



UNIVERSIDAD CÉSAR VALLEJO

ESCUELA DE POSGRADO

**PROGRAMA ACADÉMICO DE MAESTRÍA EN INGENIERÍA
DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA
INFORMACIÓN**

**Machine Learning en la Mejora de la Gestión del Conocimiento en
el Consejo Nacional de Ciencia y Tecnología e Innovación
Tecnológica - CONCYTEC, Lima 2021**

TESIS PARA OBTENER EL GRADO ACADÉMICO DE:
Maestro en Ingeniería de Sistemas con Mención en Tecnologías de la Información

AUTOR:

Quipas Bellizza, Mirtha Jesús (ORCID: 0000-0002-2229-0681)

ASESOR:

Dr. Visurraga Agüero, Joel Martin (ORCID: 0000-0002-0024-668X)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LIMA – PERÚ

2021

Dedicatoria

A mis padres por su guía y amor incondicional, que me impulsan cada día a alcanzar mis metas.

Agradecimiento

Agradezco a mi institución laboral por su apoyo en el desarrollo de mi investigación

Índice de contenidos

	Pág.
Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de gráficos y figuras	vi
Resumen	vii
Abstract	viii
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	6
III. METODOLOGÍA	21
3.1. Tipo y diseño de investigación	21
3.2. Variables y operacionalización	22
3.3. Población, muestra y muestreo	23
3.4. Técnicas e instrumentos de recolección de datos	24
3.5. Procedimientos	26
3.6. Método de análisis de datos	26
3.7. Aspectos éticos	27
IV. RESULTADOS	28
V. DISCUSIÓN	39
VI. CONCLUSIONES	47
VII. RECOMENDACIONES	49
REFERENCIAS	50
ANEXOS	57

Índice de tablas

	Pág.
Tabla 1. Matriz de operacionalización de la variable dependiente – Gestión del conocimiento	23
Tabla 2. Población de la investigación	24
Tabla 3. Ficha técnica del instrumento	25
Tabla 4. Expertos que validaron el instrumento de recolección de datos cuantitativos	26
Tabla 5. Medidas descriptivas del indicador 1 tiempo de ejecución	28
Tabla 6. Medidas descriptivas del indicador 2 índice de expedientes procesados	29
Tabla 7. Medidas descriptivas del indicador 3 índice de expedientes procesados en el tiempo previsto	31
Tabla 8. Pruebas de Wilcoxon –Rangos del indicador 1 tiempo de ejecución	33
Tabla 9. Pruebas de Wilcoxon – Estadísticos de contraste indicador 1 tiempo de ejecución.	34
Tabla 10. Pruebas de Wilcoxon – Rangos del indicador 2 índice de expedientes procesados	35
Tabla 11. Pruebas de Wilcoxon – Estadísticos de contraste del indicador 2 índice de expedientes procesados	35
Tabla 12. Pruebas de Wilcoxon – Rangos del indicador 3 índice de expedientes procesados en el tiempo previsto	37
Tabla 13. Pruebas de Wilcoxon – Estadísticos de contraste del indicador 3 índice de expedientes procesados en el tiempo previsto	37

Índice de figuras

	Pág.
Figura 1. Comparación de valores medios del indicador tiempo de ejecución	28
Figura 2. Comparación de valores medios del indicador índice de expedientes procesados	30
Figura 3. Comparación de valores medios del indicador índice de expedientes procesados en el tiempo previsto	31

Resumen

El objetivo de la presente investigación es determinar que machine learning mejora la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021. Para lo cual, se midieron los resultados del proceso de gestión de conocimiento en forma tradicional y haciendo uso de una aplicación machine learning para establecer la mejora a través de indicadores de tiempo, así como de índices de documentos procesados normalmente y en tiempos previstos.

El tipo de investigación empleada fue aplicada, con diseño experimental puro, utilizando una población de 50 observaciones y con un muestreo probabilístico por cada indicador. Para la recolección de datos se utilizó como instrumento de recolección de datos la guía de observación y para la prueba de hipótesis en el análisis inferencial, se utilizó la prueba de Wilcoxon.

Se concluyó que con la implementación de machine learning se mejora significativamente la gestión de conocimiento, teniendo como puntos fuertes de mejora sus indicadores, demostrando que el tiempo de ejecución disminuyó en promedio 63.68%, los expedientes procesados se incrementaron en 25.13% y expedientes procesados en el tiempo previsto se incrementaron en 42.15%, respecto a sus situaciones iniciales.

Palabras clave: aprendizaje automático, gestión del conocimiento, proceso, algoritmos de predicción.

Abstract

The objective of this research is to determine that Machine Learning improves knowledge management in the National Council for Science and Technology and Technological Innovation - CONCYTEC, Lima 2021. For which, the results of the knowledge management process were measured in a traditional way and made use of a Machine Learning application to establish the improvement through time indicators, as well as indexes of documents processed normally and in anticipated times.

The type of research used was applied, with a pure experimental design, using a population of 50 observations and with a probability sampling for each indicator. For data collection, the observation guide was used as a data collection instrument and for the hypothesis test in the inferential analysis, the Wilcoxon test was used.

It was concluded that with the implementation of Machine Learning, it significantly improves knowledge management, having its indicators as strengths of improvement, showing that the execution time decreased by an average of 63.68%, the processed files increased by 25.13% and the files processed in the expected time increased by 42.15%, compared to their initial situations.

Keywords: machine learning, knowledge management, process, prediction algorithms.

I. INTRODUCCIÓN

Actualmente con el acelerado avance tecnológico, a las organizaciones se les presentan oportunidades para que a través del uso de tecnologías de información puedan desarrollar su conocimiento empresarial y lograr incrementar su competitividad en este mundo globalizado que se encuentra en constante crecimiento.

Organizaciones internacionales como la Organización para la Cooperación y Desarrollo Económicos (OCDE), a través de los años ha realizado estudios que demuestran que una gestión de conocimientos puede hacer que una empresa sea productiva, innovadora y por ende sea exitosa. De igual manera, el Programa de las Naciones Unidas para el Desarrollo (PNUD) ha desarrollado y viene desarrollando proyectos basados en la gestión del conocimiento en diferentes países del mundo y en sus evaluaciones se evidencia que los resultados mejoran las organizaciones.

La gestión del conocimiento a través del tiempo ha ido incorporando componentes actuales como innovación y el indudable aporte de nuevos productos que proporcionan las tecnologías emergentes adaptándose a las nuevas realidades, que incluso con fenómenos no advertidos como COVID-19 han impactado positivamente en las aplicaciones digitales, la gestión del conocimiento de las organizaciones y su productividad, al tener que adaptarse rápidamente a nuevas formas de brindar sus servicios. La ICEMD de la ESIC Business & Marketing School publicó que las tecnologías más disruptivas que presentan mayor impacto en la sociedad son inteligencia artificial y cognitiva, machine learning, deep learning, asistentes virtuales, chatbots, reconocimiento de imágenes, reconocimiento de voz entre otros.

Desde una perspectiva del impacto social-técnico Young, Majehrzak y Kane (2021) opina que los incentivos económicos y la innovación están impulsando el

incremento de dependencia de las herramientas de aprendizaje automático más allá de las redes sociales, ya que éstas incursionan en la vida laboral diaria en sus diversas aplicaciones como asistentes virtuales, pilotos automáticos de aviones, vehículos de transporte autónomos, sistemas de monitoreo de fraudes financieros, procesos logísticos predictivos por indicar algunos.

En el Perú la Presidencia del Consejo de Ministros como parte de sus políticas de modernización del estado, tiene a la gestión del conocimiento como un eje transversal y pilar importante para el éxito de la modernización del Estado, así como el marco jurídico que facilita el uso de tecnologías en empresas del sector público. En este sentido, las Entidades del Estado lo incluyen en sus estrategias gubernamentales para el logro de sus objetivos.

Actualmente, debido a los últimos acontecimientos de la pandemia en las diferentes regiones del País se viene experimentando un crecimiento en los servicios digitales tanto privados como gubernamentales. Las aplicaciones de estos servicios se han realizado por grandes inversiones en tecnología que permiten mantener la operatividad, salvaguardar la salud y plantearse mejores estrategias para gestionar la información y tomar decisiones.

El Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica (Concytec) de acuerdo a la política gubernamental del país, tiene en sus principales instrumentos de gestión, componentes de gestión del conocimiento que le permiten el cumplimiento de sus objetivos. Desde el inicio de la Pandemia la mayoría del personal viene realizando labores de tele-trabajo, trabajando permanentemente en ofrecer mejores y mayores servicios digitales al ciudadano. Sin embargo, existen actividades especializadas al interno que toman mucho tiempo y esfuerzo de realizar y cuyo resultado es conocimiento importante para la toma de decisiones.

En Concytec se realizan actividades de cálculos especializados, simulaciones y predicciones tanto para generar instrumentos de gestión como para

para la toma de decisiones, y generalmente por las características de éste y no estar completamente automatizada, su elaboración toma mucho tiempo y esfuerzo realizarlo, debido al alto volumen de información que maneja y a la naturaleza del trabajo que realiza como ente rector. Por lo que se propone hacer uso de la tecnología machine learning para que le ayude en su gestión.

En este sentido, la presente investigación responderá a la pregunta: ¿De qué manera machine learning mejora la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?.

En cuanto a los problemas específicos, se indican a continuación: (a) ¿De qué manera machine learning mejora el tiempo de ejecución de la gestión del conocimiento, en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?, (b) ¿De qué manera machine learning mejora el índice de expedientes procesados de la gestión de conocimiento, en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?, (c) ¿De qué manera machine learning mejora el índice de expedientes procesados en el tiempo previsto de la gestión de conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?.

La presente investigación se justifica porque a través del proyecto se obtiene un modelo de trabajo utilizando tecnologías actuales como es machine learning aplicada a sistemas de cobertura nacional para empresas del estado, que facilita responder a nuevos tiempos, mejorar la eficiencia y adecuarse a las nuevas exigencias del ciudadano y del entorno. Asimismo, la justificación epistemológica se basa en el enfoque de racionalismo de la corriente de Buffón, quien veía al hombre como un ser con capacidad de pensar y formar conocimiento (Arellano, 2014).

En el campo teórico la investigación se justifica porque permite generar conocimientos de los fundamentos de la tecnología machine learning y de la gestión del conocimiento, de forma tal que facilita su entendimiento para posteriores aplicaciones en otros sectores del gobierno y/o empresas privadas; y en el ámbito práctico el resultado de la investigación permite generar un conocimiento valioso al interno de la organización y adicionalmente podrá utilizarse para otras finalidades que permitan beneficiar a más personas de la comunidad científica.

La presente investigación utilizó el método experimental, con pre y post test en sus mediciones, y hace uso de instrumentos de recolección de datos confiables y validados, que permiten la obtención de datos y posterior generación de resultados, que podrán servir de base para futuros estudios.

Por otro lado, tiene como propósito alcanzar ciertos objetivos, los cuales se encuentran planteados a partir del problema. En este sentido, se formuló el siguiente objetivo general: Determinar que machine learning mejora la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021. Asimismo, los objetivos específicos son los siguientes: (a) Determinar que machine learning mejora el tiempo de ejecución de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021, (b) Determinar que machine learning mejora el índice de expedientes procesados de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021, (c) Determinar que machine learning mejora el índice de expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

De igual modo, se realizó suposiciones sobre los resultados a obtener de la investigación. La hipótesis general planteada fue la siguiente: Machine learning

mejora significativamente la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica –CONCYTEC, Lima 2021.

Respecto a las hipótesis específicas, tenemos las siguientes: (a) Machine learning mejora significativamente el tiempo de ejecución de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica –CONCYTEC, Lima 2021, (b) Machine learning mejora significativamente el Índice de expedientes procesados de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica –CONCYTEC, Lima 2021, (c) Machine learning mejora significativamente el índice de expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica –CONCYTEC, Lima 2021.

II. MARCO TEÓRICO.

Para efectos de sustentar la presente investigación, en el ámbito nacional se tienen los siguientes estudios previos:

Neyra (2021) realizó una investigación que tuvo como objetivo desarrollar métodos de predicción basados en machine learning, para estimar el contenido de cadmio en granos de cacao, a partir de sus firmas hiperespectrales en el proceso productivo del cacao para su detección en tiempo real y con ello reducir el riesgo de contaminación por metales pesados como el cadmio en el producto final. El estudio se realizó en el departamento de Automática y Control de la Universidad de Piura, tomando datos de muestras de cacao de algunas locaciones de la región, se obtuvieron resultados de diversos algoritmos de predicción de machine learning permitiendo comparar los modelos para ese caso y con un monitoreo en tiempo real, se tomaron medidas correctivas logrando la mitigación de contaminación por metales pesados como el cadmio. Concluyó que las imágenes multiespectrales para uso de la industria agrícola se complementan muy bien con la aplicación de machine learning como estrategia en la detección de cadmio en el cacao.

Gamarra (2020) en su investigación, tuvo como propósito la elaboración de modelos de predicción machine learning relacionados con la mortalidad perinatal para la Dirección Ejecutiva de epidemiología de la DIRESA Junín, con la finalidad que los funcionarios puedan tomar decisiones de prevención y evitar eventos no deseados. Para ello se utilizó datos históricos de muertes prenatales y luego de la fase de recolección de datos, se procesaron para entrenamiento y prueba aplicando a varios modelos predictivos. Resultado de la evaluación se determinó que el algoritmo W-J48 fue quien presentó el mejor modelo de predicción relacionado con el número de pacientes con muerte prenatal en la región Junín, obteniendo métricas de precisión de 88.09%, sensibilidad de 78.53% y precisión de clase de 97.11%.

Por otro lado, López (2019) tuvo como objeto de su investigación, la comparación e identificación del mejor modelo de aprendizaje automático para predecir el incumplimiento de pagos en el sector bancario, específicamente en instituciones financieras. Se aplicaron varios modelos de aprendizaje automático haciendo uso de los datos proporcionados por una empresa del sector financiero del Perú, logrando realizar las predicciones utilizando el modelo XGBoost, con una exactitud del 97.53% y un F1-Score de 0.1278. A través de éste modelo se puede reducir el tiempo en la identificación de mal pagadores y buenos pagadores el porcentaje de exactitud indicado anteriormente.

Asimismo, Medina (2018) en su tesis de maestría, realizó una investigación relacionada con la detección de depresión en alumnos aplicando un análisis textual y el modelo predictivo basado en aprendizaje automático. La investigación permitió identificar patrones personalizados de personas con depresión utilizando técnicas TF-IDF en una muestra de 596 alumnos de la Universidad César Vallejo, utilizando Microsoft Azure machine learning con R, como herramienta para implementar el modelo de predicción, el cual proporciona resultados con una alta probabilidad de 91.7% y con resultados reportados en menos de dos horas. Se comprobó que con la aplicación web basado en análisis textual y aprendizaje automático el tiempo de atención para pacientes con depresión considerable disminuye siendo oportuno y permitiendo apoyar a la disminución de casos de suicidio en cualquier parte del mundo.

En cuanto a los antecedentes internacionales, se tienen los siguientes: Zhao, Zhang, Liu, Wang, Zhu *et al.* (2021) efectuaron una investigación en Beijing-China, exponiendo los algoritmos y técnicas del aprendizaje automático para la exploración y evaluación de biomasa de peces, su identificación y clasificación, así como el análisis de comportamiento y predicción de parámetros de calidad del agua, destacando los problemas actuales en el sector y su tendencia de desarrollo. Se describieron y utilizaron métodos para aplicación del aprendizaje automático comparado con el aprendizaje tradicional, logrando identificar que el aprendizaje

automático mejoró la eficiencia en beneficio de la cría, identificando algunas deficiencias en el aprendizaje automático como la complejidad e incompletitud de los conjuntos de datos, baja calidad de imágenes, y costos altos en la implementación técnica. Asimismo, se observó que los factores externos como el clima, la geografía, la composición biológica de los ambientes de acuicultura limitan el aprendizaje automático; como parte del análisis del comportamiento y predicción de la calidad del agua, se crearon modelos de datos y una base de datos pesquera para mejorar la transparencia de la información. El autor concluye que el aprendizaje automático puede ser aplicado en la detección y clasificación de peces, posibilitando la prevención de enfermedades, análisis de comportamiento e investigación ambiental en el sector de la acuicultura, así como que se está convirtiendo en un elemento vital para la acuicultura y que en el futuro debe superar sus limitaciones.

Ashok y Pellakuri (2020) realizaron su investigación en Solapur India, donde implementaron algoritmos predictivos basados en machine learning que permiten predecir el riesgo de salud de niños menores a 5 años, ya que son una población vulnerable y requieren tratamientos tempranos. Con estas predicciones se puede evitar muertes y realizar programas de tratamientos tempranos. Adicionalmente, como parte de la investigación se abordan temas relacionados con las formas de aprendizaje automático existentes, sus tipos y atributos, así como sus necesidades, haciendo énfasis en la comparación de los algoritmos K-mean y el algoritmo K-nearest neighbour, resaltando que el último es más sencillo de clasificar.

Li, Yang, Wang y Wu (2018) en su investigación aplicó modelos de aprendizaje automático para predecir como las inversiones de empresas en I + D influyen en el proceso de innovación en el futuro, derivados de estrategias de innovación actuales. Para realizar la investigación se aplicó y recomendó el uso del método XGBoost para sus estimaciones, tanto por su precisión como por la interpretación de datos en comparación a árbol de decisiones, los bosques aleatorios y la red neuronal; asimismo, para la muestra utilizaron datos completos

de las empresas que pertenecieron al Parque Científico Zhongguancun – ZGC en China, de los años 2005 hasta el 2015. Este modelo de aprendizaje automático puede aplicarse a cualquier tamaño de empresa y de diferentes clústeres, para predecir fácilmente su desempeño en innovación, así como para realizar evaluaciones de proyectos que ayuden en la toma de decisiones para la elección de financiación limitada y para investigar sobre los efectos de nuevas estrategias en el desempeño de la empresa.

Knoll, Prügmeier y Reinhart (2016) efectuaron su investigación en la Universidad de Múnich Alemania con el propósito proporcionar un enfoque que permita predecir escenarios de la logística de entrada, mediante el uso aprendizaje automático. En este enfoque se propone para apoyar principalmente a las actividades de planificación logísticas y hace una revisión técnica detallada de lineamientos y procedimiento a seguir. Propone dividir el trabajo en etapa de configuración y creación de una ontología logística, mediante identificación y relaciones relevantes, modelar el sistema, evaluación de capacidad para implementar modelos, identificar los algoritmos para machine learning y evaluar modelos con datos para reducir los esfuerzos en proceso de planificación logístico.

En relación a teorías tomadas como base para la presente investigación, se ha considerado a la teoría del conectivismo y la teoría institucional. Según Torres y Barnabé (2020) la teoría de conectivismo emerge por la presencia de las nuevas tecnologías y nuevas formas de comunicación en el proceso de aprendizaje de la era digital, y presenta posibilidades de cambios para los enfoques pedagógicos, currículo y paradigmas metodológicos. Considera que el conectivismo tiene como base la interacción del sujeto con otro medio generando una cultura mediática y digital. Además, indica que el aprendizaje se encuentra inmerso en un contexto sistémico con nuevos conceptos como ecologías de aprendizaje y ecosistemas de aprendizaje.

Por su parte, Montoya, Parra, Arias, Cabello y Coloma (2019) realizaron un análisis sobre la teoría del conectivismo promovida por George Siemens y Stephen Downes, quienes consideran que se trata de una teoría de aprendizaje aplicada a la actual era digital, que sustenta, que el aprendizaje se puede dar en cualquier lugar. Definen el conocimiento como un patrón especial de relaciones y aprendizaje, así como la habilidad para poder maniobrar en el entorno las redes o patrones existentes, tratando los principios de aprendizaje a nivel biológico-neuronal, conceptual y social-externo. Consideran que el conectivismo inicia en la persona y posteriormente su conocimiento interno se desarrolla de una red, que transmite información a organizaciones y luego se retroalimentan de información dentro de la misma red, para proveer un nuevo aprendizaje a la persona; esto permite que el profesor para los alumnos ya no será una fuente de conocimientos, progresivamente se convertirá en un guía, en un referente para generar en los alumnos autonomía y emancipación. La teoría de conectivismo destaca cuatro conceptos que proporcionan una explicación sobre el aprendizaje en un entorno social mediado por tecnologías de información y comunicaciones, estos son la apertura, la diversidad, la interacción y la autonomía. El estudio del aprendizaje autónomo se basa los fundamentos de ésta teoría.

Sánchez (2019) basado en las ideas de Dowes y Siemens, indica que para el conectivismo el cerebro y la tecnología son quienes proponen los contenidos y funciones mentales y que el aprendizaje puede almacenarse en entes no humanos, idea cuestionada por algunos autores que consideran a la tecnología como un agente activo del proceso de aprendizaje. Considera que internamente las redes de aprendizaje se perciben como estructuras que viven en la mente, que conectan y crean patrones de comprensión, ya que estas conexiones se adaptan en el cerebro para realizar procesamientos en diversos entornos como el de la tecnología, evidenciando una relación entre la operatividad del cerebro y la organización, así como las herramientas y lo que se aprende, resaltando las habilidades que se desarrolla para adquisición y almacenamiento de la información en estructuras de redes y dispositivos informáticos.

En el proceso de aprendizaje plantea los siguientes pasos: contexto (ubicación de fuentes de información en la red), relevancia (patrones recurrentes que activan otras conexiones), aparición (activación del patrón relevante), memoria (persistencia en aparición de patrones de conectividad), que hacen que el aprendizaje se presente como un proceso que se centra en el reconocimiento de patrones dentro de una red.

Para Cabero y Lorente (2015), las teorías tradicionales que han aportado mayor explicación al fenómeno educativo del aprendizaje son: conductismo, cognitvismo y constructivismo. Sin embargo, se presentan como parte de la evolución humana nuevos escenarios formativos y una de las más significativas es la teoría conectivista, que nos explica cómo se produce el aprendizaje en entornos tecnológicos caracterizando. Su teoría se basa en los enfoques de Lepi, Prensky y Siemens. Resalta el tema que el conocimiento se relaciona básicamente con la capacidad de conectarlo con sus saberes previos y desde la perspectiva de aprendizaje se considera como un proceso que conecta nodos especializados o llamados también recursos de información. Por eso se propicia que el alumno realice sus propias conexiones investigando en internet, teniendo como objetivo abrir sus mentes no llenarlas. Asimismo, menciona que Cochrane considera que el aporte de la teoría del conectivismo explica claramente lo relacionado con el acceso al conocimiento distribuido, pero falta profundizar en cómo se aprende. Al respecto, considero que los nuevos aportes a ésta teoría podrían ir consolidando el conocimiento acerca del aprendizaje en nuestra nueva realidad facilitando su aplicación en todos los niveles de la educación.

Asimismo, Calvani (2008) como parte de la teoría conectivista identifica características para la educación: la estructura de presentación de información requiere procesamiento, se requiere hacer uso de aplicaciones web y servicios web como blogs, wikis, etc, es transformable dentro de su ecosistema, con tendencias a generar flujos de información y participación en la red, el flujo de información y base de conocimiento se distribuye, los grupos de trabajo se dan por intereses

personales y son espontáneos en comparación con las clases tradicionales, el currículo debe ser negociado en conjunto. También hace mención la posición de Verhagen, Kop y Hill que se consideran a esta teoría como una corriente educativa o una propuesta a una nueva realidad, esto por considerar que aún está presente lo que conocemos como escuela actual. Sin embargo, vemos que ésta teoría es la base del futuro educacional y seguirá transformándose a través del tiempo.

Sobre la teoría institucional, Falaster, Zanin y Guerrazzi (2017) señalan que la teoría institucional ha sido clave en investigaciones como ciencias sociales y destaca el elemento de isomorfismo para explicar elementos comunicacionales y de estrategias. Cabe señalar que los autores consideran que a través de la teoría institucional se puede explicar las respuestas que se dan ante las presiones institucionales, precisando que las instituciones son elementos que aseguran estabilidad, que generan expectativas en las futuras acciones y comportamientos en el tiempo. Asimismo, indican que Scott en su enfoque sobre la teoría de institucional, refiere que existe una evolución en la propia teoría, pero esto es válido para una conceptualización inicial, ya que considera que las acciones pasadas podrían generar en la vida diaria normas y reglas, sean estas formales o no, pero generan nuevos patrones de futuro.

Jardim y Bronzo (2016) en su análisis sobre la teoría institucional toma información de los principales exponentes de ésta Teoría, para entender su aplicación en estudios en organizaciones agroindustriales. Nos indica que el enfoque de North es macro desarrollador identificando el origen, la estructuración y los cambios institucionales, y distingue dos agentes económicos las instituciones formales e informales ambos con características particulares construidas por la sociedad. Para North las Instituciones se consideran como reglas de juego y las organizaciones como sus jugadores, tratando de reducir la incertidumbre. Por otro lado, Williamson presenta un enfoque micro institucional ya que su estudio se dirige a la explicación de diversos arreglos institucionales. También destaca el enfoque de Zylbersztajn que se refiere a las organizaciones como entes dependientes del

aparato institucional para establecerse, es decir que existe dependencia de su entorno institucional. De acuerdo a ello podemos destacar que no es lo mismo una organización que una institución, las organizaciones como las que pertenecen al Estado, que tienen marcos jurídicos, políticas, aspectos culturales entre otros dan las pautas necesarias para que se asegure un adecuado desempeño en su organización, la cual se afecta permanente por factores del entorno e internos.

Scott (2014) aporta a la Teoría institucional en la identificación de tres pilares dentro de las instituciones: a) un sistema regulativo, que se componen de leyes y reglas y de acuerdo a su cumplimiento se generan recompensas o no entre los actores del sistema; b) un sistema normativo, que genera estándares a seguir convirtiéndose en fuente de legitimidad para el grupo ya que los asocia y comparte intereses comunes; c) una sistema cultural-cognitivo, que orientan la vida social del grupo. En ese sentido, podemos decir que las instituciones se componen por elementos simbólicos, recursos materiales y actividades sociales.

Restrepo y Rosero (2002) consideran que la característica principal de la teoría institucional es su enfoque social, alejándose del aspecto teórico de las organizaciones. La atención de la teoría está centrada en el estudio de la institución como una estructura social integrada por un grupo de individuos u organizaciones que ejercen colectivamente acciones para restringir un ambiente que cambiará en el tiempo. Se plantea que las instituciones son entes con valor y significado, con miembros legítimos y líderes, que reflejan los mitos de sus ambientes institucionales y para evitar la ilegitimidad institucional, la sociedad induce prácticas y comportamientos racionales y adecuados para sobrevivir con éxito y como construcciones sociales requieren que sus actores interactúen permanentemente con el objetivo de mantener estable a la institución. Considera que las necesidades de cambio se originan del contexto y las respuestas de las instituciones son el producto de la inmersión organizativa de dicho entorno.

Otro tema a tratar en la presente investigación, es sobre machine learning o aprendizaje automático. Para Zhao, Zhang, Liu, Zhu *et al.* (2021) el aprendizaje automático, permite que las máquinas adquieran nuevas habilidades, nuevos conocimientos y puedan reorganizar los conocimientos ya existentes, en forma similar a un proceso de aprendizaje, para establecer y describir estructuras de patrones en un universo de datos proporcionados utilizando como núcleo los algoritmos para realizar el análisis de datos y tomar decisiones sobre los nuevos. En términos generales, se deben ingresar masivamente datos para entrenar al modelo, de tal forma que éste pueda comprender las leyes potenciales que se encuentran en los datos y lograr clasificarlos o predecir con precisión los datos recién ingresados.

Según Jordan y Mitchell (2015) el aprendizaje automático está centrado en resolver problemas sobre el aprendizaje de datos y algoritmos, para crear modelos matemáticos que mejoren el rendimiento de los sistemas informáticos. También refiere que el desarrollo de éstos dos elementos (aprendizaje de datos y algoritmos) y la explosión de datos en línea, así como la computación de bajo costo han impulsado el avance del aprendizaje automático. También indican, que es uno de los campos técnicos que actualmente se encuentra en rápido crecimiento y que se relaciona con la informática, estadística, el núcleo de la inteligencia artificial, así como con la ciencia de datos.

Krishna, Amirali, Swathi y Madhavi (2021), consideran que el aprendizaje automático es una técnica que forma parte de la inteligencia artificial (IA), que facilita el aprendizaje de los programas y su desarrollo sin programación directa, donde las instrucciones de las computadoras trabajan principalmente para crear nuevos programas de computadoras y que utilizan el conocimiento para su uso, en otras palabras el objetivo principal del aprendizaje automático es hacer que las computadoras puedan aprender y que alteren su comportamiento sin la ayuda o interferencia del ser humano. Uzunoma, Khan, Venkatesan y Guillard (2021) señalan que el proceso de aprendizaje inicia con una evaluación y la evidencia,

como ejemplos, conocimiento directo o instrucción para evaluar patrones de datos y tomar decisiones inteligentes basadas en ejemplos en el futuro para permitir que las computadoras aprendan y modifiquen automáticamente el comportamiento sin ayuda o interferencia humana. Arias (2020) confirma que el uso de tecnologías impacta al proceso de aprendizaje.

Para Mora (2018); Arunakranthi, Rajkumar, Chandra y Harshavardhan (2020) el aprendizaje automático enfocado en un uso práctico, se refiere a la utilización de grandes volúmenes de datos para aprender de ellos y posteriormente hacer predicciones o sugerencias sobre algo; requiere gran disponibilidad de datos para alimentar a los motores de algoritmos y es considerado como una de las áreas de investigación más destacadas en el descubrimiento de conocimientos y en la identificación de patrones ocultos de los conjuntos de datos.

Según Murali, Barkat, Arelli y Madhavi (2021) el aprendizaje automático es una técnica de la inteligencia artificial (IA), que permite que la máquina tenga capacidad de pensar y aprender sin programación explícita. Su aprendizaje se centra en la creación de programas de computadoras que observan y utilizan el conocimiento que generan para uso propio. Considera que el aprendizaje automático es uno de los inventos más prometedores jamás conocidos.

Adicionalmente Vázquez-Marrufo, Sarrias-Arrabala, García-Torres, Martín-Clemente y Izquierdo (2020) indican que Machine Learning presenta características que se pueden medir como la precisión, sensibilidad y especificidad, sin embargo, no todos los estudios aplican estos tal cual pues se tiende a crear sus propias mediciones de acuerdo al contexto.

Robert (2014) define el aprendizaje automático como un conjunto de métodos que detectan patrones en los datos en forma automática y luego se utilizan para predecir datos futuros o para ayudar a la toma de decisiones con baja incertidumbre. Al respecto Wu (2020) realiza una interesante reflexión sobre la

aplicación y potencial que tiene el aprendizaje automático o machine learning (ML) especialmente en sectores críticos con alto nivel de riesgo como el sector financiero. Considera que la adopción de técnicas Machine Learning mejora la gestión de riesgos y ofrece un gran potencial en este campo, ya que permite analizar y explorar datos masivamente e identificar intrincados patrones no lineales que facilitan la construcción de un modelo de riesgos y convirtiéndose actualmente en una herramienta importante para el sector empresarial.

Finalizando este bloque de ideas, Uhunoma, Khan, Venkatesan y Guillard (2021) indican que, aunque el aprendizaje automático y la inteligencia artificial tienen usos diferentes, el aprendizaje automático es un subconjunto de la inteligencia artificial, que utiliza algoritmos o procedimientos por parte de un sistema artificial para aprender de los datos y puede ser un aprendizaje supervisado (se reciben datos de entrada y salida para relacionarlos y predecir resultados), aprendizaje no supervisado (se reciben datos sin precisar su tipo intenta predecir su relación) y aprendizaje reforzado (aplica el descubrimiento en base a la experiencia). En esta investigación, se trabajó con un aprendizaje supervisado, ya que, de acuerdo a Quispe, Celi y Campos (2020); Zhao, Li, Zuo y Zhang (2021) la aplicación de este tipo de aprendizaje es recomendable cuando se cuenta con un conjunto de datos de entrada que se alimentará por sí mismo para obtener las mejoras en los procesos empresariales. Asimismo, señala que en la creación de cualquier modelo de machine learning se utilizan los siguientes pasos: carga de datos, análisis y procesamiento de los mismos.

En cuanto a la variable dependiente esta se refiere al proceso de gestión de conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica, en donde de acuerdo al manual de organización y funciones la Dirección de Evaluación y Gestión del Conocimiento del Concytec, es el órgano encargado entre otras funciones, del desarrollo y operación del proceso de seguimiento y evaluación de los planes, programas y proyectos de CTI del Sistema Nacional de Ciencia y Tecnología, así como también de generar los procesos que

permitan crear, compartir y hacer uso del conocimiento en el Sistema Nacional de Ciencia y Tecnología.

Dalmarco, Maehler, Trevisan y Mortari (2017) describen la gestión del conocimiento como una práctica dentro de la empresa liderada por un individuo responsable de su gestión. Asimismo, señalan que la teoría de las organizaciones del conocimiento consiste en formular e implementar estrategias para la construcción, incorporación, distribución y utilización del conocimiento.

Ode y Ayavoo (2019); OCDE (2014) indican que la gestión del conocimiento (GC), abarca un proceso llamado práctica internacional y sistemática que adquiere, capta, comparte y produce conocimiento, que se utiliza para describir los procesos dentro y fuera de la organización, de tal manera que puedan mejorar el rendimiento de las misma y conducir al logro de los objetivos organizacionales.

Según Costa, Johann, Passeri y Scott (2017) en el siglo XXI, se considera al conocimiento como el principal factor de producción en las organizaciones, el cual se encuentra afectado por un gran volumen de información a través de diversas fuentes, en comparación con la etapa de la sociedad industrial. Indica que las organizaciones deben adaptarse para sobrevivir con éxito en este mundo competitivo, a través de un enfoque más auto organizado con formación de equipos multidisciplinarios y con procesos que realicen esfuerzos colaborativos que traspasen las fronteras de las áreas funcionales.

Para el autor, una gestión del conocimiento que trabaja en un entorno de red es un proceso que organiza y mantiene la colaboración en forma eficiente, por lo que es importante tener en cuenta el propósito de acceder a ideas, tecnologías y compartir información, experiencias, conocimientos técnicos y comerciales entre las organizaciones. Para capitalizar el conocimiento en el contexto de su creación, validación, presentación y aplicación en organizaciones, debe haber una interacción entre tecnología, técnica y personas ya que la acción conjunta crea un

entorno de aprendizaje en el hacer, que permite a la organización gestionar eficazmente su conocimiento, así como crear y mantener en el tiempo las ventajas competitivas. Desde una perspectiva operativa considera que la gestión de conocimiento combina el conocimiento (explícito) y know how (tácito) en los procesos y productos de la organización para la creación de valor, por lo que necesario combinar las habilidades y el conocimiento, así como implantar una cultura de buenas prácticas en la organización, así como ampliar las redes de relaciones y mejorar los activos intelectuales. Al respecto, es importante mencionar que las herramientas tecnológicas se deben aprovechar para sustentar los procesos de gestión con la finalidad de iniciar la migración vivencial en forma ordenada de actividades manuales a electrónicas o digitales (Barrera, 2018).

Ansari (2019) desde una perspectiva estratégica y basado en la definición de North y Majer, indica que en ésta era de rápida innovación y permanentes cambios tecnológicos la gestión del conocimiento en una economía digital es un habilitador clave para la creación del valor y esto se logra haciendo uso de tecnologías digitales para la creación e intercambio del conocimiento en una sociedad digital.

Asimismo, Córdova, Durán, Pinchera, Palominos y Galindo (2021) opinan que actualmente existe mucha presión de los altos ejecutivos para obtener resultados en el corto plazo, lo que se refleja en sus indicadores de desempeño, obligando a realizar acciones para mejorar sus procesos, desarrollar sus recursos humanos y sus tecnologías de información. Por consiguiente, las organizaciones necesitan introducir globalmente productos y servicios innovadores acompañados de procesos eficaces y ágiles.

En este contexto, Beltrán (2017) indica que los indicadores de gestión representan aspectos críticos de la organización y de su monitoreo, ya que le permite identificar ciertas condiciones y síntomas que se derivan de un normal

comportamiento de sus actividades, convirtiéndose en una herramienta clave para el logro de la competitividad organizacional y personal.

Para la presente investigación se consideró que la variable gestión del conocimiento se mida por los siguientes indicadores: (a) Tiempo de ejecución, (b) Índice de expedientes procesados, (c) Índice de expedientes procesados en el tiempo previsto.

En este sentido, Álvarez y Lobo (2020) señalan que el tiempo de ejecución corresponde al tiempo que se emplea para realizar diversas actividades que forman parte del proceso de gestión de conocimiento, enmarcados dentro de un periodo final y un periodo anterior o inicial, es decir desde el registro expediente hasta su calificación. Adicionalmente, Beltrán (2017) precisa que el tiempo de ejecución mide la rapidez del proceso de Gestión dentro de sus límites es decir de inicio a fin.

Según la RP N° 007 - 2011 - CONCYTEC-P (2011) el tiempo de ejecución es aquel tiempo que se da cuando se recepciona y registra en el sistema informático los documentos o expedientes provenientes de usuarios externos a la organización, el cual genera un número de registro, fecha y hora de inicio del proceso hasta la etapa de cierre donde el documento de respuesta resultado del proceso es enviado a dicho usuario.

Respecto al Índice de expedientes procesados, la RDE No.38-2019-MEF (2018), indica que los expedientes procesados son un conjunto de documentos electrónicos debidamente numerados y generados en un sistema informático de la institución, que se encuentran asociados a un tema o asunto específico, que permite dar inicio a la atención y pasar por varias etapas dentro del proceso hasta que finaliza cuando el área haya cumplido con el propósito de la atención y comunique el resultado al usuario.

Para la Secretaría de Gobierno Digital, en su RSEGDI No.001-2017-PCM/SEGDI (2017) se indica que un expediente procesado es aquel expediente digitalizado que transita a lo largo del ciclo vital del expediente, desde su recepción hasta su disposición final, conforme a la tecnología, procedimientos informáticos y normatividad vigente que forman parte de su proceso.

ISO 15489-1 (2016) se refiere a los expedientes procesados como la documentación trabajada que pasa por una o más secuencias de acciones necesarias para producir un resultado como son los expedientes procesados los cuales cumplen con las reglas de gobierno en respuesta a una atención requerida.

Según Calvo, Pelegrin y Gil (2018) el índice de expedientes procesados en el tiempo previsto, es una herramienta importante en el análisis y control de cumplimiento de metas, ya que permite medir el grado de relación que existe entre los expedientes procesados (resultado de procesamiento en varias etapas) y la meta esperada en un tiempo determinado.

Para Beltrán (2017) el índice de expedientes procesados en el tiempo previsto refleja el grado de cumplimiento de los objetivos planteados con anterioridad, en un tiempo determinado, en donde se realizan las actividades necesarias para el procesamiento de expedientes que permitan generar un resultado o atención.

OIM (2016) precisa que el índice de expedientes procesados en el tiempo es un valor que ayuda a medir objetivamente, en un espacio temporal la evolución de actividades relacionadas con el procesamiento de expedientes regulados por la institución, para el cumplimiento de un compromiso adquirido con anterioridad como protocolo o estándar.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

Tipo de investigación

La presente investigación es del tipo aplicada, de acuerdo con Hernández, Fernández y Baptista (2014), la investigación aplicada es aquella investigación que se realiza con el propósito de resolver problemas reales de la vida. Según Thomas y Manz (2017) la investigación aplicada permite comprender qué tan bien se utiliza el conocimiento para diseñar un sistema que resuelva problemas urgentes y genere resultados predecibles.

Diseño de investigación

El diseño utilizado en la investigación es experimental del tipo puro y de acuerdo a Hernández, Fernández y Baptista (2014) el diseño experimental puro se refiere al estudio donde una o varias variables independientes se manipulan intencionalmente para analizar su efecto sobre una o varias variables dependientes en una situación controlada. Asimismo, se trabajó con grupos de comparación y equivalencia entre ellos. El diagrama del diseño que se utilizó es el siguiente:

RG: O1→X→O2

Pre-test→ Machine Learning →Post-Test

R=Asignación al azar

G=Grupo Experimental

X=Tratamiento

O1-O2=mediciones pre-test/ post-test del diseño

3.2. Variables y operacionalización

Definición conceptual de la variable independiente: Machine learning

Krishna, Amirali, Swathi y Madhavi (2021) indican que machine learning o aprendizaje automático es una técnica que forma parte de la inteligencia artificial (IA), que tiene como objetivo facilitar el aprendizaje de los programas y su desarrollo sin programación directa, donde las instrucciones de las computadoras trabajan principalmente para crear nuevos programas de computadoras y que utilizan el conocimiento para su uso, en otras palabras el objetivo principal del aprendizaje automático es hacer que las computadoras puedan aprender y que alteren su comportamiento sin la ayuda o interferencia del ser humano.

Definición conceptual de la variable dependiente: Gestión del conocimiento

Según OCDE (2014) La gestión del conocimiento abarca un proceso llamado práctica internacional y sistemática, que adquiere, capta, comparte y produce conocimiento con la finalidad de mejorar el aprendizaje y rendimiento de las organizaciones.

Definición operacional de la variable dependiente: Gestión del conocimiento

Para el diseño de la Gestión de Conocimiento se ha considerado tres indicadores de medida: (a) Tiempo de ejecución, (b) Índice de expedientes procesados, (c) Índice de expedientes procesados en el tiempo previsto. Siendo la unidad de medida para el primer indicador el tiempo y para los otros dos el porcentaje. Para los tres indicadores se utilizó la guía de observación como instrumento de recolección de datos.

Tabla 1

Matriz de operacionalización de la variable dependiente – Gestión del conocimiento.

Indicador	Instrumento	Frecuencia de toma	U.M.	Fórmula
Tiempo de ejecución	Guía de observación	10 veces por semana	Tiempo (en minutos)	$x = \text{Tiempo Final} - \text{Tiempo de Inicio}$
Índice de expedientes procesados	Guía de observación	10 veces por semana	Porcentaje (%)	$x = \frac{\text{Expedientes procesados}}{\text{Expedientes pendientes de atención}} * 100$
Índice de expedientes procesados en el tiempo previsto.	Guía de observación	10 veces por semana	Porcentaje (%)	$x = \frac{\text{Expedientes procesados en el tiempo previsto}}{\text{Expedientes pendientes de atención}} * 100$

La matriz de operacionalización de la variable gestión del conocimiento se muestra en el Anexo 2.

3.3. Población, muestra y muestreo

Población

Hernández Fernández y Baptista (2014) indica que la población es un conjunto de elementos, con características iguales, que forman parte de la investigación. Para este estudio de investigación se ha considerado como población el número de datos observados, para lo cual se realizó con 50 observaciones para cada indicador.

El detalle de la población de la investigación se presenta en la siguiente tabla:

Tabla 2

Población de la investigación

Población	Cantidad		Indicador
	Pre-test	Post-test	
Observaciones	50	50	Tiempo de ejecución
Observaciones	50	50	Índice de expedientes procesados
Observaciones	50	50	Índice de expedientes procesados en el tiempo previsto

Muestreo

Para la presente investigación se utilizará un tipo de muestreo probabilístico, de acuerdo con Hernández, Fernández y Baptista (2014), en el muestreo probabilístico permite elegir elementos físicos de la población, seleccionarlos en una pequeña parte de la muestra a investigar, considerando que toda la población (expedientes) posee igual probabilidad de ser escogidos para la muestra; se obtiene estableciendo en la población sus propiedades o características y el tamaño de muestra. La técnica que se utilizó fue muestro aleatorio simple sin reemplazo basado en números aleatorios generados por el software Excel.

3.4. Técnicas e instrumentos de recolección de datos

Técnicas de recolección de datos

Según Hernández, Fernández y Baptista (2014). Para la presente investigación se utilizó la técnica de recolección de datos la observación, la cual facilita la recolección de la información sobre el comportamiento y procesos observables para nuestra investigación de forma ordenada y confiable.

Instrumentos de recolección de datos

Para la investigación se hizo uso como instrumento de recolección de datos la Guía de observación de campo, con la cual se realizarán las mediciones en el

pre-test y post-test. Según Hernández, Fernández y Baptista (2014) los instrumentos de medición de recolección de datos son recursos que permiten registrar los datos de la realidad observable.

Tabla 3

Ficha técnica del instrumento

Nombre del instrumento	Guía de observaciones de medición del indicador
Autor:	Mirtha Jesús Quipas Bellizza
Año:	2021
Descripción	
Tipo de instrumento:	Guía de observación de campo
Objetivo:	Determinar que Machine Learning mejora la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.
Indicadores:	a) Tiempo de Ejecución. b) Índice de expedientes procesados c) Índice de expedientes procesados en el tiempo previsto
Número de observaciones a recolectar:	50
Aplicación:	Directa

Validez

La validez del instrumento de recolección de datos, de la presente investigación se determinó a través de juicio de expertos, considerando tres profesionales relacionados con la temática. Hernández, Fernández y Baptista (2014) indica que la validez se refiere al grado en que un instrumento realiza la medición de una variable que intenta demostrar.

En la siguiente tabla mostramos el detalle de los expertos que participaron en la validación:

Tabla 4.

Expertos que validaron el instrumento de recolección de datos cuantitativos.

DNI	Grado Académico Apellidos y nombres	Institución donde labora	Calificación
10271988	Doctor Del Carpio Salinas Jorge Alberto	Universidad de Ciencias y Humanidades	Aplicable
42248986	Magister Pérez Grández Violeta	Concytec	Aplicable
10192315	Doctor Visurraga Agüero Joel Martín	Universidad César Vallejo	Aplicable

3.5. Procedimientos

En la presente investigación se siguieron las siguientes etapas: primero se elaboró los instrumentos de recolección de datos, luego se determinó la validez de los instrumentos a través de juicio de expertos y posteriormente desarrollamos la aplicación informática del presente trabajo.

3.6. Método de análisis de datos

Para el análisis de datos registrados de la presente investigación, tanto en el pre-test como en el post-test, se utilizaron las herramientas digitales Microsoft Excel y el software estadístico IBM SPSS v25. Asimismo, para el análisis descriptivo se utilizaron tablas y figuras, exponiendo medidas de tendencia central usando la media, para la interpretación o lectura de los datos emitidos por el instrumento por cada indicador.

Seguidamente, en el análisis inferencial, se comprobó la normalidad de los datos mediante la prueba Test de Shapiro Wilk, por ser considerada como una de

las pruebas estadísticas más sólidas para este análisis (Pedrosa, Juarros-Basterretxea, Robles-Fernández, Basteiro y García-Cueto, 2015; Yap y Sim, 2011; Royston, 1992) y luego para la contrastación de la hipótesis se utilizó la prueba no paramétrica de los rangos con signo de Wilcoxon.

3.7. Aspectos éticos

Con la finalidad de garantizar la integridad en la presente investigación, se cumplió con aplicar las principales normas asociadas a aspectos éticos, como los estándares de ética de la Universidad Cesar Vallejo-Resolución de Consejo 0262-2020UCV, que contienen las directivas para una correcta transparencia y veracidad de la información, la ley 29733 Ley de protección de datos personales y su reglamento, Reglamento de calificación, clasificación y registro de investigadores del SINACYT y su reglamento; y tomando en cuenta la veracidad de todo lo exhibido en presente proyecto, se asume la responsabilidad y el compromiso de las políticas de uso jurídico y ético, respetando y manteniendo la privacidad de las mismas. Además, para la autenticidad de los datos recolectados y para respetar las políticas anti plagio, se utilizó el software Turnitin.

IV. RESULTADOS

Análisis descriptivo

Medidas descriptivas del indicador 1: Tiempo de ejecución

Tabla 5

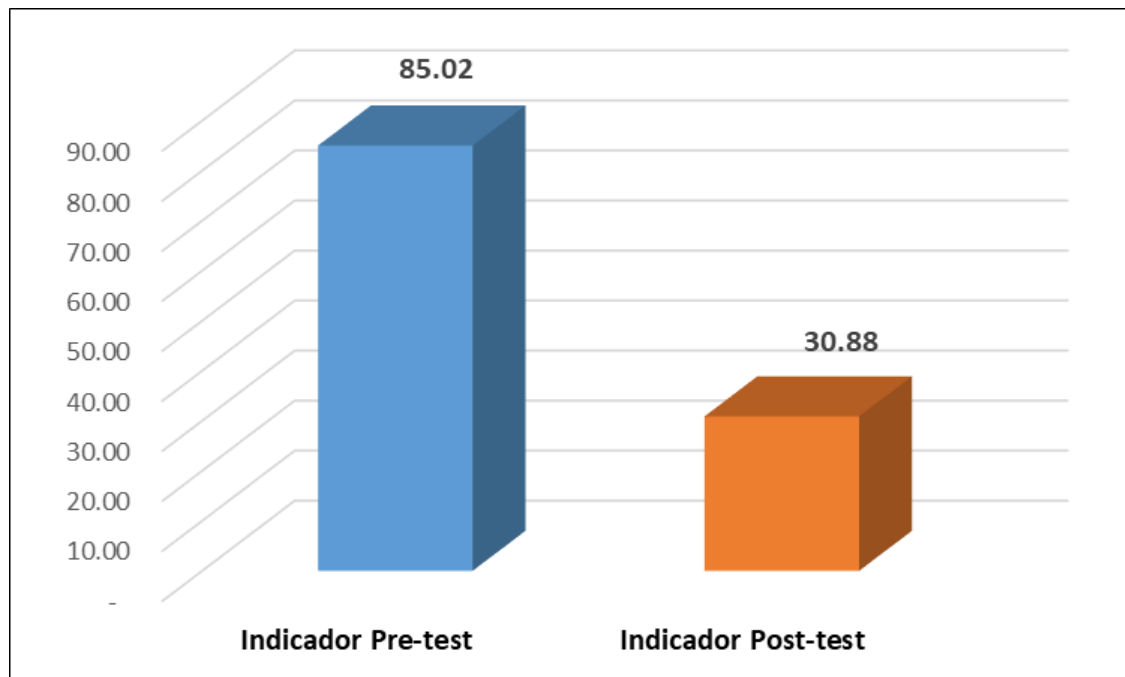
Medidas descriptivas del indicador 1: tiempo de ejecución

	N	Mínimo	Máximo	Media	Desviación Estándar
Indicador 1 Pre-test	50	10	592	85.02	121.10250
Indicador 1 Post-test	50	4	148	30.88	37.83025

Nota. Datos asistidos en el Software IBM SPSS V25.

Figura 1

Comparación de valores medios del indicador tiempo de ejecución.



En la tabla 5 se muestra los datos resultantes del análisis descriptivo del indicador tiempo de ejecución, en el pre-test de la muestra la media fue de 85.02 veces y el valor del post-test fue de 30.88 veces que se redujo los tiempos. En este sentido, podemos afirmar que existe una mejora significativa en la disminución del tiempo en la ejecución del proceso de gestión de conocimiento, después de implementar la tecnología machine learning. Asimismo, se aprecia una diferencia de medias entre el pre-test y pos-test de 54.14 minutos, que se traduce en 63.68% en promedio.

En la figura 1 se refleja el comportamiento del indicador tiempo de ejecución antes y después de la aplicación de la tecnología machine learning, en base a los datos obtenidos en la guía de observación y como se puede observar se concluye que el tiempo de ejecución del proceso de gestión de conocimiento se redujo considerablemente.

Asimismo, en el anexo 8a se puede observar las tendencias del comportamiento de este indicador producto de las observaciones realizadas.

Medidas descriptivas del indicador 2: Índice de expedientes procesados

Tabla 6

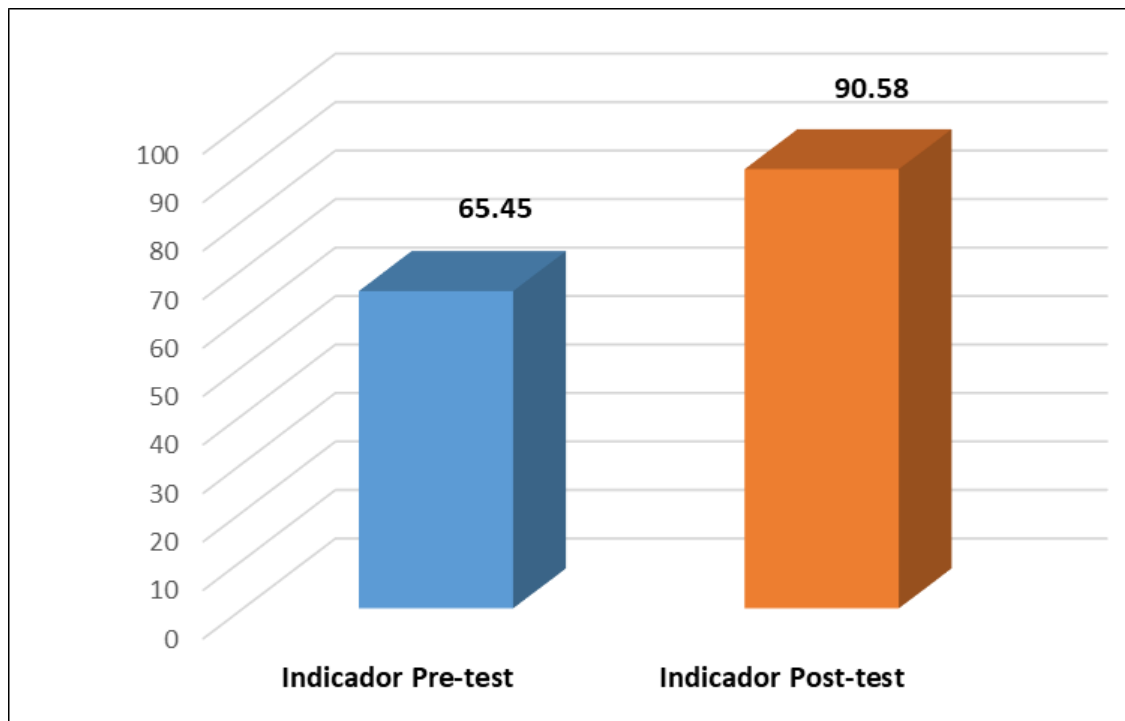
Medidas descriptivas del indicador 2: Índice expedientes procesados

	N	Mínimo	Máximo	Media	Desviación Estándar
Indicador 2 Pre-test	50	31.03	100	65.45	20.29708
Indicador 2 Post-test	50	76.47	100	90.58	6.21661

Nota. Datos asistidos en el Software IBM SPSS V25.

Figura 2

Comparación de valores medios del indicador Índice de expedientes procesados



En la tabla 6, se muestran los datos resultantes del análisis descriptivo del indicador índice de expedientes procesados; en el pre-test se obtuvo una media porcentual de 65.45% y en el post-test se incrementó a 90.58%, evidenciando una mejora significativa con una diferencia de 25.13% después de implementar la tecnología machine learning.

En la figura 2 se refleja el comportamiento de los valores medios del indicador índice de expedientes procesados antes y después de la aplicación de la tecnología machine learning, en base a los datos obtenidos en la guía de observación, concluyendo que el índice de expedientes procesados del proceso de Gestión de conocimiento se incrementó respecto al pre-test.

Asimismo, en el anexo 8b se puede observar las tendencias del comportamiento de este indicador producto de las observaciones realizadas.

Medidas descriptivas del indicador 3: Índice de expedientes procesados en el tiempo previsto.

Tabla 7

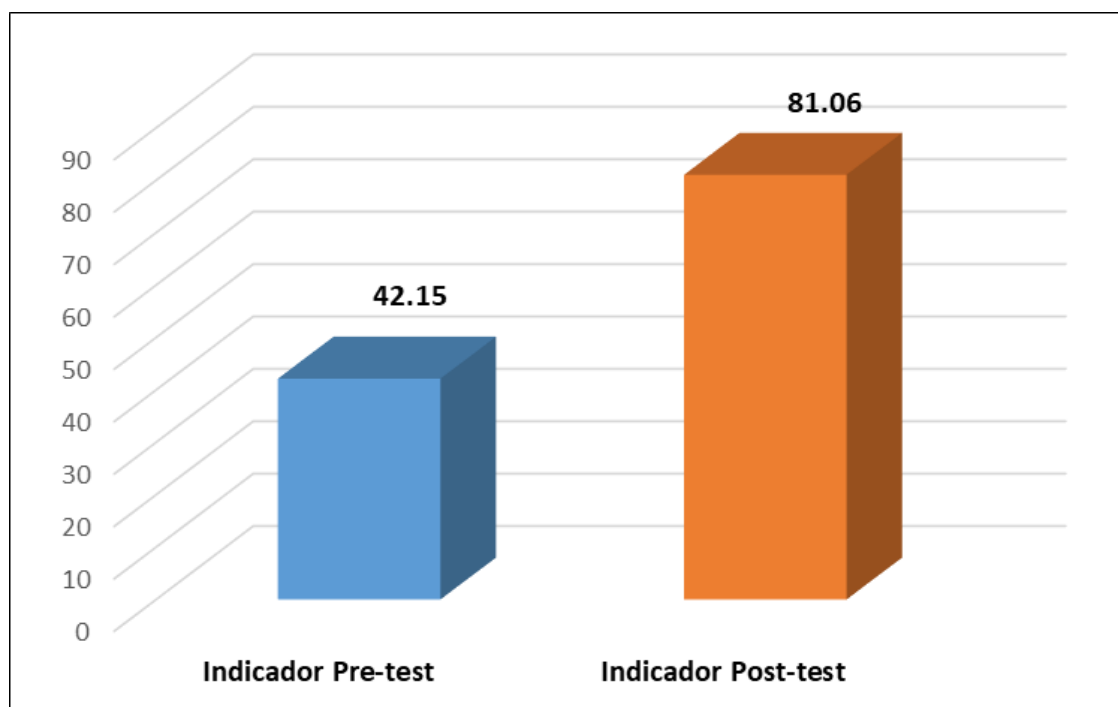
Medidas descriptivas del indicador 3: Índice de expedientes procesados en el tiempo previsto

	N	Mínimo	Máximo	Media	Desviación Estándar
Indicador 3 Pre-test	50	0	100	42.15	31.47594
Indicador 3 Post-test	50	63.16	100	81.06	10.33298

Nota. Datos asistidos en el software IBM SPSS V25.

Figura 3

Comparación de valores medios del indicador Índice de expedientes procesados en el tiempo previsto.



En la tabla 7, se muestran los datos resultantes del análisis descriptivo del indicador índice de expedientes procesados en el tiempo previsto; en el pre-test se obtuvo una media porcentual de 42.15% y en el post-test se incrementó a 81.06%, evidenciando una mejora significativa con una diferencia de 38.91% después de implementar la tecnología machine learning.

En la figura 3 se refleja el comportamiento del indicador índice de expedientes procesados en el tiempo previsto antes y después de la aplicación de la tecnología machine learning, en base a los datos obtenidos en la guía de observación, concluyendo que el índice de expedientes procesados en el tiempo previsto del proceso de Gestión de Conocimiento se incrementó respecto al pre-test. Asimismo, en el anexo 8c se puede observar las tendencias del comportamiento de este indicador producto de las observaciones realizadas.

Análisis Inferencial

Prueba de Hipótesis

De acuerdo con Pedrosa Juarros-Basterretxea, Robles-Fernández, Basteiro y García-Cueto (2015) es recomendable utilizar pruebas no paramétricas cuando los datos a analizar no cumplen los supuestos de normalidad y considerando que el resultado de la prueba Shapiro Wilk indica que los indicadores no presentan una distribución normal, se decidió aplicar la prueba de Rango de Wilcoxon, utilizada para contrastar hipótesis no paramétricas.

Esta prueba tal como señalan Berlanga y Rubio (2012); Ramírez y Polack (2020), permite contrastar la hipótesis de igualdad entre dos medianas poblacionales, considerando que los datos formen parte de la misma muestra con medición de pre-test y pos-test.

Prueba Hipótesis específica 1: Indicador tiempo de ejecución.

Formulación de hipótesis estadística:

H₀: Machine Learning no mejora significativamente el tiempo de ejecución de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

H₁: Machine Learning mejora significativamente el tiempo de ejecución de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

Considerando que el resultado de la prueba de normalidad del indicador tiempo de ejecución no es normal (ver Anexo 7), se aplicó la prueba de Wilcoxon.

Tabla 8

Pruebas de Wilcoxon – Rangos del indicador 1 tiempo de ejecución

		N	Rango promedio	Suma de rangos
Postest - Pretest	Rangos negativos	50 ^a	25.50	1275.00
	Rangos positivos	0 ^b	0.00	0.00
	Empates	0 ^c		
	Total	50		

Nota. Datos asistidos en el software IBM SPSS V25.

a. Postest < Pretest

b. Postest > Pretest

c. Postest = Pretest

Tabla 9

Pruebas de Wilcoxon – Estadísticos de contraste indicador 1 tiempo de ejecución

	Prueba de rangos con signo de Wilcoxon	
	Z	Sig. Asintótica (bilateral)
Índice Tiempo de ejecución		
Postest – Pretest	-6.216	0.000

Nota. Datos asistidos en el software IBM SPSS V25.

Para contrastar la hipótesis se utilizó la prueba no paramétrica de Rango de Wilcoxon. Como puede apreciarse en la tabla de valores de rangos (Tabla 8) los 50 pares de elementos analizados pertenecen a los rangos negativos, que indica que el indicador tiempo de ejecución después de la aplicación de machine learning (postest) es menor al tiempo de ejecución aplicado en situación inicial (pretest).

Asimismo, en la tabla de estadísticos de contraste (Tabla 9) se muestra que el valor Z obtenido fue de -6.216, el cual se ubica en la zona de rechazo de la hipótesis nula; y el valor de significancia fue 0.000 el cual es menor al valor asumido de 0.05, por tanto, se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1).

Prueba Hipótesis específica 2: Indicador Índice de expedientes procesados.

Formulación de hipótesis estadística:

- H_0 : Machine Learning no mejora significativamente el índice de expedientes procesados de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.
- H_1 : Machine Learning mejora significativamente el índice expedientes procesados de la gestión del conocimiento en el Consejo Nacional de

Considerando que el resultado de la prueba de normalidad del indicador índice de expedientes procesados no es normal (ver Anexo 7), se aplicó la prueba de Wilcoxon.

Tabla 10

Pruebas de Wilcoxon – Rangos del indicador 2 índice de expedientes procesados.

		N	Rango promedio	Suma de rangos
Postest - Pretest	Rangos negativos	6 ^a	6.67	40.00
	Rangos positivos	42 ^b	27.05	1136.00
	Empates	2 ^c		
	Total	50		

Nota. Datos asistidos en el software IBM SPSS V25.

- a. Postest < Pretest
- b. Postest > Pretest
- c. Postest = Pretest

Tabla 11

Pruebas de Wilcoxon – Estadísticos de contraste del indicador 2 índice de expedientes procesados

	Prueba de rangos con signo de Wilcoxon	
	Z	Sig. Asintótica (bilateral)
Índice de expedientes procesados	-5.621	0.000
Pretest – Postest		

Nota. Datos asistidos en el software IBM SPSS V25.

Para contrastar la hipótesis se utilizó la prueba no paramétrica de Rango de Wilcoxon. Como se muestra en la tabla de valores de rangos (Tabla 10) se analizaron 50 pares de elementos, 6 pertenecen a rangos negativos, 42 a rangos positivos y 2 empates, que indica el mayor número de casos se encuentran en los rangos positivos por lo que el indicador índice de expedientes procesados después de la aplicación de machine learning (postest) es mayor que el índice de expedientes procesados aplicado en situación inicial (pretest).

Asimismo, en la tabla de estadísticos de contraste (Tabla 11) se muestra que el valor Z obtenido fue de -5.621, el cual se ubica en la zona de rechazo de la hipótesis nula; y el valor de significancia fue 0.000 el cual es menor al valor asumido de 0.05, por tanto, se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1).

Prueba Hipótesis específica 3: Indicador Índice de expedientes procesados en el tiempo previsto.

Formulación de hipótesis estadística:

H_0 : Machine Learning no mejora significativamente el índice de expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

H_1 : Machine Learning mejora significativamente el índice de expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

Considerando que el resultado de la prueba de normalidad del indicador índice de expedientes procesados en el tiempo previsto no es normal (ver Anexo 7), se aplicó la prueba de Wilcoxon.

Tabla 12

Pruebas de Wilcoxon - Rangos del indicador 3 índice de expedientes procesados en el tiempo previsto

		N	Rango promedio	Suma de rangos
Postest - Pretest	Rangos negativos	6 ^a	8.17	49.00
	Rangos positivos	42 ^b	26.83	1127.00
	Empates	2 ^c		
	Total	50		

Nota. Datos asistidos en el software IBM SPSS V25.

a. Postest < Pretest

b. Postest > Pretest

c. Postest = Pretest

Tabla 13

Pruebas de Wilcoxon – Estadísticos de contraste del indicador 3 índice de expedientes procesados en el tiempo previsto.

	Prueba de rangos con signo de Wilcoxon	
	Z	Sig. Asintótica (bilateral)
Índice de expedientes procesados en el tiempo previsto Pretest – Postest	-5.528	0.000

Nota. Datos asistidos en el software IBM SPSS V25.

Para contrastar la hipótesis se utilizó la prueba no paramétrica de Rango de Wilcoxon. Como se muestra en la tabla de valores de rangos (Tabla 12) se analizaron 50 pares de elementos, 6 pertenecen a rangos negativos, 42 a rangos positivos y 2 empates, que indica el mayor número de casos se encuentran en los

rangos positivos por lo que el indicador índice de expedientes procesados en el tiempo previsto después de la aplicación de machine learning (postest) es mayor que el índice de expedientes procesados en el tiempo previsto aplicado en situación inicial (pretest).

Asimismo, en la tabla de estadísticos de contraste (Tabla 13) se muestra que el valor Z obtenido fue de -5.528, el cual se ubica en la zona de rechazo de la hipótesis nula; y el valor de significancia fue 0.000 el cual es menor al valor asumido de 0.05, por tanto, se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1).

V. DISCUSIÓN

De acuerdo a los resultados obtenidos en la presente investigación, se observa que se han producido cambios favorables en los tres indicadores planteados de la variable dependiente gestión del conocimiento, después de la implementación de la variable independiente machine learning en el Consejo Nacional de Ciencia y Tecnología – Concytec.

Respecto al indicador tiempo de ejecución, tenemos según sus resultados obtenidos, que después del experimento el tiempo de ejecución disminuyó respecto a la situación inicial. Es así que, en su análisis descriptivo, el cual se realizó con una toma muestral de 50 observaciones, se observó una disminución en el tiempo de ejecución de 55.14 minutos, producto de la diferencia de medias del pre-test y pos-test, que se traduce en un 63.68% en promedio. Por tanto, podemos afirmar que, el tiempo de ejecución disminuye, después de implementar la tecnología machine learning en comparación con la forma tradicional.

Asimismo, en análisis inferencial de este indicador tiempo de ejecución, se aplicó la prueba Shapiro-Wilk para determinar la normalidad, en donde resultado confirmó que el indicador tiempo de ejecución no presenta una distribución normal, por lo que se decidió realizar la prueba no paramétrica de Rango de Wilcoxon para contrastar la hipótesis. En la tabla de Rangos (Tabla 8) se aprecia que el indicador tiempo de ejecución después de la aplicación de machine learning (postest) es menor al tiempo de ejecución aplicado en situación inicial (pretest). Asimismo, respecto a los estadísticos de contraste (Tabla 9) se obtuvo un valor Z de -6.216, el cual se ubica en la zona de rechazo de la hipótesis nula con un valor de significancia de 0.000 menor al valor asumido de 0.05 por tanto, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 , por lo que se concluye que machine learning mejora significativamente el tiempo de ejecución de la gestión del conocimiento en Concytec.

Estos resultados concuerdan con los resultados obtenidos por Medina (2018) quien en su investigación comprueba que haciendo uso de modelos predictivos machine learning el tiempo de atención para pacientes con depresión disminuyó considerablemente. De similar manera López (2019) en su investigación comprobó que aplicando el modelo predictivo XGBoost machine learning se puede identificar en un menor tiempo los malos y buenos pagadores en empresas del sector financiero. Otro antecedente interesante es el de Zhao, Zhang, Liu, Wang, Zhu *et al.* (2021) quienes realizaron investigaciones ambientales con estudios predictivos de aprendizaje automático para mejorar la eficiencia del sector acuicultor; demostraron que aplicación de modelos predictivos machine learning, disminuyen los tiempos en la detección de enfermedades de peces de manera oportuna, en un uso preventivo. Neyra (2021) en su investigación desarrolló modelos de predicción machine learning para estimar el contenido de cadmio en granos de cacao a partir de firmas hiperespectrales en el proceso productivo del cacao para detectar en tiempo real el riesgo de contaminación por metales pesados como el cadmio en el producto final. Demostró que aplicando modelos predictivos se lograba predecir que los tiempos de ejecución para detección del riesgo se disminuían considerablemente, mejorando así la gestión del riesgo operacional.

Dichos resultados se encuentran dentro del marco conceptual del indicador y según Álvarez y Lobo (2020), Beltrán (2017), RP N° 007 - 2011 - CONCYTEC-P (2011) señalan que el tiempo de ejecución mide la rapidez del proceso, es el tiempo que se da durante el proceso, que se inicia cuando los expedientes o documentos provenientes de usuarios se reciben y registran en el sistema, transcurre en el desarrollo de diversas actividades hasta la etapa de cierre donde se genera un resultado final del proceso y es comunicado.

Respecto al indicador índice de expedientes procesados observamos que, en el resultado del análisis descriptivo, en el pre-test la media porcentual fue de 65.45% y en el post-test se incrementó a 90.58%, evidenciando una mejora significativa con una diferencia de 25.13% después de implementar la tecnología

machine learning, demostrando que existe una mejora significativa en el proceso de gestión del conocimiento después de implementar la tecnología machine learning.

Para análisis inferencial de este indicador, se aplicó la prueba Shapiro-Wilk para determinar la normalidad, en donde el resultado confirmó que el indicador índice de expedientes procesados no presenta una distribución normal, por lo que se decidió realizar la prueba no paramétrica de Rango de Wilcoxon para contrastar la hipótesis. En la tabla de Rangos (Tabla 10) se aprecia que el indicador índice de expedientes procesados después de la aplicación de machine learning (postest) es mayor al índice de expedientes procesados aplicado en situación inicial (pretest). Asimismo, respecto a los estadísticos de contraste (Tabla 11) se obtuvo un valor Z de -5.621, el cual se ubica en la zona de rechazo de la hipótesis nula con un valor de significancia de 0.000 menor al valor asumido de 0.05 por tanto, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 : Machine learning mejora significativamente el índice de expedientes procesados de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021, demostrando así que machine learning mejora significativamente el índice de expedientes procesados de la gestión del conocimiento en Concytec.

Estos resultados concuerdan con los resultados obtenidos por Knoll, Prüglmeier y Reinhart *et al* (2016) quien investigó sobre predicciones de escenarios logísticos basados en modelos predictivos machine learning, para el procesamiento de expedientes de contratación, concluyendo que mediante la aplicación de varios modelos predictivos de machine learning se logró reducir esfuerzo y mejorar el proceso logístico. Asimismo, Gamarra (2020), investigó sobre la aplicación de modelos de predicción machine learning relacionados con mortalidad perinatal en la Región de Junín, donde se procesaron expedientes clínicos de pacientes para entrenamiento y pruebas del modelo machine learning basado en el algoritmo

WJ48, obteniendo con alta precisión la identificación de pacientes con riesgo de muerte prenatal para mejorar el proceso de salud.

También tenemos a Li, Yang, Wang y Wu (2018), quien investigó sobre la predicción del desempeño de innovación de las Empresas de base tecnológicas del Parque Científico Zhongguancun – ZGC en China. Demostrando que con el uso del método predictivo XGBoost de machine learning y en base al procesamiento de expedientes de los proyectos de innovación de los años 2005 al 2015, se pudo predecir el comportamiento y evaluación de los proyectos innovación mejorando la toma de decisiones en la elección de financiación limitada y con altas probabilidades de inversión. Asimismo, Ashok y Pellakuri (2020) en su investigación implementaron algoritmos predictivos machine learning para estimar el riesgo de salud de niños menores a 5 años de las poblaciones vulnerables con necesidad de tratamiento temprano. En la implementación del modelo predictivo se procesaron expedientes clínicos de pacientes menores a cinco años, demostrando que machine learning mejoró el proceso de salud al identificar en la población a los pacientes con riesgo alto para su atención por medio de programas de tratamientos tempranos, reduciendo así los niveles de mortalidad.

Dichos resultados se encuentran alineados con el marco conceptual del indicador índice de expedientes procesados. Así tenemos, La RSEGDI No.001-2017-PCM/SEGDI (2017), ISO 15489-1 (2016), RDE No.38-2019-MEF (2018) señalan que un expediente procesado es un conjunto de documentación agrupada en un tema o asunto específico que es trabajada en un expediente digitalizado que transita a lo largo del ciclo vital del proceso, desde su recepción hasta su disposición final donde se cumple el propósito de la atención, dentro del marco tecnológico, procedimientos informáticos y normatividad vigente que forman parte de su proceso.

Respecto al Indicador 3: Índice de expedientes procesados en el tiempo previsto, el cual está referido al cumplimiento de metas para lograr el propósito de la atención

en el tiempo requerido, obtuvo como resultado de su análisis descriptivo, que en el pre-test la media porcentual fue de 42.15% y en el post-test se incrementó a 81.06%, evidenciando una mejora significativa con una diferencia de 38.91% después de implementar la tecnología machine learning.

Para análisis inferencial de este indicador, primero se aplicó la prueba Shapiro-Wilk para determinar la normalidad, en donde resultado confirmó que el indicador índice de expedientes procesados en el tiempo previsto, no presenta una distribución normal, por lo que se utilizó la prueba no paramétrica de Rango de Wilcoxon para contrastar la hipótesis. En la tabla de Rangos (Tabla 12) se aprecia que el indicador índice de expedientes procesados en el tiempo previsto después de la aplicación de machine learning (postest) es mayor al índice de expedientes procesados en el tiempo previsto aplicado en situación inicial (pretest). Asimismo, respecto a los estadísticos de contraste (Tabla 13) se obtuvo un valor Z de -5.528, el cual se ubica en la zona de rechazo de la hipótesis nula con un valor de significancia de 0.000 menor al valor asumido de 0.05, por tanto, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 : Machine learning mejora significativamente el índice de expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

Estos resultados se encuentran acordes con los resultados obtenidos por López (2019) en su investigación quien comprobó que un modelo predictivo machine learning con alta exactitud del 97.53% mediante el procesamiento de expedientes crediticios en el tiempo previsto, permite identificar oportunamente a los clientes que son buenos y malos pagadores en las empresas del sector financiero mejorando significativamente su proceso empresarial y sus resultados financieros, para lo cual incluyó estos parámetros de cumplimiento de acuerdo a metas en su diseño. Procesos como éste requieren culminarse oportunamente para que no pierdan su relevancia e importancia, es decir proporcionar la atención, pero dentro de los tiempos y/o requisitos establecidos.

Respecto al objetivo general, podemos indicar que la implementación machine learning aplicado a la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021, obtuvo resultados positivos, tal como se indican en los siguientes párrafos.

El indicador tiempo de ejecución, confirmó que se produjo una disminución significativa del tiempo de ejecución en el proceso de gestión del conocimiento, que representa un 63.68% respecto al escenario anterior; asimismo en el análisis inferencial se obtuvieron evidencias que permiten rechazar la hipótesis nula y aceptar la hipótesis alterna que Machine Learning mejora significativamente el tiempo de ejecución de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

De manera similar, se confirmó que el indicador índice de expedientes procesados mejoró significativamente después de implementar la tecnología machine learning, evidenciado en el incremento obtenido de 25.13% respecto a la situación inicial y como resultado del análisis inferencial se aceptó la hipótesis alterna: Machine Learning mejora significativamente el índice expedientes procesados de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

El último indicador acorde con los resultados anteriores, confirmó que el índice de expedientes procesados en el tiempo previsto mejoró significativamente después de la implementación de la tecnología machine learning, obteniendo un incremento de 38.91% respecto al escenario anterior y como resultado del análisis inferencial se aceptó la hipótesis alterna: Machine Learning mejora significativamente el índice expedientes procesados en el tiempo previsto de la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.

En consecuencia, por lo indicado anteriormente machine learning mejora significativamente la gestión del conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica, Concytec-2021.

Este resultado se contrasta con lo indicado por Medina (2018), López (2019), Zhao, Zhang, Liu, Wang, Zhu *et al.* (2021), Knoll, Prüglmeier y Reinhart (2016), Gamarra (2020), Li, Yang, Wang y Wu (2018), Ashok y Pellakuri (2020), Neyra (2021) quienes coinciden que la aplicación de modelos predictivos de machine learning mejora los procesos que generan conocimiento en las organizaciones, y se ha convertido en una herramienta poderosa para las Organizaciones en diversas Áreas del Conocimiento tanto en la mejora de tiempos, cumplimientos, eficacia de sus operaciones, toma de decisiones entre otras. También coinciden que es un campo que aún falta explorar, difundir y explotar.

La metodología de investigación que se utilizó permitió fortalecer la investigación, debido a que el diseño de investigación experimental puro, facilita el control de la validez interna del experimento mediante la asignación aleatoria. Adicionalmente, mediante las pruebas pre-test y post-test se pudo medir con mayor exactitud el cambio aplicado, con la finalidad de describir sus resultados, identificando la relación de causa y efecto, es decir, la relación directa entre las variables de la investigación. Asimismo, permitió establecer el actual flujo de trabajo de la organización en relación a los indicadores.

También es importante señalar, que las guías de observación como instrumento de recolección de datos, nos ayudó en el control y seguimiento durante el proceso de recolección de datos. Asimismo, las herramientas tecnológicas utilizadas como el SPSS v25 para el procesamiento de los modelos predictivos fueron de mucho apoyo para el éxito del trabajo.

De similar forma, los indicadores que forman parte de nuestra investigación nos permitieron conocer el proceso de Gestión Conocimiento transversal a varias Áreas y crear una base de datos para ser explotada y continuar su uso.

Respecto a la relevancia social científica, esta investigación proporciona el conocimiento necesario para que sea aplicado en las organizaciones especialmente en las del sector público, como alternativa de solución en el tratamiento de grandes volúmenes de expedientes.

VI: CONCLUSIONES

Primera: De acuerdo a los resultados obtenidos en la presente investigación efectuada en la Empresa Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica – Concytec, se concluye que con la implementación de Machine Learning se mejora significativamente la Gestión de Conocimiento, teniendo como puntos fuertes de mejora de sus tres indicadores: tiempo de ejecución, expedientes procesados y expedientes procesados en el tiempo previsto respecto a sus situaciones iniciales. Estos resultados fueron contrastados con sus respectivas pruebas de hipótesis.

Segunda: El primer indicador tiempo de ejecución, dio como resultado una mejora significativa después de la aplicación de machine learning, ya que disminuyó en un 63.68% en promedio el tiempo de procesamiento de expedientes y obtuvo un valor estadístico de prueba Z de -6.216 con un p valor de 0.000 de probabilidad de significancia que determinaron el rechazo de la hipótesis nula y la aceptación de la hipótesis alterna. Por lo que podemos afirmar que el tiempo empleado en el proceso de gestión de conocimiento con la aplicación de Machine Learning es significativamente menor.

Tercera: En el segundo indicador índice de expedientes procesados, tuvo como resultado una mejora significativa después de la aplicación de machine learning, ya que se incrementó en promedio un 25.13% respecto a la situación inicial, y obtuvo un valor estadístico de prueba Z de -5.621 con un p valor de 0.000 de probabilidad de significancia que determinaron el rechazo de la hipótesis nula y la aceptación de la hipótesis alterna. Por tanto, podemos afirmar que en un periodo se podrán resolver mayor número de expedientes.

Cuarta : Para el tercer indicador índice de expedientes procesados en el tiempo previsto, se dio como resultado una mejora significativa después de la aplicación machine learning, ya que se comprobó un incremento promedio de 42.15% respecto al escenario inicial, y obtuvo un valor estadístico de prueba Z de -5.528 con un p valor de 0.000 de probabilidad de significancia que determinaron el rechazo de la hipótesis nula y la aceptación de la hipótesis alterna, por lo que podemos

indicar que se mejoró el cumplimiento de metas establecidas de acuerdo a lo planificado por periodos, permitiendo contar con un mejor control del avance de los expedientes procesados.

VII. RECOMENDACIONES

Primera: Para el sostenimiento de los resultados positivos en los tres indicadores, obtenidos por la investigación realizada a la Empresa Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica, después de la implementación de Machine Learning en el proceso de Gestión de Conocimiento, se sugiere que el Sub- Director del Área usuaria, solicite como parte de la explotación de datos, la incorporación de nuevos indicadores como índice de apelaciones e índice de reconsideraciones, que permitan optimizar el proceso de Gestión del Conocimiento.

Segunda: Respecto al indicador de tiempos de ejecución, a la luz de los resultados de disminución significativa de tiempos, se sugiere que el especialista del Área usuaria, realice un sinceramiento de los tiempos para realizar ajustes a los estándares de programación de trabajos relacionados con los tiempos de ejecución del proceso Gestión del Conocimiento.

Tercera: Sobre el indicador expedientes procesados, se sugiere que el supervisor del procesamiento de expedientes solicite el desarrollo de nuevos modelos predictivos que trabajen con otros escenarios, que permitan atender la mayor cantidad de posibilidades de los expedientes pendientes de procesar, como son los expedientes improcedentes.

Cuarta: Respecto al indicador expedientes procesados en el tiempo previsto, se sugiere que el supervisor del procesamiento de expediente considere los resultados de la mejora significativa del índice de expedientes procesados en el tiempo previsto, para la actualización de la programación de asignación de recursos humanos para atención de expedientes e incorporación de nuevos hitos de control para facilidad del seguimiento.

REFERENCIAS

- Álvarez. y Lobo, G. (2020) El tiempo, su uso y abuso. *Información Tecnológica*, (31). Recuperada de: <http://dx.doi.org/10.4067/S0718-07642020000200073>
- Ansari, F. (2019). Knowledge Management 4.0: Theoretical and Practical Considerations in Cyber Physical Production Systems. *Revista IFAC paper on line* . Recuperado en: <https://doi.org/10.1016/j.ifacol.2019.11.428>
- Arellano, A. (2014) Epistemología antropológica como conocimiento del hombre: El papel de la antropología de la tecnociencia. *Revista Acta Sociológica*, (63), 15-39. Recuperada de: [https://doi.org/10.1016/S0186-6028\(14\)70474-8](https://doi.org/10.1016/S0186-6028(14)70474-8)
- Arias, R. y J. Mejía (2020) Knowledge management in two universities before and during the COVID-19 effect in Peru. *Revista Technology in Society*, (64). Recuperado de: <https://doi.org/10.1016/j.techsoc.2020.101479>
- Arunakranthi, G., Rajkumar, B., Chandra, V. y Harshavardhan, A (2020). Advanced patterns of predictions and cavernous data analytics using machine learning. *Revista Materials Today: Proceedings*. Recuperado de: <https://doi.org/10.1016/j.matpr.2020.11.062>
- Ashok, M, y Pellakuri, V. (2021). Innovative supervised machine learning techniques for classification of data. *Revista Materialstoday Proceeding*. Recuperado de: <https://doi.org/10.1016/j.matpr.2020.11.092>
- Berlanga, V. y Rubio, M.(2012) Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *REIRE - Revista d'Innovació i Recerca en Educació*, 5(2), 101-113. Recupero de: <https://doi.org/10.1344/reire2012.5.2528>
- Barrera, M.. (2018). Juzgado sin papel, un paso más de la justicia electrónica. *Revista IUS*. Recuperado de: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-21472018000100133
- Beltrán. J. (2017) *Indicadores de Gestión. Guía práctica para estructurar acertadamente esta herramienta clave para el logro de la competitividad. (4ta ed.)*. Colombia: Panamericana Editorial Ltda.

- Cabero, J. y Lorente, M (2015). Tecnologías de la Información y la Comunicación (TIC): escenarios formativos y teorías del aprendizaje. *Revista Lasallista de Investigación*, 12, 186-193. Recuperado de: <https://www.redalyc.org/pdf/695/69542291019.pdf>
- Calvani, A. (2008). Connectivism: New paradigm or fascinating pot-pourri?. *Journal of E-Learning and Knowledge Society-English Version*, 4(1). Disponible en https://www.je-lks.org/ojs/index.php/Je-LKS_EN/article/view/60
- Calvo, J, Pelegrin, A. y Gil, M. (2018). Enfoques teóricos para la evaluación de la eficiencia y eficacia en el primer nivel de atención médica de los servicios de salud del sector público. *Retos de la Dirección*, 12(1), 96-118. Recuperado de: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2306-91552018000100006
- Córdova, F., Durán, C., Pinchera, M., Palominos, F. y Galindo, R. (2020). Knowledge Management of Intangible Actives in Service. *Procedia Computer Science*, 162, 596-603. Recuperado en: <https://www.sciencedirect.com/science/article/pii/S1877050919320381>
- Costa,K., Johann, J., Passeri, P. y Scott, P. (2017) Gestão do conhecimento nas organizações. Uma aprendizagem em rede colaborativa. *Perspectivas em Gestão & Conhecimento*, 7, 145-159. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=5908254>
- Dalmarco, G., Maehler, A., Trevisan, M. y Mortari, J. (2017). The use of knowledge management practices by Brazilian startup companies. *RAI Revista de Administração e Inovação*, 14, 226-234. Recuperado de: <http://dx.doi.org/10.1016/j.rai.2017.05.005>
- Falaster, C., Zanin, L. y Guerrazzi, L. (2017). Teoria institucional na pesquisa em turismo: novas oportunidades de uma teoria em evolução. *RETUR Revista Brasileira de pesquisa em Turismo Brazilian*, 11(2), 270-293. Recuperado en: <http://dx.doi.org/10.7784/rbtur.v11i2.1310>
- Gamarra, W. (2020). *Modelos de aprendizaje automático basado en técnicas supervisadas para la predicción de la mortalidad perinatal en la Región*

- Junín. [Tesis de Doctorado, Universidad Nacional Federico Villareal].
<http://repositorio.unfv.edu.pe/handle/UNFV/4355>
- Hernández, R, Fernández, C. y Baptista, P. (2014). *Metodología de la Investigación (6ta)*. México D.C.: MacGraw Hill – Interamericana Editores S.A.
- ISO 15489-1 (2016). *Information and documentation — Records management — Part 1: Concepts and principles*. Recuperado de:
<https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-2:v1:en>
- Jardim, G. y Bronzo, M. (2016). Teorias institucionais aplicadas aos estudos de sistemas agroindustriais no contexto do agronegócio café: uma análise conceitual. *REGE - Revista de Gestão*, 23(2), 159-171. Recuperado de:
<https://doi.org/10.1016/j.rege.2015.12.005>
- Jordan, M. y Mitchell, T. (2015). Machine learning: trends, perspectives, and prospects. *Science*, 349, 255–260. Recuperado de:
<https://doi.org/10.1126/science.aaa8415>
- Knoll, D., Prüglmeier, M. y Reinhart, G. (2016). Predicting Future Inbound Logistics Processes using Machine Learning. *Revista Procedia CIRP*, 52, 145-150. Recuperado de: <https://doi.org/10.1016/j.procir.2016.07.078>
- Krishna, M, Amirali, B., Swathi, A. y Madhavi, V. (2021). High performance machine learning and data science based implementation using Weka. *Materials Today: Proceedings*. Recuperado de:
<https://doi.org/10.1016/j.matpr.2021.01.470>
- Li, Y., Yang, L., Yang, B., Wang, N. y Wu, T. (2018). Application of interpretable machine learning models for the intelligent decision. *Revista Neurocomputing*, 333, 273-283. Recuperado de :
<https://doi.org/10.1016/j.neucom.2018.12.012>
- López, J. (2019). *Comparación de modelos de aprendizaje de máquina en la predicción del incumplimiento de pago en el sector de las microfinanzas*. [Tesis de Maestría, Pontificia Universidad Católica del Perú].
<http://repositorio.pucp.edu.pe/index/handle/123456789/179093>
- Medina, O. (2018). *Detección de depresión a través de análisis textual utilizando aprendizaje automático 2017*. [Tesis de Maestría, Universidad Ricardo

Palma].

<https://repositorio.urp.edu.pe/bitstream/handle/URP/1674/AMOru%c3%a9M.pdf?sequence=1&isAllowed=y>

- Montoya, L., Parra, M., Arias, M., Cabello, O. y Coloma, G. (2019). Teorías pedagógicas que sustentan el aprendizaje con el uso de las Tecnologías de la Información y las Comunicaciones. *Revista Información científica*, 98(2). Recuperado de: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1028-99332019000200241
- Mora, E. (2018). Algoritmo Machine Learning en los sistemas de filtrado caso práctico SPAM de google en las cuentas de correo Institucionales de la FAFI. *Pro Sciences*, 2(12), 17-23. Recuperado de: <https://doi.org/10.29018/issn.2588-1000vol2iss12.2018pp17-23>
- Murali, K., Barkat, A., Arelli, S y Madhavi, V. (2021). High performance machine learning and data science based implementation using Weka . *Revista Materials Today: Proceedings*. Recuperado de: <https://doi.org/10.1016/j.matpr.2021.01.470>
- Neyra, J. (2021). *Determinación en tiempo real de presencia de cadmio en cultivo de cacao aplicando Machine Learning*. [Tesis de Maestría, Universidad de Piura]. <https://pirhua.udep.edu.pe/handle/11042/4990>
- OCDE (2014). *Medición de la gestión de conocimientos en las empresas: primeros resultados*. Recuperado de: <https://www.oecd.org/education/innovation-education/31540414.pdf>
- Ode, E. y Ayavoo, R. (2019). The mediating role of knowledge application in the relationship between knowledge management practices and firm innovation. *Journal of Innovation & Knowledge*, 5(3), 210-218. Recuperado de: <https://doi.org/10.1016/j.jik.2019.08.002>
- OIM (2016). *Indicadores de Cumplimiento. Organismo internacional para las migraciones*. Recuperado de: https://kmhub.iom.int/sites/default/files/indicadores_de_cumplimiento_-

_compromisos_adquiridos_en_materia_de_abordaje_integral_de_la_trata_de_personas_.pdf

- Pedrosa, I., Juarros-Basterretxea, J., Robles-Fernández, A., Basteiro, J., y García-Cueto, E. (2015). Pruebas de bondad de ajuste en distribuciones simétricas, ¿qué estadístico utilizar? *Universitas Psychologica*, 14(1), 245-254. Recuperado de: <https://doi.org/10.11144/Javeriana.upsy14-1.pbad>
- Quispe, A., Celi, L. y Campos, R. (2020). Uso de Machine Learning en la creación de páginas web a medida de los usuarios. *Campus*, 25(30), 337-344. Recuperado de: <https://doi.org/10.24265/campus.2020.v25n30.09>
- Ramírez, A. y Polack, A.(2020). Estadística inferencial. Elección de una prueba estadística no paramétrica en investigación científica. *Horizonte de la Ciencia*, 10(19), 191-208. Recuperado de: <https://doi.org/10.26490/uncp.horizonteciencia.2020.19.597>
- RDE No.38-2019-MEF (2018). Directiva del proceso de trámite documentario en pro inversión del Ministerio de Economía y Finanzas. Recuperado de: <https://www.proinversion.gob.pe/RepositorioAPS/0/1/JER/RDE2018/ResolucionDEN0382018.pdf>
- Restrepo, M. y Rosero, X. (2002). Teoría Institucional y proceso institucional y proceso de internacionalización de las empresas colombianas. *Estudios Gerenciales* (18). Recuperado de: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-59232002000300006#:~:text=Seg%C3%BAn%20la%20teor%C3%ADa%20institucional%2C%20DiMaggio,organizaci%C3%B3n%20por%20su%20ambiente%20externo.
- RP N° 007 - 2011 - CONCYTEC-P (2011). Directiva 001-2011-CONCYTEC-OPP (2011). Directiva de recepción, registro, derivación, cierre y archivo de la documentación en CONCYTEC. Recuperado de: <https://www.gob.pe/institucion/concytec/normas-legales/605133-007-2011-concytec-p>
- RSEGDI No.001-2017-PCM/SEGDI (2017). Aprueban modelo de gestión documental en el marco del decreto legislativo No.1310. *Diario Oficial El*

- Peruano*. Recuperado de: <https://www.gob.pe/institucion/pcm/normas-legales/292301-001-2017-pcm-segdi>
- Robert, C. (2014). Machine Learning, a Probabilistic Perspective. *Chance*, 27, 62-63. Recuperado de: <https://doi.org/10.1080/09332480.2014.914768>
- Royston, J. (1982). Approximating the Shapiro–Wilk W test for non-normality. *Royal Statistical society*, 31(2), 115-124. Recuperado de: <https://doi.org/10.2307/2347973>
- Sánchez, J. (2019). Desarrollo de un entorno digital de aprendizaje desde el Conectivismo y su posterior análisis utilizando algoritmos de machine learning. EDUTEC. *Revista Electrónica de Tecnología Educativa* (64). Recuperado de: DOI: <https://doi.org/10.21556/edutec.2019.69>
- Scott, W. (2014). *Institutions and Organizations: Ideas, Interests, and Identities* (4a). London: Sage Publications Ltd.
- Thomas, E. y Manz, D. (2017). Research Methods for Cyber Security. Synggress, 269. Recuperado de: <https://doi.org/10.1016/B978-0-12-805349-2.00035-2>
- Torres, J. y Barnabé T. (2020). Aspectos pedagógicos del conectivismo y su relación con las redes sociales y ecologías del aprendizaje. *Revista Brasileira de Educação*. Recuperado de: <http://dx.doi.org/10.1590/S1413-24782020250026>
- Uhunoma, F. Khan, R. Venkatesan y P.Guillard (2021). Review and analysis of supervised machine learning algorithms for hazardous events in drilling. *Process Safety and Environmental Protection*. Recuperado de: <https://doi.org/10.1016/j.psep.2020.09.038>
- Vázquez-Marrufo, M., Sarrias-Arrabala, E., García-Torres, M., Martín-Clemente, R. y Izquierdo, G. (2020). Revisión sistemática de la aplicación de algoritmos de «machine learning» en la esclerosis múltiple. *Neurología*. Recuperado de: <https://doi.org/10.1016/j.nrl.2020.10.017>
- Wu, Y. (2020). Bank Risk Control based on FPGA and Machine Learning. *Revista Microprocessors and Microsystems*. Recuperado de: <https://doi.org/10.1016/j.micpro.2020.103472>

- Yap, B. y Sim, C. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141-2155. Recuperado de: <https://doi.org/10.1080/00949655.2010.520163>
- Young, A., Majehrzak y Kane, G. (2021). Organizing workers and machine learning tools for a less oppressive workplace. *Revista International Journal of Information Management*. Recuperado de: <https://doi.org/10.1016/j.ijinfomgt.2021.102353>
- Zhao, S., Zhang, S., Liu, J., Wang, H., Zhu, J., Li, D. y Zhao, R (2021). Application of machine learning in intelligent fish aquaculture: A review . *Revista Aquaculture* (540). Recuperado de: <https://doi.org/10.1016/j.aquaculture.2021.736724>
- Zhao, T., Li, Y., Zuo, L. y Zhang, K. (2021). Machine-learning optimized method for regional control of sound fields. *Revista Materials Today: Proceedings*. Recuperado de: <https://doi.org/10.1016/j.eml.2021.101297>

ANEXOS

Anexo 1: Matriz de Consistencia

TÍTULO: Machine Learning en la mejora de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021				
AUTOR: Mirtha Jesús Quipas Bellizza				
PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES	
<p>Problema principal: ¿De qué manera Machine Learning mejora la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?</p> <p>Problemas específicos: PE1: ¿De qué manera Machine Learning mejora el tiempo de ejecución de la Gestión del Conocimiento, en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?</p>	<p>Objetivo principal: Determinar que Machine Learning mejora la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p> <p>Objetivos específicos: OE1: Determinar que Machine Learning mejora el tiempo de ejecución de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p>	<p>Hipótesis principal: Machine Learning mejora significativamente la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p> <p>Hipótesis específicos: HE1: Machine Learning mejora significativamente el tiempo de ejecución de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p>	Variable Independiente: Machine Learning	
			Variable Dependiente: Gestión del Conocimiento	
			Indicadores	Unidad de medida
			Tiempo de ejecución	Tiempo
			Índice de expedientes procesados	Porcentaje
			Índice de expedientes procesados en el tiempo previsto	Porcentaje

TÍTULO: Machine Learning en la mejora de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021

AUTOR: Mirtha Jesús Quipas Bellizza

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES
<p>PE2: ¿De qué manera Machine Learning mejora el índice de expedientes procesados de la Gestión del Conocimiento, en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?</p> <p>PE3: ¿De qué manera Machine Learning mejora el índice de expedientes procesados en el tiempo previsto de la Gestión del Conocimiento, en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021?</p>	<p>OE2: Determinar que Machine Learning mejora el índice de expedientes procesados de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p> <p>OE3: Determinar que Machine Learning mejora el índice de expedientes procesados en el tiempo previsto de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p>	<p>HE2: Machine Learning mejora significativamente el índice de expedientes procesados de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p> <p>HE3: Machine Learning mejora significativamente el índice de expedientes procesados en el tiempo previsto de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021.</p>	

Metodología

TIPO Y DISEÑO	POBLACIÓN Y MUESTRA	TÉCNICAS E INSTRUMENTOS	ESTADÍSTICA POR UTILIZAR
<p>Tipo: Aplicada</p> <p>Diseño: Experimental – Experimental puro</p>	<p>Población: 50 Observaciones</p> <p>Muestreo: Probabilístico del tipo Aleatorio simple.</p>	<p>Técnicas: Observación</p> <p>Instrumentos: Guía de Observación</p>	<p>Descriptiva: Para el análisis descriptivo, se usará tablas y figuras, exponiendo medidas de tendencia central usando la media, se realizará su interpretación o lectura por cada indicador, datos emitidos por el instrumento, lo cual ayudará a fijar de manera visual y estructurada la comprensión sencilla de todos los datos numéricos.</p> <p>Inferencial: Para el análisis inferencial, se comprobó la normalidad de los datos obtenidos mediante la prueba Test de Shapiro Wilk; Además, se utilizó para la contratación de la hipótesis las pruebas de los rangos con signo de Wilcoxon.</p>

Anexo 2: Matriz de Operacionalización de Variables

TÍTULO: Machine Learning en la mejora de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021. AUTOR: MIRTHA JESUS QUIPAS BELLIZZA				
INDICADOR	DEFINICIÓN	INSTRUMENTO	UNIDAD DE MEDIDA	FÓRMULA
Tiempo de ejecución	Según Álvarez y Lobo (2020), Beltrán (2017) el tiempo de ejecución corresponde al tiempo que se emplea para realizar diversas actividades que forman parte del proceso de Gestión de Conocimiento y que se encuentra enmarcado dentro de un periodo final y un periodo anterior o inicial, es decir desde el registro expediente hasta su calificación asimismo, este indicador se utiliza para medir la rapidez del proceso de gestión.	Guía de observación	Tiempo	$x = \text{Tiempo Final} - \text{Tiempo de Inicio}$
Índice de expedientes procesados	La RDE No.38-2019-MEF (2018), indica que los expedientes debidamente numerados y generados en un sistema informático de la institución, que se encuentran asociados a un tema o asunto específico, que permite dar inicio a la atención y pasar por varias etapas dentro del proceso hasta que finaliza cuando el área haya cumplido con el propósito de la atención y comunique el resultado al usuario.	Guía de observación	Porcentaje	$x = \left(\frac{\text{Expedientes procesados}}{\text{Expedientes pendientes de atención}} \right) * 100$
Índice de expedientes procesados en el tiempo previsto.	La RDE No.38-2019-MEF (2018), Calvo <i>et al</i> (2018) indican que los expedientes procesados son un conjunto de documentos electrónicos debidamente numerados y generados en un sistema informático de la institución, que se encuentran asociados a un tema o asunto específico, que permite dar inicio a la atención y pasar por varias etapas dentro del proceso hasta que finaliza cuando el área haya cumplido con el propósito de la atención y comunicación del resultado al usuario, de acuerdo a la meta esperada en un tiempo determinado.	Guía de observación	Porcentaje	$x = \left(\frac{\text{Expedientes procesados en el tiempo previsto}}{\text{Expedientes pendientes de atención}} \right) * 100$

Anexo 3: Instrumento de Recolección de Datos

Guía de observación N° 1: Tiempo de ejecución

Guía de observación de medición del indicador Tiempo de ejecución					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión de Conocimiento		
Pre-Test					
N° de Obs.	Turno	Fecha	Tiempo de inicio	Tiempo final	Tiempo de ejecución = Tiempo final – Tiempo de inicio
1					
2					
3					
4					
5					
6					
...					
n					

Guía de observación de medición del indicador Tiempo de ejecución					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión de Conocimiento		
Post-Test					
N° de Obs.	Turno	Fecha	Tiempo de inicio	Tiempo final	Tiempo de ejecución = Tiempo final – Tiempo de inicio
1					
2					
3					
4					
5					
6					
...					
n					

Guía de observación N° 2. Índice de expedientes procesados

Guía de observación de medición del indicador índice de expedientes procesados					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión del Conocimiento		
Pre-Test					
N° de Obs.	Turno	Fecha	Expedientes procesados	Expedientes pendientes de atención	Índice de expedientes procesados = (expedientes procesados/expedientes pendientes de atención) *100
1					
2					
3					
4					
5					
6					
...					
n					

Guía de observación de medición del indicador índice de expedientes procesados					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión de Conocimiento		
Post-Test					
N° de Obs.	Turno	Fecha	Expedientes procesados	Expedientes pendientes de atención	Índice de expedientes procesados = (expedientes procesados/expedientes pendientes de atención) *100
1					
2					
3					
4					
5					
6					
...					
n					

Guía de observación N° 3. Índice de expedientes procesados en el tiempo previsto.

Guía de observación de medición del indicador índice de expedientes procesados en el tiempo previsto.					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión del conocimiento		
Pre-Test					
N° de Obs.	Turno	Fecha	Expedientes procesados en el tiempo previsto	Expedientes pendientes de atención	Índice de expedientes procesados en el tiempo previsto= (expedientes procesados en el tiempo previsto / expedientes pendientes de atención)* 100
1					
2					
3					
4					
5					
6					
...					
n					

Guía de observación de medición del indicador índice de expedientes procesados en el tiempo previsto.					
Investigador:			Mirtha Quipas		
Proceso observado:			Gestión del conocimiento		
Post-Test					
N° de Obs.	Turno	Fecha	Expedientes procesados en el tiempo previsto	Expedientes pendientes de atención	Índice de expedientes procesados en el tiempo previsto= (expedientes procesados en el tiempo previsto / expedientes pendientes de atención)* 100
1					
2					
3					
4					
5					
6					
...					
n					

Anexo 4: Certificado de Validación del Instrumento de Recolección de Datos

Validación del Experto N°1

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: GESTIÓN DEL CONOCIMIENTO

N°	INDICADORES	Claridad ¹		Pertinencia ²		Relevancia ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO DE EJECUCIÓN Fórmula: Tiempo Final – Tiempo de Inicio	X		X		X		
2	INDICADOR 2: INDICE DE EXPEDIENTES PROCESADOS Fomula: (Expedientes procesados / expedientes pendientes de atención) * 100	X		X		X		
3	INDICADOR 3: INDICE DE EXPEDIENTES PROCESADOS EN EL TIEMPO PREVISTO. Fomula: (Expedientes procesados en el tiempo previsto / Expedientes pendientes de atención) * 100	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable [x] Aplicable después de corregir [] No aplicable []

21.de Mayo del 2021

Apellidos y nombres del juez evaluador: DEL CARPIO SALINAS JORGE ALBERTO DNI: 10271988

Especialista: Metodólogo [x] Temático []

Grado: Maestro [] Doctor [X]

¹ Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

² Pertinencia: Si el ítem pertenece a la dimensión.

³ Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión



Firma del Experto Informante

Validación del Experto N°2

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: GESTIÓN DEL CONOCIMIENTO

N°	INDICADORES	Claridad ¹		Pertinencia ²		Relevancia ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO DE EJECUCIÓN Fórmula: Tiempo Final – Tiempo de Inicio	X		X		X		
2	INDICADOR 2: ÍNDICE DE EXPEDIENTES PROCESADOS Formula: (Expedientes procesados / expedientes pendientes de atención) * 100	X		X		X		
3	INDICADOR 3: ÍNDICE DE EXPEDIENTES PROCESADOS EN EL TIEMPO PREVISTO. Formula: (Expedientes procesados en el tiempo previsto / Expedientes pendientes de atención) * 100	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable Aplicable después de corregir No aplicable

21.de Mayo del 2021

Apellidos y nombres del juez evaluador: PEREZ GRÁNDEZ VIOLETA DNI: 42248986

Especialista: Metodólogo Temático

Grado: Maestro Doctor

¹ Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

² Pertinencia: Si el ítem pertenece a la dimensión.

³ Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión


Firma del Experto Informante

Validación del Experto N°3

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: GESTIÓN DEL CONOCIMIENTO

N°	INDICADORES	Claridad ¹		Pertinencia ²		Relevancia ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO DE EJECUCION Fórmula: Tiempo Final – Tiempo de Inicio	X		X		X		
2	INDICADOR 2: ÍNDICE DE EXPEDIENTES PROCESADOS Formula: (Expedientes procesados / expedientes pendientes de atención) * 100	X		X		X		
3	INDICADOR 3: ÍNDICE DE EXPEDIENTES PROCESADOS EN EL TIEMPO PREVISTO. Formula: (Expedientes procesados en el tiempo previsto / Expedientes pendientes de atención) * 100	X		X		X		

Observaciones (precisar si hay suficiencia): SUFICIENTE

Opinión de aplicabilidad: Aplicable [X] Aplicable después de corregir [] No aplicable []

29 de mayo del 2021

Apellidos y nombres del juez evaluador: Visurraga Agüero Joel Martín

DNI: 10192315

Especialista: Metodólogo [X] Temático [X]

Grado: Maestro [] Doctor [X]

¹ Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

² Pertinencia: Si el ítem pertenece a la dimensión.

³ Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión



Dr. Joel Martin Visurraga Agüero

Anexo 5: Base de datos

	Tiempo de ejecución		Índice de expedientes procesados		Índice de expedientes procesados en el tiempo previsto	
	I1PreTest	I1PostTest	I2PreTest	I2PostTest	I3PreTest	I3PostTest
1	115	66	93.33	100	93.33	93.33
2	87	37	66.67	93.33	66.67	93.33
3	162	85	62.5	87.5	56.25	81.25
4	12	4	50	93.75	25	87.5
5	19	14	57.14	76.47	28.57	64.71
6	361	77	50	93.75	25	87.5
7	210	68	57.89	84.21	36.84	63.16
8	31	14	100	93.75	60	87.5
9	275	64	65.22	83.33	30.43	66.67
10	70	40	100	87.5	100	81.25
11	25	8	69.57	83.33	34.78	66.67
12	182	87	66.67	94.12	66.67	82.35
13	183	80	54.55	94.44	22.73	77.78
14	210	115	40	87.5	15	81.25
15	268	117	48	93.75	8	87.5
16	19	6	74.07	88.24	29.63	76.47
17	16	4	64.29	87.5	17.86	81.25
18	19	8	45.45	76.47	13.64	64.71
19	27	14	72	88.24	32	76.47
20	18	4	74.07	84.21	29.63	63.16
21	26	8	63.64	88.24	31.82	76.47
22	29	14	54.55	88.89	22.73	72.22
23	29	14	43.48	82.35	8.7	70.59
24	28	10	40	88.24	0	76.47
25	20	8	60.71	88.89	14.29	72.22
26	18	13	33.33	94.12	0	82.35
27	30	9	90.91	82.35	90.91	70.59
28	25	8	75	100	50	93.75
29	27	14	93.75	83.33	87.5	66.67
30	253	119	100	93.33	75	93.33
31	27	12	62.5	88.89	90.91	72.22
32	24	11	80	100	80	93.75
33	37	17	57.14	88.24	68.75	76.47

	Tiempo de ejecución		Índice de expedientes procesados		Índice de expedientes procesados en el tiempo previsto	
	I1PreTest	I1PostTest	I2PreTest	I2PostTest	I3PreTest	I3PostTest
34	16	7	50	93.33	33.33	93.33
35	10	6	41.67	88.24	31.58	76.47
36	31	12	58.33	87.5	16.67	81.25
37	21	7	31.03	88.24	0	76.47
38	22	9	84	94.12	44	82.35
39	36	21	40	88.24	0	76.47
40	397	87	100	100	78.95	93.75
41	31	7	50	94.12	6.45	82.35
42	31	5	93.33	100	93.33	100
43	34	12	33.33	93.75	0	87.5
44	592	148	100	93.33	93.75	93.33
45	22	6	43.24	100	0	93.75
46	22	8	86.67	87.5	86.67	81.25
47	28	16	69.44	100	29.03	100
48	30	6	82.35	82.35	70.59	70.59
49	13	5	53.85	100	38.1	100
50	33	13	88.89	100	72.22	83.33

Anexo 6: Autorización de la investigación



CONSTANCIA DE AUTORIZACIÓN DE INVESTIGACIÓN

Mediante el presente dejamos constancia que la Srta. Mirtha Jesús Quipas Bellizza identificada con DNI 08715701, cuenta con autorización para llevar a cabo su investigación denominada: "Machine Learning en la mejora de la Gestión del Conocimiento en el Consejo Nacional de Ciencia y Tecnología e Innovación Tecnológica - CONCYTEC, Lima 2021" y aplicar los instrumentos de recolección de datos asociados a ella en nuestra institución, para optar el grado académico de Maestro en ingeniería de sistemas con mención en tecnologías de información en la universidad César Vallejo.

Se expide el presente documento a solicitud del interesado para los fines que estime conveniente.

Lima, 11 de Junio del 2021



.....
Ing. Juan Manuel Rojas Ochante
(Firma)
Oficina de Tecnologías de Información - OTI
CONSEJO NACIONAL DE CIENCIA, TECNOLOGÍA
E INNOVACIÓN TECNOLÓGICA

Anexo 7: Prueba de Normalidad

Para la prueba de normalidad se aplicó Shapiro-Wilk, ya que es considerada como una de las pruebas estadísticas más sólidas y que presenta mayor potencia estadística que otras; asimismo, muestra mejores resultados con tamaños de muestras hasta 50 (Pedrosa, Juarros-Basterretxea, Robles-Fernández, Basteiro y García-Cueto, 2015); (Yap y Sim, 2011; Royston, 1982).

Prueba normalidad del indicador 1: Tiempo de ejecución

Formulación de hipótesis estadística:

H₀: Los datos del indicador tiempo de ejecución presentan una distribución normal.

H₁: Los datos del indicador tiempo de ejecución no presentan una distribución normal.

Tabla 1

Pruebas de normalidad del indicador tiempos de ejecución

	Estadístico	Shapiro-Wilk	
		gl	Sig.
Tiempo de ejecución Pre-test	0.624	50	0.000
Tiempo de ejecución Pos-test	0.699	50	0.000

Nota: Datos asistidos en el software IBM SPSS V25.

De acuerdo al resultado de la prueba Shapiro Wilk podemos observar que el valor de significancia de los datos del indicador tiempo de ejecución fue 0.000 para el pre-test y el pos-test, por lo que siendo estos valores menores al error asumido de 0.05 se rechaza la hipótesis nula (H₀) y se acepta la hipótesis alterna (H₁), es decir que el indicador tiempo de ejecución no presenta una distribución normal.

Prueba normalidad del indicador 2: Índice de expedientes procesados

Formulación de hipótesis estadística:

H₀: Los datos del indicador índice de expedientes procesados presentan una distribución normal.

H₁: Los datos del indicador expedientes procesados no presentan una distribución normal.

Tabla 2

Pruebas de normalidad del indicador índice de expedientes procesados

	Shapiro-Wilk		
	Estadístico	gl	Sig.
Índice de expedientes procesados Pre-test	0.951	50	0.036
Índice de expedientes procesados Pos-test	0.932	50	0.007

Nota: Datos asistidos en el software IBM SPSS V25.

De acuerdo al resultado de la prueba Shapiro Wilk podemos observar que el valor de significancia de los datos del indicador índice de expedientes procesados fue 0.036 para el pre-test y de 0.007 para el post-test, y dado que ambos no son superiores al error asumido de 0.05, se rechaza la hipótesis nula (H₀) y se acepta la hipótesis alterna (H₁), es decir que el indicador índice de expedientes procesados no presenta una distribución normal.

Prueba normalidad del indicador 3: Índice de expedientes procesados en el tiempo previsto

Formulación de hipótesis estadística:

H₀: Los datos del indicador índice de expedientes procesados en el tiempo previsto presentan una distribución normal.

H₁: Los datos del indicador índice de expedientes procesados en el tiempo previsto no presentan una distribución normal.

Tabla 3

Pruebas de normalidad del indicador índice de expedientes procesados en el tiempo previsto.

	Estadístico	Shapiro-Wilk gl	Sig.
Índice de expedientes procesados en el tiempo previsto Pre-test	0.916	50	0.002
Índice de expedientes procesados en el tiempo previsto Pos-test	0.956	50	0.061

Nota: Datos asistidos en el software IBM SPSS V25.

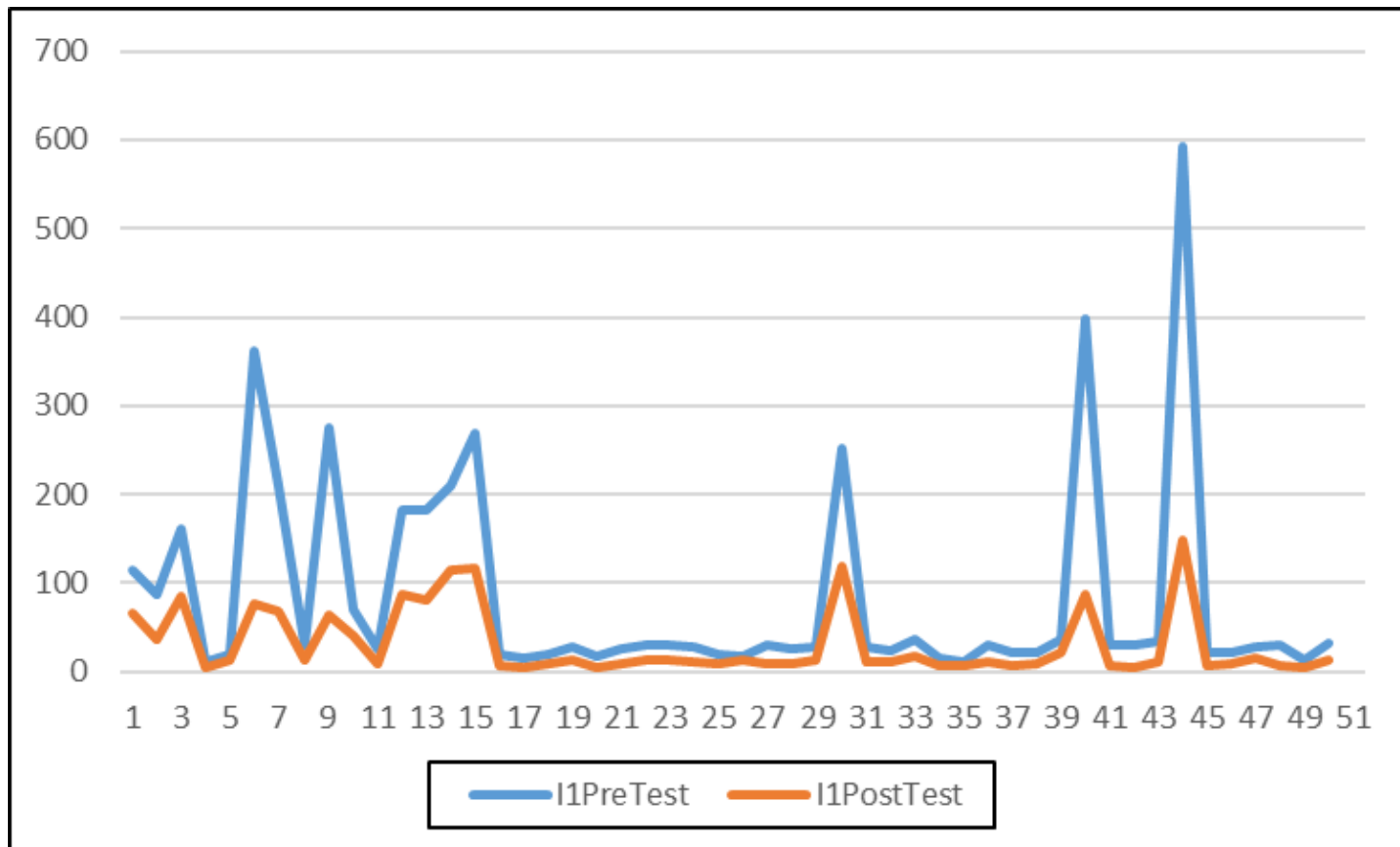
De acuerdo al resultado de la prueba Shapiro Wilk podemos observar que el valor de significancia de los datos del indicador índice de expedientes procesados en el tiempo previsto fue 0.002 para el pre-test y de 0.061 para el pos-test, por lo que dado que ambos no son superiores al error asumido de 0.05, se rechaza la hipótesis nula (H₀) y se acepta la hipótesis alterna (H₁), es decir que el indicador índice de expedientes procesados en el tiempo previsto no presenta una distribución normal.

Anexo 8: Comportamiento de las medias descriptivas

a) Indicador 1: Tiempo de ejecución

Figura 1

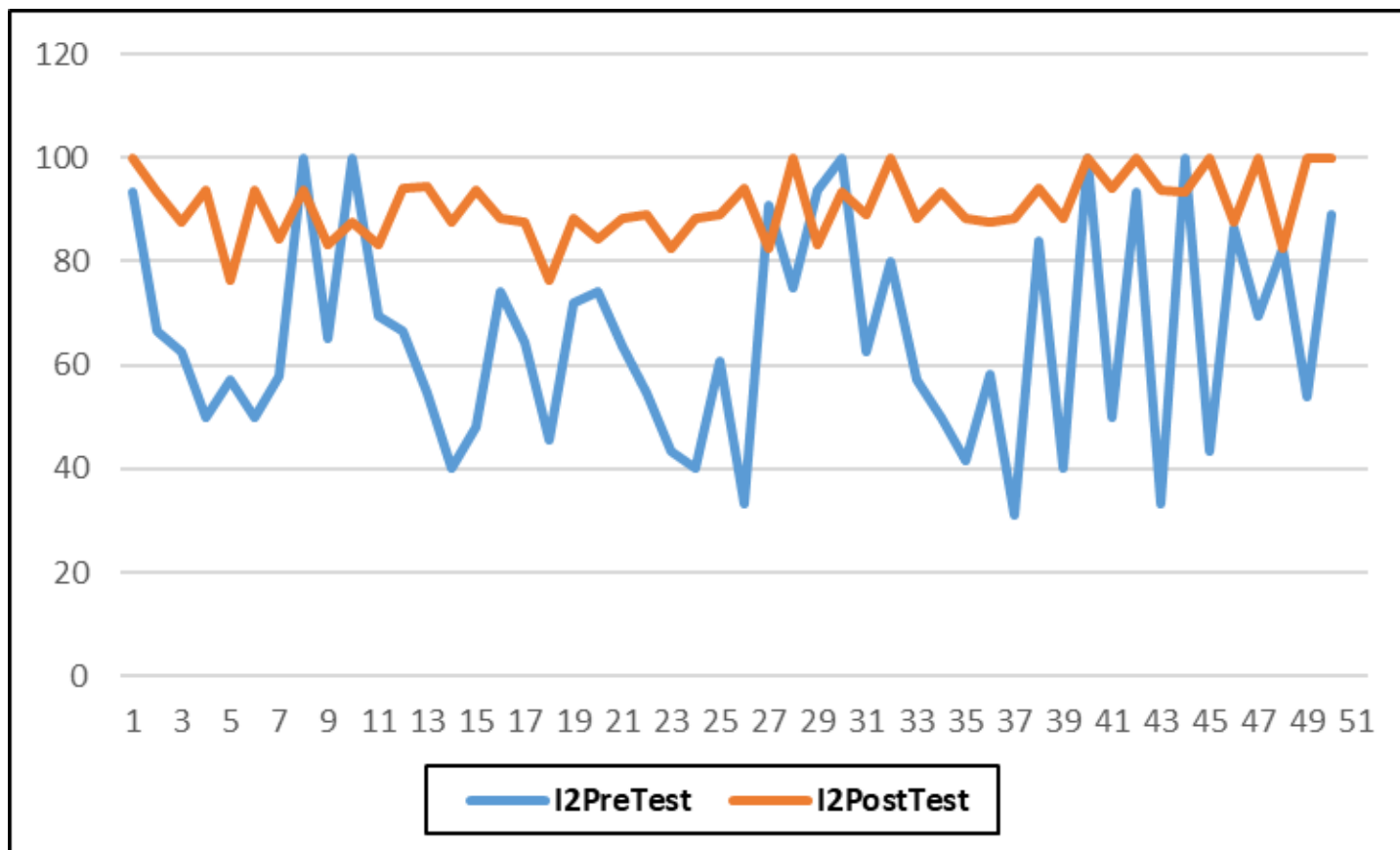
Comparación del comportamiento del indicador tiempo de ejecución



b) Indicador 2: Índice de expedientes procesados

Figura 2

Comparación del comportamiento del indicador índice de expedientes procesados



c) **Indicador 3: Índice de expedientes procesados en el tiempo previsto.**

Figura 3

Comparación del comportamiento del indicador índice de expedientes procesados en el tiempo previsto.

