o%20an%C3%B3malas%20sobre%20los%20datos.
- Chinlli, C. (2021). *Modelización de Series Temporales modelos clásicos y SARIMA*. masteres.ugr.es/TFM_MIRANDA_CHINLLI_CARLOS.pdf
- EDSRobotics. (2020, December 5). *¿Qué es la Industria 4.0?* [https://www.edsrobotics.com/blog/que-es-la-industria-4-0/#:~:text=y%20menos%20desperdicios.-,Desventajas%20de%20la%20industria%204.0%3A,de%20interferencia%20en%20las%](https://www.edsrobotics.com/blog/que-es-la-industria-4-0/#:~:text=y%20menos%20desperdicios.-,Desventajas%20de%20la%20industria%204.0%3A,de%20interferencia%20en%20las%20)

20comunicaciones

El Blog Ruralvía. (2023). *Desafíos y oportunidades para las empresas en la era de la cuarta revolución industrial*. <https://blog.ruralvia.com/empresas-en-la-cuarta-revolucion-industrial/>

Flores, C., & Ortiz, M. (2018). *Revisión de algoritmos para la detección de valores atípicos* [Universidad Católica de Cuenca].

pdfs.semanticscholar.org/a8f5c67aea4861b45caea912ace8da248b9f.pdf

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles and practice; [a comprehensive introduction to the latest forecasting methods using R; learn to improve your forecast accuracy using dozens of real data examples]*. Otexts.

Iglesias, I. (2022). *ANÁLISIS Y DETECCIÓN DE ANOMALÍAS USANDO TÉCNICAS DE DEEP LEARNING: CASO DE ESTUDIO EN COTIZACIONES DE PFIZER Y MODERNA*.

www.repositorio.comillas.edu/retrieve

José Alberto Mauricio. (2007). *Introducción al Análisis de Series Temporales*. Universidad Complutense Madrid. www.ucm.es/518-2013-11-11-JAM-IAST-Libro

Lee, M.-C., Lin, J.-C., & Gran, E. G. (2023). *RePAD: Real-time Proactive Anomaly Detection for Time Series* (arXiv:2001.08922). arXiv. <https://doi.org/10.48550/arXiv.2001.08922>

Mendoza, J. (2020, November 24). *XGBoost o Extreme Gradient Boosting en R*.

<https://estadisticamente.com/xgboost-o-extreme-gradient-boosting-en-r/>

Muñoz, J. (2020). *Codificación y procesamiento de series temporales multivariantes atendiendo a su estructura temporal*. repositorio.uam.es/munoz_aguado_jorge_tgf.pdf

Patel, K. (2023). *Implicaciones de la Industria 4.0 para la infraestructura de red*.

<https://www.computerweekly.com/es/opinion/Implicaciones-de-la-Industria-40-para-la->

infraestructura-de-red

RAE. (n.d.). *Definición anomalía*. <https://dle.rae.es/anomal%C3%ADa>

Redacción APD. (2021, September 21). *Pros y contras de la Cuarta Revolución Industrial*.

<https://www.apd.es/ciberindustria-ventajas-desventajas/>

Sanchez, J. (2023, February 27). *Industria 4.0: Qué es, beneficios y ejemplos*.

[https://www.iebschool.com/blog/industria-cuarta-revolucion-industrial-business-tech-](https://www.iebschool.com/blog/industria-cuarta-revolucion-industrial-business-tech-logistica/#:~:text=En%20la%20Industria%204.0%2C%20las,necesidades%20espec%C3)

[logistica/#:~:text=En%20la%20Industria%204.0%2C%20las,necesidades%20espec%C3%ADficas%20de%20los%20consumidores.](https://www.iebschool.com/blog/industria-cuarta-revolucion-industrial-business-tech-logistica/#:~:text=En%20la%20Industria%204.0%2C%20las,necesidades%20espec%C3%ADficas%20de%20los%20consumidores.)

Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series: A

comprehensive evaluation. *Proceedings of the VLDB Endowment*, 15(9), 1779–1797.

<https://doi.org/10.14778/3538598.3538602>

Shumway, R. H. (2017). *Time series analysis and its applications: With r examples*. Springer

Science+Business Media.

TIBCO. (n.d.). *¿Qué es la detección de valores anómalos?* [https://www.tibco.com/es/reference-](https://www.tibco.com/es/reference-center/what-is-outlier-detection)

[center/what-is-outlier-detection](https://www.tibco.com/es/reference-center/what-is-outlier-detection)

Torres-Domínguez, O., Sabater-Fernández, S., Bravo-Ilisatigui, L., Martín-Rodríguez, D., &

García-Borroto, M. (2019). Detección de anomalías en grandes volúmenes de datos.

Revista Facultad de Ingeniería, 28(50), 62–75.

<https://doi.org/10.19053/01211129.v28.n50.2019.8793>

Ulloa, G., Arroyo, S., & Idme, R. (2022, June 5). *Modelos ARIMA, SARIMA y Método de*

Selección de Variables LASSO para Series Temporales.

<https://datasciencepe.substack.com/p/modelos-arima-sarima-y-metodo-de>

ANEXOS

Anexo 1 Código Utilizado

Carga de Librerías a Utilizar

```
library(readxl)
library(tidymodels)
library(modeltime)
library(tidyverse)
library(timetk)
library(thief)
library(ggplot2)
```

Carga de Base de Datos

```
database <- read_xlsx("BD Eficiencias Historicas 2023 (activos).xlsx")
head(database)
```

Generación de Datos por Empleado, filtro, limpieza y exploración

```
sample_data_individual <- database %>%
  filter(!(idunidad %in% c("TY MAT", "TY-INSPECT"))) %>%
  select(-c(Periodo, exp_nombres_apellidos, idunidad, idcelda, idtipoemp, tiempo_on_normal, Textra25, `Tiempo Off`, SAMS, `SAMS en Off`)) %>%
  group_by(idempleado)%>%
  filter(!(n() < 9)) %>%
  ungroup()
```

```
sample_data_individual %>%
  filter(idempleado %in% c("40228", "513122", "49134", "40481"))%>%
  group_by(idempleado) %>%
  plot_time_series(
    .date_var = Fecha,
    .value = `Eficiencia Operador`,
    .facet_ncol = 2,
    .interactive = FALSE,
    .smooth = TRUE
  )
```

```
sample_data_individual %>%
  filter(idempleado %in% c("40228", "513122", "49134", "40481"))%>%
  group_by(idempleado) %>%
  plot_anomaly_diagnostics(
    .date_var = Fecha,
    .value = `Eficiencia Operador`,
    .facet_ncol = 2,
    .interactive = FALSE
  )
```

Preparación de Dataset y Creación de Modelos

Nested Time Series

```
parallel_start(6)

data_individual_agrupada <- sample_data_individual %>%
  group_by(idempleado) %>%
  extend_timeseries(
    .id_var = idempleado,
    .date_var = Fecha,
    .length_future = 4
  ) %>%
  nest_timeseries(
    .id_var = idempleado,
    .length_future = 4
  ) %>%
  split_nested_timeseries(
    .length_test = 4
  )

parallel_stop()
```

Receta XGBoost

```
rec_xgb <- recipe(`Eficiencia Operador` ~ ., extract_nested_train_split(data_
individual_agrupada) ) %>%
  step_timeseries_signature(Fecha)%>%
  step_rm(Fecha) %>%
  step_zv(all_predictors()) %>%
  step_dummy(all_nominal_predictors(), one_hot = TRUE)
```

Modelos XGBoost

Se generan dos modelos XGBoost modificando el parametro "learn_rate"

```
wflow_xgb_1 <- workflow()%>%
  add_model(boost_tree("regression", learn_rate = 0.35)%>%
    set_engine("xgboost")) %>%
  add_recipe(rec_xgb)

wflow_xgb_2 <- workflow()%>%
  add_model(boost_tree("regression", learn_rate = 0.50)%>%
    set_engine("xgboost")) %>%
  add_recipe(rec_xgb)
```

Modelo THIEF (Temporal HIERarchical Forecasting)

```
wflow_thief <- workflow() %>%  
  add_model(temporal_hierarchy(  
    mode = "regression",  
    combination_method = "mse",  
    use_model = "arima" ) %>%  
    set_engine("thief")) %>%  
  add_recipe(recipe(`Eficiencia Operador` ~ ., extract_nested_train_split(data_individual_agrupada)))
```

Prueba y Entrenamiento de Modelos

Prueba con Una Serie Temporal

```
serie_muestra <- data_individual_agrupada %>%  
  slice(1) %>%  
  modeltime_nested_fit(  
  
    model_list = list(  
      wflow_xgb_1,  
      wflow_xgb_2,  
      wflow_thief  
    ),  
  
    control = control_nested_fit(  
      verbose = TRUE,  
      allow_par = FALSE  
    )  
  )  
  
serie_muestra %>%  
  extract_nested_test_forecast() %>%  
  group_by(idempleado) %>%  
  plot_modeltime_forecast()
```

Entrenamiento de Modelos a Escala

```
parallel_start(6)

serie_escalacompleta <- data_individual_agrupada %>%
  modeltime_nested_fit(

  model_list = list(
    wflow_xgb_1,
    wflow_xgb_2,
    wflow_thief
  ),

  control = control_nested_fit(
    verbose = TRUE,
    allow_par = TRUE
  )
)
parallel_stop()
```

Revisar Resultados - Accuracy

```
serie_escalacompleta %>%
  extract_nested_test_accuracy() %>%
  table_modeltime_accuracy() %>%
  save.image()
```

Visualizar Resultados

```
serie_escalacompleta %>%
  extract_nested_test_forecast(.id_subset = 512558, .include_actual = FALSE)
%>%
  ##filter(idempleado == 512558) %>%
  group_by(idempleado) %>%
  plot_modeltime_forecast()
```

Selección de Mejor Modelo por Serie

```
mejores <- serie_escalacompleta %>%
  modeltime_nested_select_best()

extract_nested_best_model_report(mejores)
```

Visualizar Mejores Modelos

```
mejores %>%
  extract_nested_test_forecast() %>%
  filter(idempleado == 49134) %>%
  group_by(idempleado) %>%
  plot_modeltime_forecast()
```

Reajustar Modelos (Refit)

```
parallel_start(6)
```

```
modelos_reajustados <- mejores %>%  
  modelftime_nested_refit(  
    control = control_refit(  
      verbose = TRUE,  
      allow_par = TRUE  
    )  
  )  
)
```

```
parallel_stop()
```

Exportar Forecast a CSV

```
modelos_reajustados %>%  
  extract_nested_future_forecast() %>%  
  filter(.key == "prediction") %>%  
  write.csv("Predicciones.csv")
```

Anexo 2 Cotización AWS por servicio de almacenamiento de data

The screenshot displays the AWS Pricing Calculator interface. The 'Estimate summary' section shows the following costs:

Category	Cost
Upfront cost	0.00 USD
Monthly cost	33.80 USD
Total 12 months cost (Includes upfront cost)	405.60 USD

The 'My Estimate' section includes a search bar for resources and a table of services:

Service Name	Status	Upfront cost	Monthly cost	Description	Region	Config Sum...
Amazon EC2	-	0.00 USD	33.80 USD	-	México (Querét...	Tenancy (Share...

At the bottom, there are links for Privacy, Site terms, and Cookie preferences, along with the copyright notice: © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Anexo 3 Cotización adquisición servicio Microsoft Office

Microsoft 365 Empresa Básico
USD\$6.00
por usuario al mes
(Suscripción anual: se renueva automáticamente)¹
El precio no incluye impuestos.
[Compra ahora](#)
Pruébalo gratis durante un mes ² >
[Más información](#) >

- ✓ Chatea, llama o reúnete con 300 asistentes
- ✓ Versiones web y celulares de las aplicaciones de Office
- ✓ 1 TB de almacenamiento en la nube por

Microsoft 365 Empresa Estándar
USD\$12.50
por usuario al mes
(suscripción anual: se renueva automáticamente)¹
El precio no incluye impuestos.
[Compra ahora](#)
Pruébalo gratis durante un mes ² >
[Más información](#) >

Todo lo que incluye la versión Empresa Básico, y además:

- ✓ Versiones de escritorio de las aplicaciones de Office con características premium

Microsoft 365 Empresa Premium
USD\$22.00
por usuario al mes
(Suscripción anual: se renueva automáticamente)¹
El precio no incluye impuestos.
[Compra ahora](#)
Pruébalo gratis durante un mes ² >
[Más información](#) >

Todo lo que ofrece Empresa Estándar y, además:

- ✓ Seguridad avanzada
- ✓ Control de datos y acceso

Anexo 4 Cotización equipos para análisis

SYCOM Te ofrece soluciones!

Deseo buscar... ACCEDER CARRITO / L0.00

CONTACTAR +504 2544 0124 MENU OFERTAS ZONA GAMER PC BUILDER SERVICIOS TIENDAS CONTÁCTENOS

INICIO / LAPTOPS

NBR Dell Latitude 5400 i7-8665U/16GB/SSD 256GB/LED 14.0"/W10P 6MG

L\$12,500.00

- 1 + [AÑADIR AL CARRITO](#)

Contáctate con un asesor si tienes alguna duda o sugerencia.

Categoría: Laptops [Chat](#)

Anexo 5 Carta de Autorización de Empresa

San Pedro Sula, Cortes, 25 de Agosto del 2023

Sergio Felipe Jones
Gerente de Recursos Humanos Corporativo
Fruit of the Loom Inc.

Bufalo, Villanueva, Cortes

Estimado Señor (a): **Sergio Felipe Jones**

Reciba un cordial y atento saludo. Por medio de la presente deseamos solicitar su apoyo, dado que somos alumnos de UNITEC y nos encontramos desarrollando el Trabajo Final de Graduación previo a obtener nuestro título de maestría en **ANALITICA DE NEGOCIOS**.

Hemos seleccionado como tema, **Análisis de Series Temporales para la Detección de Anomalías de Eficiencia en la Nómina Diaria en una Planta de Manufactura**, por lo que estaríamos muy agradecidos de contar con el apoyo de la empresa que usted representa para poder desarrollar nuestra investigación. En particular, dicha solicitud se circunscribe a petitionar se nos autorice a realizar consultas a la base de datos de eficiencias de la nómina.

A la espera de su aprobación, me suscribo de usted,

Atentamente,


Gustavo Ernesto Avila Griffin


Sebastián Núñez Flores

No. de cuenta: 22213018

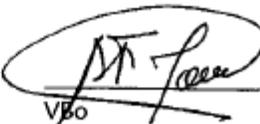
No. de cuenta: 22213065

Por este medio, Fruit of the Loom Inc
(Empresa / Institución)

Autoriza la realización dentro de las condiciones el proyecto de investigación de Postgrado antes mencionado.

Felipe Jones
(Nombre y sello Gerente)




VBo