UNIVERSIDAD NACIONAL DEL ALTIPLANO

FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA, ELECTRÓNICA Y SISTEMAS

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

ANÁLISIS DE LA GESTIÓN DOCUMENTARIA MEDIANTE EL ÁRBOL DE DECISIONES J48 PARA UNA DATA SET NORMALIZADA USANDO LA HERRAMIENTA WEKA, EN LA UGEL CHUCUITO, JULI 2023.

TESIS

PRESENTADA POR:

Bach. JHENERY ANYELA CARBAJAL PARI

Bach. ROGER WILBER TAPIA BARRIOS

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

PUNO – PERÚ

2024

JHENERY ANYELA CARBAJAL PARI ROGER WILBE...

ANÁLISIS DE LA GESTIÓN DOCUMENTARIA MEDIANTE EL ÁRBOL DE DECISIONES J48 PARA UNA DATA SET NORMALIZ...

Universidad Nacional del Altiplano

Detalles del documento

Identificador de la entrega
trn:oid:::8254:415836766

Fecha de entrega
13 dic 2024, 12:24 p.m. GMT-5

Fecha de descarga
13 dic 2024, 12:26 p.m. GMT-5

Nombre de archivo
ANÁLISIS DE LA GESTIÓN DOCUMENTARIA MEDIANTE EL ÁRBOL DE DECISIONES J48 PARA UNA D....pdf

Tamaño de archivo
2.3 MB

110 Páginas

18,939 Palabras

109,962 Caracteres

turnitin Página 1 of 116 - Portada Identificador de la entrega trn:oid:::8254:415836766

16% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Coincidencias menores (menos de 12 palabras)

Fuentes principales

14%	Fuentes de Internet
1%	Publicaciones
5%	Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Dr. Miguel Romilio Aceituno Rojo
INGENIERO DE SISTEMAS

V°B°
Dra. Qulina Guadalupe Sotomayor Alzamora
INGENIERO DE SISTEMAS

turnitin Página 2 of 116 - Descripción general de integridad Identificador de la entrega trn:oid:::8254:415836766

DEDICATORIA

A mi papá, por ser mi ejemplo de esfuerzo y perseverancia. Gracias por apoyarme y enseñarme a nunca rendirme.

A mi mamá, mi mayor pilar y apoyo incondicional. Te agradezco por tu amor, tu paciencia y por creer en mí incluso cuando yo no lo hacía.

A mi hermanita, por llenar mi vida de alegría y por recordarme lo importante que es soñar.

Y a mis queridas mascotas, por acompañarme en mis momentos de soledad y estrés. Sin su cariño y comprensión, este logro no habría sido posible.

Jhenery Anyela Carbajal Pari

DEDICATORIA

A mis padres, quienes han sido mi mayor fuente de motivación y fortaleza, brindándome su apoyo constante y sabios consejos, permitiéndome superar cada desafío con determinación.

A mi hermana, cuyo respaldo incondicional y aliento han sido fundamentales en cada etapa de este proceso.

A mis docentes y asesor, por la invaluable transmisión de conocimientos y la orientación precisa que me han guiado a lo largo de mi formación profesional.

Roger Wilber Tapia Barrios

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a la UGEL Chucuito, por brindarme la oportunidad de llevar a cabo este proyecto de investigación, facilitando el acceso a los recursos y el apoyo necesario para el desarrollo de la misma.

Mi agradecimiento también va dirigido a la Universidad Nacional del Altiplano, por proporcionar un entorno académico que favorece la investigación y el aprendizaje, permitiéndome desarrollar mis habilidades y conocimientos.

Finalmente, mi más profundo agradecimiento a mi asesor y a los miembros del jurado, por su invaluable orientación, apoyo constante y conocimientos, que fueron fundamentales para el éxito de este trabajo. Su dedicación y compromiso han sido esenciales en todo el proceso de investigación.

Jhenery Anyela Carbajal Pari

Roger Wilber Tapia Barrios

ÍNDICE GENERAL

Pág.

CAPÍTULO II

REVISIÓN DE LITERATURA

CAPÍTULO III

MATERIALES Y MÉTODOS

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

ÍNDICE DE TABLAS

Pág.

ÍNDICE DE FIGURAS

Pág.

ÍNDICE DE ANEXOS

Pág.

ACRÓNIMOS

UGEL:	Unidad de Gestión Educativa Local
ISO:	Organización Internacional de Normalización
BPMI:	Business Process Management Initiative
OMG:	Object Management Group
JDBC:	Java Database Connectivity
WEKA:	Waikato Environment for Knowledge Analysis

RESUMEN

En los tiempos actuales, la importancia de automatizar los procesos ha cobrado importancia en las organizaciones, con el fin de aumentar el rendimiento y eficiencia de la entidad, que puede ser tomado desde una percepción de la parte física (hardware) o parte lógica (software). Por ello, este estudio tuvo como objetivo realizar mejoras de la gestión documentaria mediante el árbol de decisiones J48 en la UGEL Chucuito, Juli, 2023. Para este estudio, se empleó una metodología de tipo aplicada de nivel descriptivo y explicativo, de diseño experimental de pre y post test. La población lo conformaron 1000 documentos; y, la muestra, 350 documentos. Para la obtención de la información, se dio a partir de los documentos ingresados a la entidad, los mismos que fueron procesados para adaptarse a la estructura de la información del árbol J48; posteriormente, se programó y se evaluó la calidad del programa bajo la norma ISO 25000. Los resultados indicaron que el árbol de decisión J48 mejore ligeramente la confiabilidad del ordenamiento de los documentos en la entidad. Se concluyo que la implementación del árbol de decisiones J48 en la gestión documentaria de la UGEL Chucuito, usando un data set normalizado en WEKA, ha mostrado buenos resultados. Con una exactitud del 84.29 %, el modelo es efectivo en clasificar documentos. Sin embargo, la tasa de error de 15.71 % sugiere que se pueden realizar ajustes para mejorar la precisión. En general, el modelo es exitoso y optimizable para una mejor clasificación.

Palabras clave: Árbol de decisión, J48, Gestión documentaria, automatización, configuración, clasificación, Machine Learning.

ABSTRACT

In current times, the importance of automating processes has gained importance in organisations, in order to increase the performance and efficiency of the entity, which can be taken from a perception of the physical part (hardware) or logical part (software). Therefore, this study aimed to make improvements in document management using the J48 decision tree in the UGEL Chucuito, Juli, 2023. For this study, a descriptive and explanatory applied methodology was used, with a pre- and post-test experimental design. The population consisted of 1000 documents and the sample consisted of 350 documents. The information was obtained from the documents entered into the entity, which were processed to adapt them to the information structure of the J48 tree; subsequently, the quality of the programme was programmed and evaluated under the ISO 25000 standard. The results indicated that the J48 decision tree slightly improves the reliability of the organisation's document management. It was concluded that the implementation of the J48 decision tree in the document management of the UGEL Chucuito, using a normalised data set in WEKA, has shown good results. With an accuracy of 84.29 %, the model is effective in classifying documents. However, the error rate of 15.71 % suggests that adjustments can be made to improve accuracy. Overall, the model is successful and optimisable for better classification.

Keywords: Decision tree, J48, Document management, automation, configuration, classification, Machine Learning.

CAPÍTULO I

INTRODUCCIÓN

En la actualidad, el Perú está saliendo poco a poco de las restricciones que produjo la pandemia, con base en ello, hoy en día se han desarrollado programas de seguimiento, monitoreo y administración de documentos, los mismos que son empleados en la UGEL-Chucuito, lo que permitirá la identificación, ordenamiento y monitoreo el estado de un documento, que es ingresado a través de la mesa partes presencial y/o virtual. Así, cada año que pasa, la cantidad de documentos que se registran en la UGEL-Chucuito son mayores, los mismos que deben ser atendidos por el personal encargado de mesa de partes y estar debidamente foliados y clasificados. Ante esto, una solución a la alta carga documentaria, es realizar una automatización de los procesos, para ello se utilizará una herramienta que clasifique correctamente los documentos según sea su prioridad y clasificación.

Para lograr esto, se requiere la automatización de los procedimientos con el objetivo de elevar la calidad del servicio proporcionado a los ciudadanos. Esto implica la búsqueda de herramientas informáticas o algoritmos, siendo estos últimos imparciales y siempre tomando decisiones de manera racional. Cumplen sus funciones de manera rápida y eficiente, de acuerdo con la programación recibida.

Por esa razón, se adopta la propuesta de optimizar la administración mediante utilización del J48-árbol de decisión, una metodología apropiada para clasificar de manera precisa los documentos de la entidad.

Esta investigación comprende los siguientes puntos:

El estudio se estructura en cinco capítulos. El Capítulo I introduce el análisis del problema. En el Capítulo II, se define el problema, se establecen los objetivos, se formulan hipótesis y se justifica la investigación. El Capítulo III aborda una revisión sistemática documental, centrada en el estado actual del conocimiento, con artículos científicos y estudios previos a nivel internacional, nacional y local. El Capítulo IV presenta los materiales, métodos, variables y la matriz de operacionalización. Finalmente, en el Capítulo V se analizan los datos, se comparan las hipótesis y se discuten conclusiones y recomendaciones para futuras investigaciones.

1.1. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

1.1.1. Descripción del problema

En el contexto global, estamos siendo testigos de un notable progreso tecnológico, donde las Tecnologías de la Información ocupan un papel fundamental importante en el ámbito diario. La continua evolución tecnológica y el crecimiento vertiginoso para el almacenamiento en los dispositivos informáticos, combinados con el aumento en la cantidad de datos gestionados por las organizaciones, introducen complejidades adicionales en el análisis del proceso e interpretación informática.

Mediante el análisis de los patrones de eventos y la aplicación de técnicas, algoritmos y procesos de validación en el ámbito de la minería de datos, es factible obtener datos sobre el origen de los incidentes informáticos y las estrategias para hacerles frente. Es esencial para una organización que los sistemas informáticos se mantengan operativos la mayor parte del tiempo, asegurando así la completa disponibilidad de estos recursos y la preservar la cualidad servicios que se brindan (Garces, 2019).

Según Henriquez (2022) en USA las tecnologías de la información han mejorado tanto que se han diseñado software con inteligencia artificial que permiten a las empresas

realizar ordenamiento bases de datos, traducciones, cálculos econométricos, etc. Lo que ha permitido a las empresas tener mayor rentabilidad ya que reducirá personal además de reducir procesos trayendo beneficio a sus clientes.

Según Cordero et al. (2020) en Europa los grandes avances tecnológicos han permitido a la industria, construcción, financiero, etc. salir de las crisis por ello cuando se produce la pandemia no les tomo mucho tiempo adaptarse a las nuevas plataformas tecnológicas lo que ayudo a que la vida se realice en forma cotidiana.

Según Barrientos et al. (2019) mientras que en América Latina en especial Colombia hemos visto que las TIC´s sobre la inteligencia artificial han mejorado notablemente ya que el sistema estatal y la empresa privada han automatizado sus procesos están entre uno de los países de la región que más adelantos tecnológicos presenta, es por ello que cuando ocurrió la pandemia les fue fácil asimilarse ya que los servicios estaban automatizados.

Según Anaya et al. (2021) en el Perú la época de la pandemia ha mostrado lo atrasado que se encuentra el país en las innovaciones tecnológicas ya que las instituciones públicas y privadas no estaban en la capacidad de atender al público en forma virtual, lo que produzco en un inicio una serie de inconvenientes por parte de la gran mayoría personas que no sabían utilizar dichas plataformas tecnológicas.

Según Ruelas (2023) en la región Puno el gran problema que tiene el ciudadano es la alta brecha tecnológica en las entidades públicas y privadas lo que demostró durante la pandemia al no poder realizar sus operaciones o actividades con normalidad lo que significó grandes pérdidas económicas.

En cuanto a la gestión documentaria, a través de un aspecto crítico en muchas organizaciones, puede presentarse de diferentes tipos de problemas como es la falta de

una estructura organizada para llevar los documentos evitando la pérdida de información importante, dificulta el ingreso que se necesita, conjunto con la ausencia de estándares claros para la nomenclatura, formato y almacenamiento de documentos puede generar confusión y/o problemas de búsqueda, finalmente la recuperación de la información.

En el mismo orden de ideas, los procesos para crear, aprobar, revisar y archivar documentos pueden ser ineficientes, lo que lleva a retrasos y pérdida de productividad, esto crea la falta de medidas de seguridad adecuadas lo que puede exponer la información confidencial a riesgos de robo, pérdida o acceso no autorizado. En organizaciones que aún utilizan documentos en papel, la pérdida o deterioro de documentos físicos puede ser un problema importante (Díaz-Landa et al., 2021).

Según Ilyas et al. (2020) a nivel mundial varios sectores que gracias a las TIC´s han experimentado avances exorbitantes, sobre todo en el área de medicina que ahora puede realizar detecciones de enfermedades en tiempo real y mucho más accesible al usuario ya que todos los servicios ahora están automatizados reduciendo tiempos elevando la rentabilidad de las empresas.

Según Valero (2021) en el Perú el desarrollo de las TIC´s ha sido lento, es una rama a la que las instituciones públicas y privadas no le han tomado mucho en cuenta, aspecto que cambia durante la pandemia hubo grandes dificultades para reanudar a las actividades cotidianas.

Según Choque (2023) en Puno es unas regiones que posee una gran brecha digital es por ello que durante la pandemia las entidades tanto públicas como privadas han experimentado pérdidas alrededor de mil millones de soles ya que Puno tiene mucha actividad comercial y financiera, las entidades al no tener sistemas informáticos que

faciliten las actividades de las personas las gestiones eran mucho más tediosas y es por ello las personas iban a entidades donde si contaban con estos servicios.

En la UGEL Chucuito, una de las principales dificultades reside en la ausencia de procesos automatizados, lo cual resultó en la interrupción de sus servicios durante la pandemia y, en algunos casos, la pérdida de documentos. Esta carencia también generó extensas colas, ya que los administrados no recibían respuestas oportunas a sus solicitudes, provocando un colapso en la gestión administrativa. La implementación de sistemas automatizados podría contribuir significativamente a superar estos desafíos, mejorando la eficiencia y la respuesta a las necesidades de la comunidad educativa.

En la actualidad, la UGEL Chucuito enfrenta un desafío significativo en lo que respecta a su gestión administrativa. Se ha observado que la documentación en todas sus áreas se encuentra desorganizada, lo que, en última instancia, provoca retrasos en los servicios a los usuarios. Por tanto, es imperativo implementar una estructura organizativa que permita clasificar la información en función de su grado de prioridad, temática, destinatario y otros criterios relevantes. De esta manera, se logrará agilizar la búsqueda y entrega de información a los ciudadanos, mejorando así la eficiencia de la institución en su conjunto.

Por ende, nace las siguientes preguntas:

1.1.2. Problema general

¿Cómo influye el árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023?

1.1.3. Problemas específicos

- ¿Cómo se puede optimizar el proceso de clasificación de documentos actual, mediante el Diagrama BPMN, en la UGEL Chucuito, Juli 2023?
- ¿Cómo se puede mejorar la precisión y relevancia del modelo predictivo J48 al enriquecer semánticamente un dataset de documentos clasificados Chucuito, mediante el análisis y clasificación de palabras clave en la UGEL Chucuito, Juli 2023?
- ¿Cómo se puede construir el árbol de decisión j48 utilizando la herramienta WEKA para clasificar los documentos en la UGEL Chucuito, Juli 2023?

1.2. JUSTIFICACIÓN DEL PROBLEMA

El árbol de decisión, como herramienta que permite a la I.A. para solucionar la problemática por clasificar, de Chucuito Juli como entidad, que se tiene una gran cantidad de documentos, que por lo general se pierden o son derivados en forma errónea a otra área, originando un gran desorden en la entidad, el árbol puede solucionar el problema de la clasificación de documentos.

Por tanto, es fundamental verificar la confiabilidad de los resultados antes de optar por el uso del árbol de decisión como una alternativa para abordar problemas. En este contexto, se delimita la zona a estudiar por el departamento administrativo-gestión en la entidad. El árbol de decisión específico a examinar es el J48, diseñado para resolver problemas relacionados con la estadística. En la actualidad, resulta crucial realizar una evaluación exhaustiva de las posibles soluciones para abordar problemas, priorizando la mejora de la calidad del servicio en beneficio de los ciudadanos y asegurando el envío

eficiente y organizado de documentos a las oficinas responsables de dar respuesta a cualquier solicitud de los ciudadanos.

La justificación practica al problema es que los usuarios reviven atención en forma manual realizándose procesos largos y tediosos en la que no se asegura que la información este reservada, resguardada y además está expuesta a quedar traspapelada entre miles archivos de usuarios por lo cual se perdería, porque lo que busca esta investigación es la atención del público en forma rápida, segura y eficiente para así mejorar la imagen de la entidad frente a los usuarios.

1.3. OBJETIVOS

1.3.1. Objetivo general

Implementar el árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.

1.3.2. Objetivos específicos

- Describir el proceso de clasificación de documentos actual, mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.
- Preparar la dataset, para el árbol de decisión j48, mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023.
- Construir el árbol de decisión j48 utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.

1.4. HIPÓTESIS

1.4.1. Hipótesis general

La construcción de un árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, mejora a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.

1.4.2. Hipótesis especifica:

- El proceso de clasificación de documentos actual, mejora mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.
- La preparación del dataset, para el árbol de decisión J48, mejora mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023.
- La construcción del árbol de decisión J48 mejora, utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES

2.1.1. Antecedentes internacionales

Zambrano (2020) en el estudio que desarrolló con el objetivo principal de determinar el impacto de la automatización de flujos de gestión en los procesos de negocios a través de la aplicación de árboles de decisión J48. Para llevar a cabo este estudio, se adoptó una metodología de enfoque mixto con un diseño de pre y post prueba. Los resultados indicaron que la implementación del árbol de decisión J48 proporciona información con una precisión significativamente mejorada, especialmente en condiciones de almacenamiento, recepción y accesibilidad, sin comprometer la seguridad de la misma.

Aporte: Se realizó la automatización de la información en cual se clasifico de acuerdo a su orden de importante por lo que la búsqueda información se logró en menos tiempo.

Ru et al. (2022) en su estudio tuvo por finalidad establecer y verificar un modelo predictivo de árbol de decisiones de nacidos vivos en pacientes después de la cirugía por adherencias intrauterinas (AIU) de moderadas a graves. Para este estudio, se utilizó una metodología observacional retrospectiva. La población estuvo conformada por 394 pacientes con AIU; y, la muestra fue censal, conformada por 394 pacientes con AIU. Se concluyó que el modelo predictivo del árbol de decisiones es útil para predecir nacidos vivos después de la cirugía de AIU y que el patrón menstrual postoperatorio es un factor clave en el modelo. Este

modelo ayudará a los médicos a tomar decisiones clínicas adecuadas durante las consultas con los pacientes.

Aporte: Se pudo automatizar la información en forma secuencial y sistematizada según sus características para realizar una revisión sistemática eficaz.

Agenjo y Hernández, (2022) el artículo tiene como objetivo aplicar estrategias de enriquecimiento y reconciliación sistemática de materiales, siguiendo los principios de la Web Semántica, para ofrecer al usuario opciones de exploración de temas igualitarios en sistemas distribuidos en la Red. Estas estrategias combinan operaciones manuales, semiautomáticas y automáticas, y colaboran diferentes programas y protocolos, lo que permite procesar grandes volúmenes de datos de manera eficiente y con costos asumibles para las instituciones culturales. Se implementan según las buenas prácticas del W3C y Europeana, siguiendo los principios FAIR. El artículo también examina vocabularios temáticos en español en Linked Open Data y su relación interlingüística, destacando la importancia de los servicios de comparación estandarizados.

Aporte: Estos sistemas en entidades culturales deben integrar capacidades que aprovechen registros de autoridad de materias enriquecidos para llevar a cabo actividades como búsqueda, navegación, visualización y exportación.

Simón et al. (2022) los autores destacan el crecimiento de los sistemas informáticos de geografía (SIG) y su valor para la toma de decisiones. Mejorar el procesamiento y recuperación de información en SIG es crucial debido a la gran cantidad de datos geográficos disponibles y la falta de formalidad en los conceptos

semánticos. Los métodos propuestos se basan en ontologías geográficas, que definen y facilitan el proceso de manejo de datos. Este enfoque semiautomático utiliza razonamiento basado en CBR (Casos), lo que mejora la capacidad y eficacia de los recursos informáticos en SIG. Este paradigma de vinculación de datos optimiza la gestión de grandes volúmenes de información, aumentando la eficiencia en su procesamiento y recuperación.

Aporte: Este enfoque, que también se implementa en el modelo de gestión de datos basado en ontología propuesto, aborda los problemas de interoperabilidad de las interfaces eléctricas relacionadas con datos geográficos y electrónicos. Además, RBC amplía el formato de la encuesta y las capacidades de recuperación de información.

Muñoz y Moreno (2020), Este proyecto de investigación surge ante la necesidad de brindar apoyo a la comunidad sordomuda en Bogotá, con el propósito de mejorar su interacción con la sociedad y reducir la brecha de comunicación entre las personas con discapacidad auditiva y el resto de la población. Para solucionar este desafío se realizó un experimento basado en aprendizaje profundo, donde se aplicaron diferentes algoritmos informáticos. Estos algoritmos se utilizaron para detectar movimientos y características en videos compuestos de fotogramas, es decir. Imágenes extraídas de vídeos. El conjunto de datos se entrenó y aprendió utilizando una red neuronal convolucional, que clasifica los videos y determina en qué categorías caen las palabras grabadas en lengua de señas colombiana. La propuesta presenta una metodología de última generación que incluye seis pasos principales: creación de conjuntos de datos, preprocesamiento, muestreo, despliegue de redes neuronales, medición y clasificación.

Aporte: En la etapa final, podrá desarrollarse un método de clasificación automática de las cinco palabras utilizadas en el proyecto. Este método proporciona resultados para más del setenta por ciento (70%) de las métricas de rendimiento generadas en las pruebas.

Jiménez (2020), en su trabajo de investigación, la información acerca de las interacciones medicamentosas (DDI) es crucial y valiosa tanto para el personal médico como para los pacientes, ya que proporciona detalles sobre los efectos que pueden surgir durante una terapia cuando se administran simultáneamente varios medicamentos a un paciente. En este estudio, se emplea un modelo convolucional por partes (PCNN) con el fin de capturar de manera eficiente la relación entre entidades farmacológicas, según lo descrito en la literatura biomédica. Adicionalmente, este modelo incorpora palabras de manera multicanal para ampliar el vocabulario y reducir la cantidad de palabras desconocidas. Además, se utiliza el optimizador estocástico Adam para aprender de manera automática los parámetros de la red. También se añade una capa de ruido gaussiano para lograr una extracción efectiva de las relaciones DDI.

Aporte: Los experimentos realizados revelan una mejora en el rendimiento del nuevo modelo en comparación con los modelos encontrados en la literatura técnica actual, específicamente relacionados con el desafío DDI Extraction 2013. Estos resultados son verificables y reproducibles.

Kastrati et al. (2019) En su artículo, el propósito central fue destacar que la mejora del proceso de clasificación se logra mediante el enriquecimiento semántico utilizando técnicas de aprendizaje profundo. Para abordar este objetivo, se implementó una metodología de enfoque mixto. La conclusión principal fue

que el uso del árbol de decisiones contribuye a mejorar el contexto de la información, proporcionándole una mayor seguridad. Este estudio adoptó una metodología de enfoque mixto que incluyó un diseño pre y post prueba.

Aporte: El árbol de decisiones ayuda a que se realiza mejoras toma de decisiones por parte de la organización.

Karabadji et al. (2018) En el desarrollo de este artículo, el objetivo principal fue construir un árbol de decisión, realizando las correcciones necesarias para asegurar su funcionamiento óptimo. Para este estudio, se adoptó una metodología de enfoque mixto que incluyó un diseño pre y post prueba. Los resultados obtenidos indican que la implementación del árbol de decisión J48 proporciona información con una precisión significativamente mejor en condiciones de almacenamiento y recepción. Además, se observó que esta solución es más accesible que las anteriores, sin comprometer su seguridad

Aporte: El árbol de problemas permite que toda organización mejore sus procesos aumentando con ello su productividad.

2.1.2. Antecedentes nacionales

Balbuena (2022), La investigación se centra en la aplicación, evaluación y selección de diversos modelos de redes neuronales recurrentes (RNN) y convolucionales (CNN) con el propósito de reconocer emociones en textos y expresiones faciales. Estos modelos tienen el potencial de integrarse como conectores en agentes de conversación en tiempo real, tales como chatbots o bots sociales. Estos componentes sensoriales permiten a los agentes conversacionales comprender los estados emocionales de las personas durante las interacciones, facilitando así una respuesta empática. En la fase inicial, se realizará una revisión

bibliográfica sobre métodos para mejorar la empatía del chatbot y la detección de emociones a través de distintos canales como el texto y las expresiones faciales. A continuación, se compilarán y procesarán bases de datos públicas a partir de la información obtenida de la revisión de la literatura para entrenar los algoritmos seleccionados.

Aporte: Estas métricas serán analizadas para determinar cuáles son los algoritmos más eficientes y adecuados para la implementación de una aplicación en tiempo real.

Diaz (2021) en su estudio buscó el emplear metodología minería de data conocida como "árbol para decisiones" a fin de que se desarrolle un modelo que visualice el rendimiento académico en la entidad. Este estudio adoptó una metodología de naturaleza aplicada, con un enfoque correlacional, un diseño no experimental y una orientación cuantitativa. La población bajo estudio consistió en 237 alumnos, y se optó por una muestra censal que incluyó a los mismos 237 estudiantes. La recolección de datos se llevó a cabo mediante encuestas utilizando un cuestionario como instrumento. El análisis de la información se llevó a cabo mediante el software WEKA. Los resultados concluyeron que la metodología de "árboles de decisión" puede ser aplicada para desarrollar un enfoque que guarda relación con la mejora del rendimiento académico de los estudiantes en la entidad. Por lo tanto, se establece una conexión entre las variables evaluadas.

Aporte: La utilización del árbol de decisiones J48 permite a las empresas mejorar el tiempo de todos sus procesos y con ello aumentar la rentabilidad de la empresa.

Garcés (2020) En la investigación llevada a cabo, el objetivo consistió en proponer una metodología que incorporara tanto el procesamiento de lenguaje natural como el algoritmo de árbol de decisiones J48. Este estudio siguió una metodología de carácter aplicado, con un enfoque cuantitativo, un nivel correlacional y un diseño no experimental. La conclusión principal resalta que el tiempo empleado para resolver las demandas de los usuarios no es adecuado, lo cual podría acarrear inconvenientes laborales de distintas magnitudes debido al impacto de los incidentes. Además, se subraya la dificultad para planificar o prever la resolución de los incidentes debido a su naturaleza imprevisible.

Aporte: El árbol de decisiones permite realizar de manera correcta mejor aprendizaje en el estudiante por medio de la sistematización de la información.

Alania (2019) El objetivo de la investigación fue prever la deserción estudiantil mediante la implementación del árbol de problemas J48. Este estudio adoptó el método de aplicación de un diseño no experimental y el método de medición cuantitativa del grado de correlación. Las conclusiones obtenidas muestran que la eficacia de los algoritmos de árbol de decisión, en particular C4.5 (J48) y árboles aleatorios, puede evaluarse en términos de precisión, mostrando que ambos algoritmos producen resultados similares.

Aporte: Por medio del árbol de problemas, se puede realizar una aproximación de la deserción estudiantil estadísticas que pueden ser modificadas en tiempo real según el contexto en el que se producen.

Espino y García (2018) La meta de la investigación fue implementar la información basada en el árbol de decisión con el fin de reducir el riesgo de morosidad en la empresa. En este estudio, se utilizó una metodología aplicada,

con un enfoque cuantitativo y un diseño no experimental a nivel correlacional. Como resultado, se determinó que el árbol de decisión mejoró la precisión en el análisis del posible retroceso de los clientes en un 26 %, lo que contribuyó al aumento de las ganancias de la empresa.

Aporte: El árbol de decisiones es necesario, porque ayuda a las empresas financieras a una correcta toma de decisiones además de realizar proyecciones económicas.

2.2. MARCO TEÓRICO

2.2.1. Inteligencia Artificial

Se conceptualiza como ciencia que las nuevas inteligencias, que ayuden al ser humano a solucionar un determinado problema, cuyos procesos son automatizados, rápidos y eficaces (Romero et al., 2007).

Para ello, será aplicada la inteligencia artificial en distintos campos como:

- **Machine learning o aprendizaje automático:** Conceptualiza como ciencia que desarrollo nuevos programas que permite al hombre soluciones rápidas a sus problemas en cualquier ámbito o campo.
- **Sistemas expertos:** Es una rama de la ciencia que se encarga en el desarrollo sistemas especializados de un campo para solución específicas de una realidad problemática.
- **Redes neuronales artificiales:** Son redes que funcionan interconectadas al sistema neurológico de animales para detectar respuestas inmediatas ante un estímulo.

- **Procesamiento del lenguaje natural:** Es una disciplina que se encarga del desarrollo de mecanismo informáticos de comunicación que le permite al hombre intercomunicarse con personas.

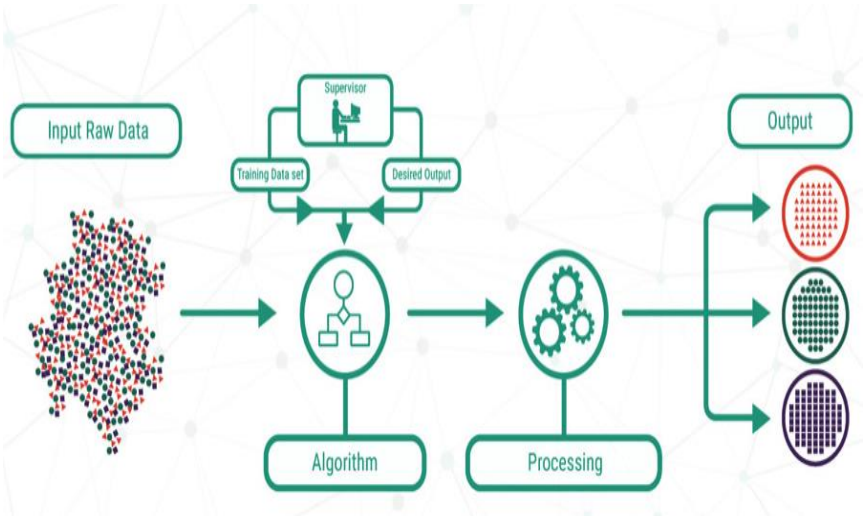
2.2.1.1. Aprendizaje Supervisado

"Es un conjunto de entrenamiento que ejecuta el algoritmo incluyendo las soluciones deseadas, conocidas como etiquetas" (Géron, 2019, p. 10).

Los métodos supervisados de machine learning se utilizan cuando se dispone de datos etiquetados previamente, es decir, datos que ya tienen una respuesta conocida. Por lo que los datos utilizan para entrenar el modelo, con el objetivo de permitir al modelo generalizar y realizar predicciones precisas sobre datos no vistos anteriormente. Los métodos supervisados incluyen árbol para decisiones, bosques aleatorios, SVM, regresión lineal y logística, etc. Como se puede apreciar en la Figura 1 mostrada a continuación:

Figura 1

Aprendizaje Supervisado



Nota: Obtenido de (Kaur et al., 2021)

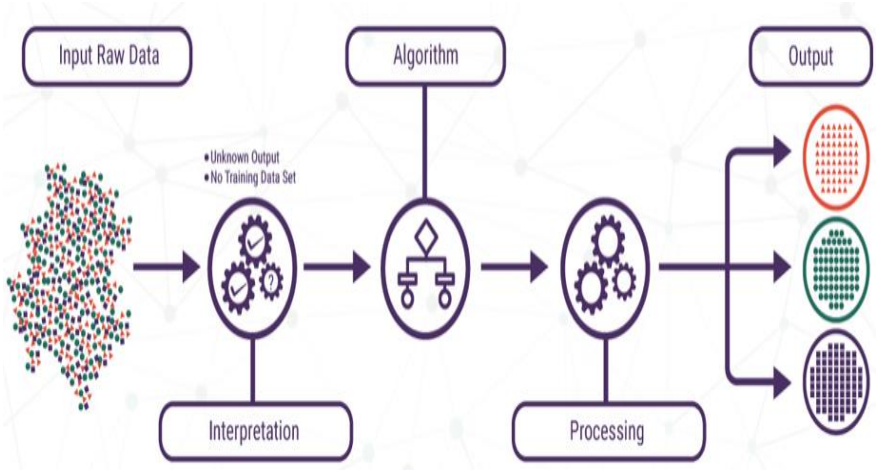
2.2.1.2. Aprendizaje No Supervisado

"Es de esperar que estos datos de entrenamiento tengan etiquetas. Debido que aquí se intenta adquirir conocimientos sin una guía explícita del profesor" (Géron, 2019, p. 12).

En este método, el modelo se entrena utilizando datos sin etiquetar, es decir, datos que no tienen una respuesta conocida. El modelo encuentre patrones y estructuras en los datos por sí mismo. Algunos ejemplos de aplicaciones del aprendizaje no supervisado son: Clasificación, reducción de dimensionalidad, detección de anomalías entre otros. Como se puede apreciar en la Figura 2 mostrada a continuación:

Figura 2

Aprendizaje No Supervisado



Nota: Obtenido de (Kaur et al., 2021)

2.2.1.3. Aprendizaje por Reforzamiento

En el libro publicado por Sutton et al. (2018), se describe es el proceso de aprender qué acciones deben realizarse y cómo asignar situaciones a estas acciones para maximizar las señales digitales de recompensa. En este enfoque, el modelo no recibe instrucciones claras sobre qué acciones debe realizar, sino que debe descubrir por sí mismo a través de las situaciones en momentos que se genera recompensas. Por ende, estas se suman especialmente más interesantes y complicadas, así también de que afectan a la recompensa inmediata sino a un momento a continuación para posteriores recompensas. Lo que estas cualidades permiten una búsqueda mejor en error, ensayo y recompensa para reforzar estos momentos en diversos indoles del aprendizaje por reforzamiento (p. 23).

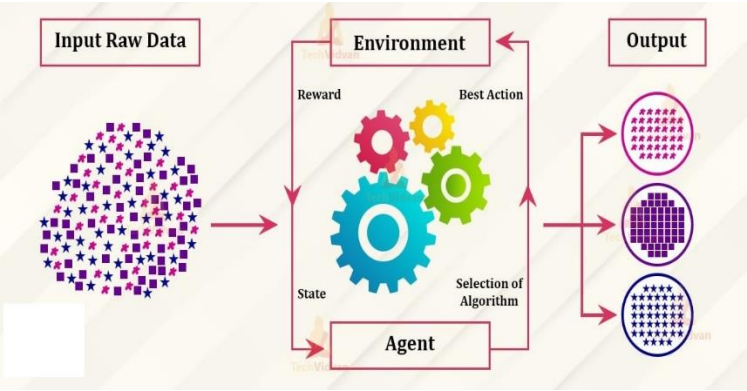
Algunos ejemplos de aplicaciones del aprendizaje por refuerzo son:

- Juegos de mesa: por ejemplo, entrenar un modelo para jugar ajedrez o Go.
- Robótica: por ejemplo, entrenar un robot para realizar tareas complejas como caminar o manipular objetos.
- Control de procesos: por ejemplo, controlar una planta de energía para maximizar la eficiencia y minimizar los costos.

Es importante tener en cuenta que estos tipos de aprendizaje no son mutuamente excluyentes y a menudo se combinan en la práctica. Se puede utilizar el aprendizaje supervisado para entrenar un modelo que clasifique datos y luego utilizar el aprendizaje por refuerzo para mejorar el rendimiento del modelo en tiempo real. Además, algunos problemas pueden ser abordados utilizando más de un tipo de aprendizaje. Por ejemplo, se puede utilizar el aprendizaje no supervisado para reprocesar los datos de entrada antes de aplicar el aprendizaje supervisado para la clasificación. Dicho diseño se observa en la Figura 3

Figura 3

Aprendizaje por Reforzamiento



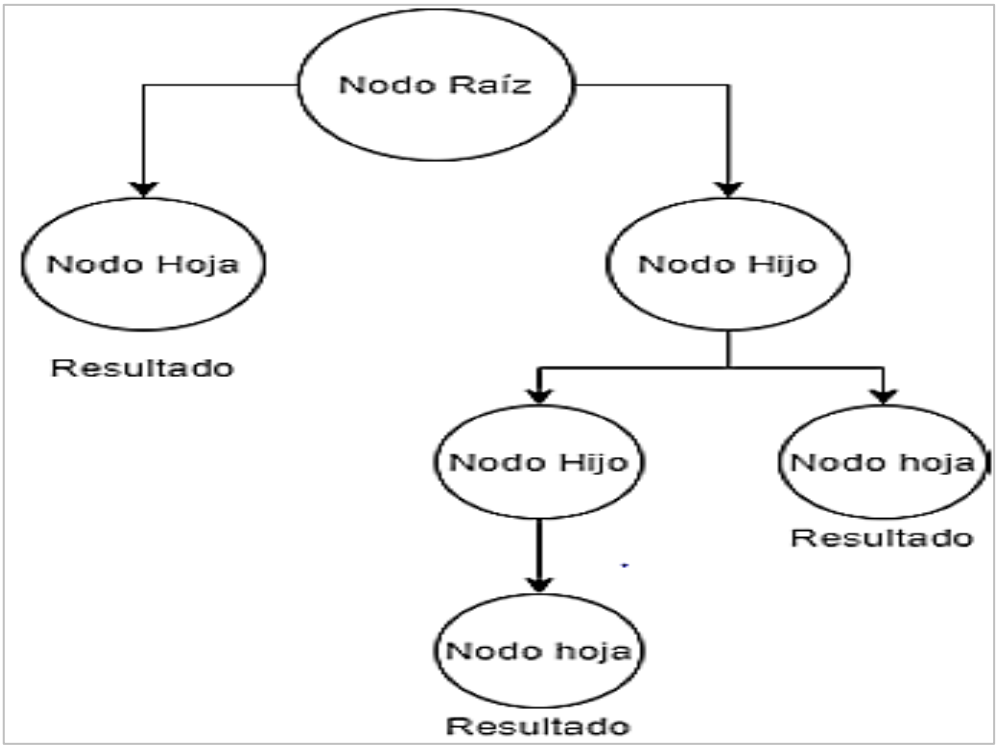
Nota: Obtenido de (TechVidvan, 2020)

2.2.2. Árboles de decisión

Es un modelo predictivo utilizado para mapear un problema para encontrar una solución final. Este diseño desarrolla inicialmente conceptualmente la problemática, puesto en un diagrama de los acontecimientos conjunto con todos los efectos interno-externo vinculados al problema.

Figura 4

Árbol de decisión



Nota: Obtenido de (Quenta Banegas, 2021)

La Figura 4 muestra que este árbol empieza en un nodo de raíz sin rama existen entrante; a continuación, las ramas se derivan en el nodo raíz (internos) o de decisión. Estos nodos analizan las características disponibles en formación de subconjuntos idénticos, representados por nodos hoja. Esto incluye salidas posibles en grupo de data. Resultante considera decisiones en la navegación a fin de establecer reglas de decisiones específicas.

El proceso de aprendizaje utiliza una estrategia de "divide y conquista" para identificar con ansia los mejores puntos de división del árbol. Esta estrategia de segmentación se aplica de forma iterativa de arriba abajo hasta que la mayoría o todas las entradas se clasifican dentro de una etiqueta de categoría específica. La homogeneidad de un conjunto de datos de clasificación está relacionada con la complejidad del árbol de decisión. Los árboles pequeños pueden mantener nodos de hojas puras, es decir, puntos de datos pertenecientes a una sola clase. Sin embargo, a medida que el árbol crece, esta pureza se hace más difícil de mantener, ya menudo resulta en demasiados pocos datos en un subárbol determinado, un fenómeno llamado fragmentación de datos, que puede provocar un sobreajuste. Por tanto, los árboles de decisión suelen ser más pequeños, de acuerdo con el principio de parsimonia extrema de Ockham que la explicación más sencilla suele ser mejor. En otras palabras, incorporan complejidad solo cuando es necesario, ya que, en la mayoría de los casos, la explicación más simple suele ser la más precisa (Mamani, 2022).

2.2.2.1. Tipos de árboles de decisión

Han surgido muchos algoritmos de árboles de decisión, muchos de los cuales se derivan del algoritmo de Hunt, que se desarrolló en la década de 1960 para imitar el proceso de aprendizaje de la psicología humana. Algunos de los principales algoritmos de árbol de decisión basados en esto son:

- ID3: La creación de ID3, que significa "Iterative Dichotomiser 3", se atribuye a Ross Quinlan. Este algoritmo utiliza medidas como la entropía y la ganancia de información para evaluar las divisiones

potenciales en el árbol de decisión. Se pueden encontrar investigaciones adicionales de Quinlan sobre este algoritmo en un documento de 1986 disponible en el siguiente enlace.

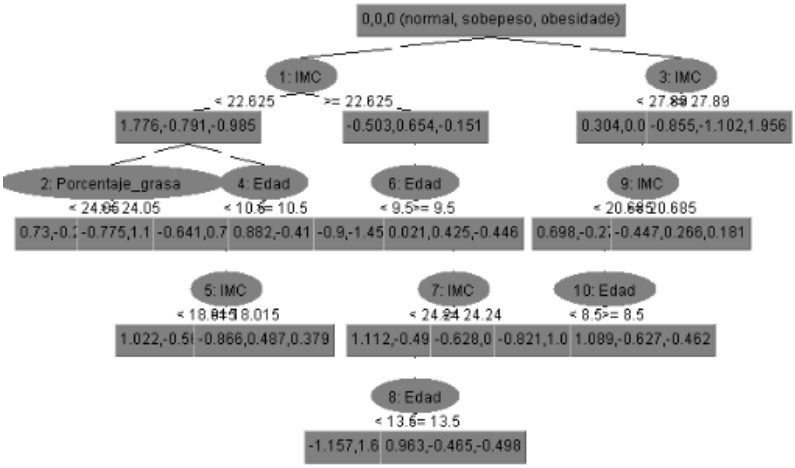
- C4.5: Este algoritmo es un avance del ID3 y también fue creado por Ross Quinlan. C4.5 Capacidad de utilizar la proporción y obtener información para analizar los puntos de división en árboles de decisión.
- CART: significa árbol de clasificación y regresión, inventado por Leo Breiman. Este algoritmo utiliza normalmente una medida de Gini de impureza para determinar las propiedades óptimas del árbol dividido. La impureza de Gini estima la probabilidad de que un atributo seleccionado aleatoriamente se clasifique mal. En el caso de impureza de Gini, los valores más bajos son más adecuados.

2.2.2.2. Árbol de decisión J48 (C4.5)

Es un algoritmo que se forma a partir de datos que se clasifican, automatizan y tienen una secuencia que van a generar reglas del árbol, con la finalidad de realizar una adecuada gestión para cada función, de manera que toda la información sea hallada en forma fácil y rápida (Kastrati et al., 2019). En la Figura 5 podemos observar el árbol de decisión J48

Figura 5

Ejemplo de árbol de decisión J48



Nota: Obtenido de (Suca et al., 2016)

2.2.2.3. Ventajas de árbol de decisión J48

Algunas de las ventajas más destacadas de los árboles de decisión J48 (C4.5) incluyen:

1. Interpretabilidad: Los árboles de decisión son modelos altamente interpretables. Son fáciles de visualizar y entender, lo que los hace adecuados para tomar decisiones basadas en reglas claras.
2. Manejo de datos mixtos: J48 puede manejar conjuntos de datos con atributos categóricos y numéricos de manera eficiente, lo que lo hace versátil para una variedad de aplicaciones.
3. Selección automática de atributos: El algoritmo J48 puede seleccionar automáticamente los atributos más relevantes para la construcción del árbol, lo que simplifica el proceso de modelado y puede mejorar la precisión del modelo.
4. Manejo de datos faltantes: Los árboles de decisión pueden manejar valores faltantes en los datos sin necesidad de imputación previa.

5. Escalabilidad: J48 es capaz de manejar grandes conjuntos de datos y puede ser utilizado en aplicaciones de minería de datos a gran escala.
6. Eficacia en la clasificación: Los árboles de decisión generados por J48 suelen ser efectivos en la clasificación de nuevos datos, y su rendimiento puede ser bastante competitivo en comparación con otros algoritmos de aprendizaje automático.
7. Tolerancia al ruido: J48 es relativamente robusto ante datos ruidosos y atípicos, ya que tiende a generar árboles que se adaptan a las tendencias generales de los datos.
8. Visualización intuitiva: Los árboles de decisión generados por J48 pueden ser representados gráficamente de manera clara, lo que facilita la comunicación de las reglas de decisión a partes interesadas no técnicas.
9. Rápido tiempo de entrenamiento: En comparación con algunos otros algoritmos más complejos, J48 generalmente tiene tiempos de entrenamiento más cortos, lo que lo hace eficiente para la construcción de modelos.
10. Poca necesidad de ajuste de hiper parámetros: A menudo, J48 no requiere una configuración de hiper parámetros extensa para obtener resultados razonables, lo que facilita su uso.

En consecuencia, el algoritmo J48 (C4.5) es una opción para la construcción de árboles de decisión debido a su interpretabilidad, capacidad para manejar una variedad de tipos de datos, y su eficacia en la clasificación de datos. Estas ventajas hacen que sea una herramienta

valiosa en el campo del aprendizaje automático y la minería de datos (Mamani, 2022).

2.2.3. Bases de datos

Es la data almacenada por el servidor siendo su característica que son clasificados de acuerdo tema, prioridad, tamaño, etc.

2.2.4. Lenguaje de bases de datos

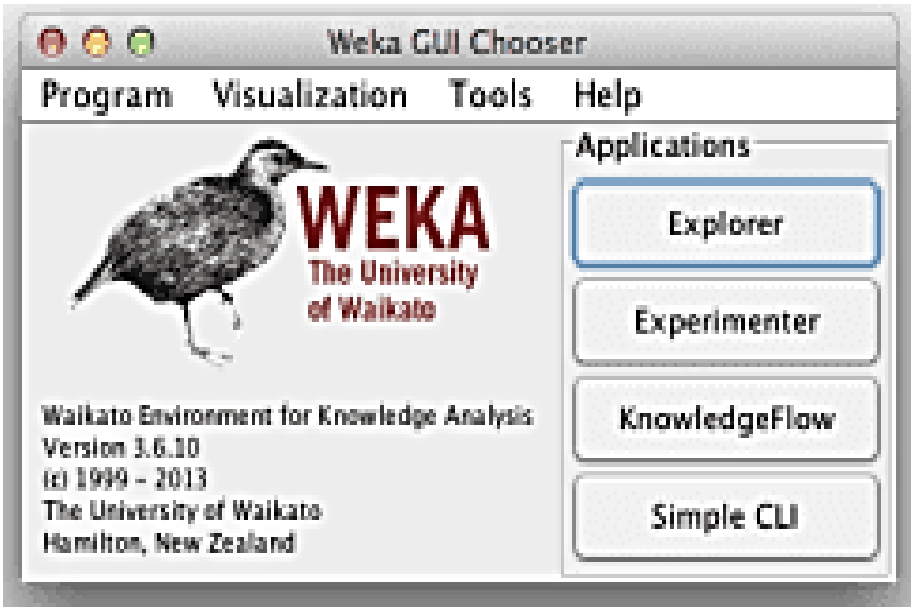
Proporciona un lenguaje de definición de datos y de operación de datos. Definiendo los datos le permite describir la estructura o el modelo de una base de datos; permitiendo formular consultas y realizar cambios en los datos de la base (Akujuobiny, 2017).

2.2.4.1. WEKA

Abreviatura de Waikato Environment for Knowledge Analysis, plataforma de software especializada en machine learning y minería de datos. Ofrece una amplia gama de para visualización y algoritmos diseñados para el data análisis y la creación de modelos predictivos. En Weka, encontramos funciones destinadas a la preparación de datos, así como algoritmos para tareas como clasificación, regresión, agrupación, identificación de reglas de asociación y diversas funcionalidades de visualización de datos (Casanova, 2022).

Figura 6

Selector de GUI de Weka



Nota: Obtenido de (Hall et al., 2009)

Como indica en la Figura 6, el selector de interfaz gráfica de usuario (GUI) de Weka ofrece opciones como Explorer, Experimenter, KnowledgeExplorer y Simple CLI (Interfaz de Línea de Comandos Sencilla). A través de esta GUI, es posible cargar conjuntos de datos y aplicar algoritmos de clasificación. Aunque también brinda funcionalidades adicionales como el filtrado de datos, la agrupación en clústeres, la extracción de reglas de asociación y diversas opciones de visualización, en este momento no utilizaremos estas características (Spositto et al., 2022).

2.2.4.2. Weka y Machine Learning

La minería de datos, o Data Mining, involucra la realización de cuatro tipos principales de tareas, que son:

- Clasificar.

- Agrupar (clustering).
- Regresión.
- Normas de asociación.

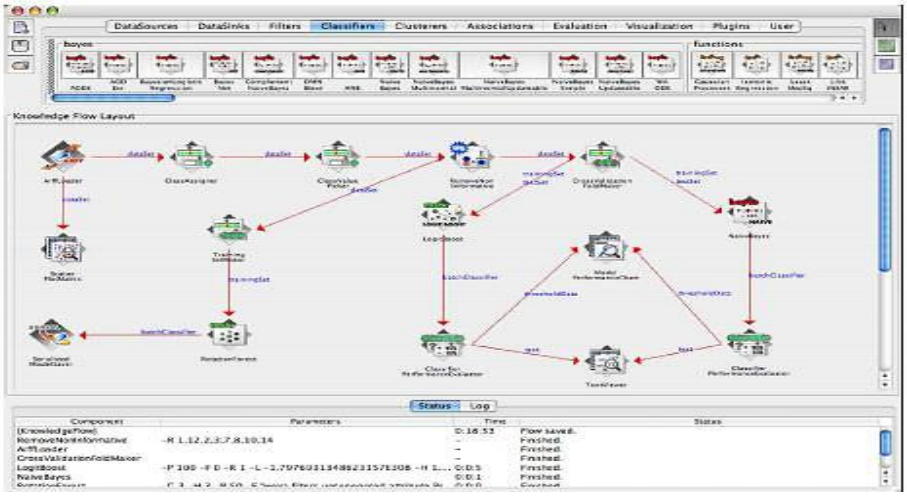
Para realizarla se utilizan técnicas estadísticas, algoritmos matemáticos y de aprendizaje automático para mejorar el rendimiento por experiencia. Como se comentado en es una subdisciplina de la ciencia de datos que implica la adquisición de conocimiento mediante la exploración, el análisis, la visualización y la visualización de datos.

En el campo del aprendizaje automático, existen tres tipos diferentes de aprendizaje:

- Aprendizaje supervisado: El algoritmo predice los valores de salida.
- Aprendizaje no supervisado: Este método suele utilizarse con fines exploratorios.
- Aprendizaje de refuerzo: El algoritmo aprende al observar su entorno.

Figura 7

Weka user Interface



Nota: Obtenido de (Hall et al., 2009)

De igual manera en la Figura 7, se observa una plataforma software desarrollada en Java que maneja varias funciones de aprendizaje automático. Estas técnicas parten de la premisa de que los datos se encuentran disponibles en un archivo plano o una relación, en la cual cada registro de datos está caracterizado por un número predeterminado de atributos. Por lo tanto, proporciona la capacidad de acceder a bases de datos mediante SQL utilizando la conexión JDBC (Java Database Connectivity) y procesar los resultados de las consultas realizadas a la base de datos.

Con respecto al algoritmo J48, que se encuentra integrado en Weka, como un algoritmo de minería de datos más ampliamente utilizados en estudios que involucran tareas por clasificación. Uno de los parámetros clave en este proceso es el nivel de confianza establecido para la poda del árbol generado, conocido como "confidence level", ya que tiene un impacto significativo en el tamaño y la capacidad predictiva del árbol resultante.

Para comprender el funcionamiento de este clasificador, por ende, se explica: que la toma de decisión del punto de corte en la iteración 'n', busca la variable predictora y punto de corte exacto donde el error cometido sea mínimo (según un criterio predefinido). Esta operación se realiza cuando estamos por encima del nivel de confianza predefinido. Después de realizar el corte, el algoritmo se repite hasta que ningún predictor alcanza un nivel de confianza superior a un determinado valor. Cabe destacar la importancia de utilizar los niveles de confianza, puesto

que el árbol resultante puede llegar a ser muy amplio en casos que contienen un gran número de temas y variables. Otra forma de controlar el tamaño del árbol es especificar un número mínimo de instancias para cada instancia por nodo (Cortés, 2019).

2.2.5. Gestión documentaria

Define un conjunto de procesos y técnicas que se integran en la administración pública. Éstos se basan en el análisis de la creación, tratamiento y valor de los documentos y tienen como objetivo planificar, controlar, utilizar, conservar, suprimir o transferir los documentos al archivo. El objetivo principal es optimizar y estandarizar su tratamiento para una gestión eficaz y rentable (Ying, 2019).

2.2.5.1. La importancia del Sistema de Gestión Documentaria

La principal tarea de las empresas en el ámbito de la gestión documental es implementar todo el ciclo de vida del documento, con el objetivo específico de convertir los documentos de formato físico a formato digital.

2.2.5.2. Ventajas del Sistema de Gestión Documentaria

En el ámbito empresarial, la organización y administración de información y documentos representan un desafío que no siempre recibe la atención adecuada. El archivo, ya sea en formato digital o físico, constituye la memoria activa de la empresa, siendo esencial para su correcto funcionamiento. Por esta razón, priorizar la gestión de la información en lugar de simplemente 'manejar' documentos se convierte

en una tarea crucial. Para alcanzar este objetivo, se centra en la creación, organización o ajuste de procesos con el propósito de optimizar los recursos y mejorar la eficacia en la labor de gestión documental.

2.2.5.3. Proceso de Trámite Documentario

El proceso de Trámite Documentario registra todos los documentos que ingresan o se generan en una organización, creando para estos y otros que se vayan añadiendo durante su trámite, una carpeta virtual o física, por medio de la cual es fácilmente identificable el usuario, el puesto de trabajo y el momento en que dicha carpeta fue procesada (Ripley, 2014).

2.2.6. ISO en la Gestión de Documentos

Según la norma ISO 30300:2011, dice: "La serie ISO 30300 proporciona un enfoque estructurado para crear y gestionar registros, coherente con los objetivos y estrategias de la organización", por lo que principalmente se usa el sistema de gestión documental (SGD) tomando en cuenta en la ejecución del estudio (Ripley, 2014).

2.2.7. Diagrama BPMN

Es una notación estándar utilizada en el campo de la gestión de procesos de negocio (BPM, por sus siglas en inglés) para representar gráficamente los procesos empresariales. Fue desarrollado por el Business Process Management Initiative (BPMI) y posteriormente adoptado por la Object Management Group (OMG) (Dos santos, 2022). En cuanto al propósito principal de BPMN es proporcionar una forma común y comprensible de representar visualmente los procesos de negocio, lo que facilita la comunicación entre las partes interesadas,

incluyendo a los usuarios comerciales, analistas de procesos y desarrolladores de software.

2.2.7.1. Características del Diagrama BPMN

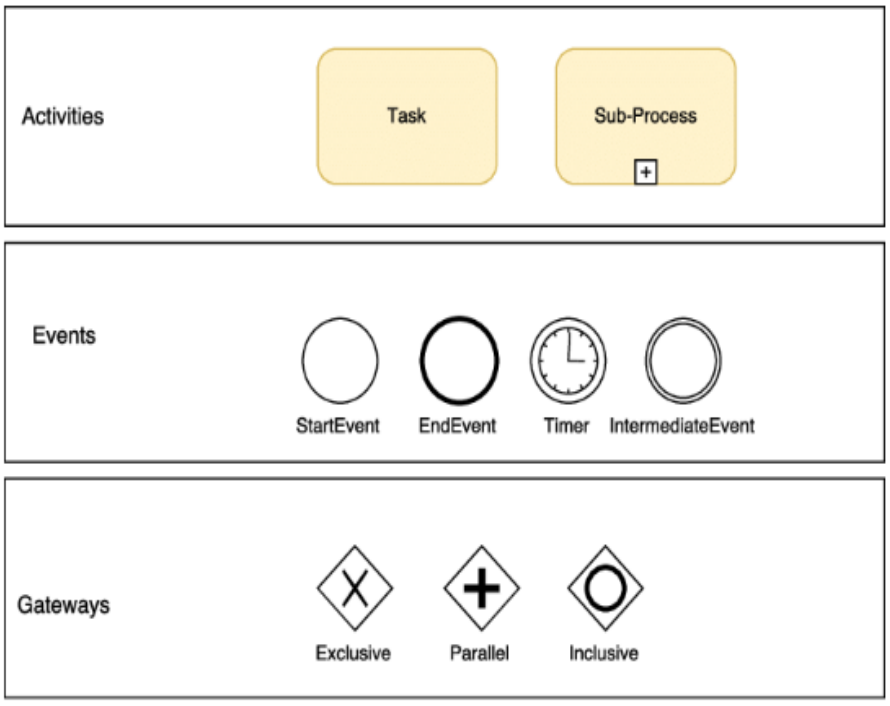
Algunas de las características clave de BPMN incluyen:

- Iconografía Normalizada: BPMN emplea una serie de esquemas gráficos normalizados para ilustrar diversos elementos dentro de un proceso, simplificando así su interpretación y comunicación. Entre los símbolos comúnmente utilizados se encuentran tareas, eventos, puertas de enlace, flujos de secuencia, entre otros.
- Escalabilidad en la Representación: BPMN es flexible y posibilita la representación de procesos en distintos niveles de detalle, abarcando desde una perspectiva general hasta un nivel de detalle específico. Esta característica facilita la comunicación tanto a un nivel ejecutivo como técnico.
- Representación Intuitiva: Los símbolos y las convenciones utilizados en BPMN están diseñados de manera intuitiva para que las personas puedan entender rápidamente cómo funciona un proceso, incluso si no tienen experiencia en notación BPMN.
- Entidades y Grupos de Trabajo: BPMN posibilita la representación de diversos participantes o actores en un proceso mediante el uso de piscinas (pools) y carriles (lanes). Esta representación visual aclara la responsabilidad de cada entidad en las distintas actividades o etapas del proceso.

- Sucesos: Los eventos desempeñan un papel esencial en BPMN, siendo elementos que indican eventos que provocan acciones dentro de un proceso. Estos pueden clasificarse como eventos de inicio, eventos intermedios o eventos de conclusión.
- Puerta: La puerta se utiliza para controlar el flujo del proceso y determinar el camino que debe seguir el proceso en función de determinadas condiciones. Éstas pueden ser puertas exclusivas, exhaustivas, paralelas, etc.
- Flujo de secuencia: el flujo de secuencia hace referencia a la dirección y el orden en el que se llevan a cabo las actividades de un proceso. Las flechas se utilizan para conectar los elementos del proceso.
- Subprocesos: BPMN permite la representación de subprocesos, que son procesos más pequeños dentro de un proceso principal. Esto ayuda a descomponer procesos complejos en partes más manejables.
- Adhesión a Estándares: BPMN es un estándar reconocido internacionalmente, lo que significa que es ampliamente aceptado y compatible con diversas herramientas y sistemas de software de BPM.
- Extensibilidad: BPMN es lo suficientemente flexible como para ser extendido para abordar requisitos específicos de una organización o industria particular.
- Soporte de herramientas de software: muchas herramientas de software de BPM admiten BPMN, lo que facilita la creación, modificación y automatización de los procesos empresariales.

Figura 8

Comprensión de los símbolos de los objetos de flujo



Nota: Obtenido de (Mroczek y Antoni, 2017)

En la Figura 8, es una notación versátil y efectiva que se utiliza ampliamente en la modelización y la gestión de procesos de negocio, gracias a sus características estandarizadas, intuitivas y adaptables (Tarr y Dur, 2022).

Los componentes fundamentales del flujo de trabajo se conocen como objetos de flujo y conforman la estructura general del proceso. Estos elementos incluyen eventos, actividades y las puertas de entrada, a continuación, su descripción:

- Los eventos, representados por símbolos circulares, actúan como disparadores en diferentes etapas del proceso. Por ejemplo, un evento de mensaje indica el envío o recepción de un correo electrónico o mensaje de texto, mientras que un evento de error

señala problemas que interrumpen el flujo del trabajo. Los eventos también pueden estar vinculados a temporizadores para marcar el inicio de acciones basadas en el tiempo o a escalones superiores en la organización para revisiones manuales.

- Las actividades, mostradas como rectángulos redondeados, describen tareas específicas realizadas por individuos o sistemas. Estas pueden ser acciones únicas, acciones repetidas o tareas sujetas a condiciones específicas. Las tareas son precisas y no se pueden dividir en sub acciones. Además, existen las transacciones, que implican procesos de pago, y los subprocesos, que comprenden un conjunto de tareas relacionadas.
- Las puertas de entrada, identificadas por símbolos de diamante, son puntos de decisión en el diagrama BPMN. Estos puntos marcan bifurcaciones en el camino del proceso. Por ejemplo, una puerta de enlace exclusiva espera la entrada correcta de un código secreto antes de decidir si permitir el acceso o negarlo. Otro tipo, basado en eventos, toma decisiones especializadas según eventos específicos, como la generación de una lista de usuarios en un día determinado. También están las puertas de enlace paralelas, que permiten acciones simultáneas sin esperar condiciones particulares. Estos elementos son esenciales para el modelado preciso y detallado de los procesos de negocio (Tineo et al., 2022).

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. TIPO DE INVESTIGACIÓN

Según Hernández y Mendoza (2018) el tipo de investigación adoptado fue aplicada, ya que se centró en resolver problemas específicos y concretos relacionados con la clasificación y optimización de la gestión documental en una entidad educativa, utilizando herramientas tecnológicas como WEKA y el algoritmo J48. Se buscó una solución práctica y eficiente a los problemas de clasificación de documentos, mejorando los procesos administrativos y la eficiencia en el manejo de la información.

Este estudio también se puede clasificar como un enfoque cuantitativo, debido a que se emplearon datos cuantificables (documentos clasificados, características semánticas de los mismos, resultados del modelo predictivo) y además se empleó un alcance de investigación explicativa.

3.2. DISEÑO DE INVESTIGACIÓN

Según Carrasco (2019) El diseño cuasi experimental es adecuado cuando, por limitaciones prácticas o éticas, no es posible realizar una asignación aleatoria de las unidades de estudio (en este caso, los documentos) a diferentes grupos. En este estudio, la intervención consistió en la implementación del modelo de clasificación basado en el árbol de decisión J48 para analizar y optimizar la gestión documental, lo que permitió evaluar los efectos de esta intervención a través de la comparación de resultados antes y después de su aplicación.

Diseño con pre test y post test para poder realizar las comparaciones entre los valores obtenidos y medir el efecto que provoca, el esquema es el siguiente:

$$O1 - X - O2$$

Donde:

$O1$= Medición antes (Pre test).

X = Implementación del árbol de decisión J48.

$O2$= Medición después (Post test).

3.3. POBLACIÓN

Según Carrasco (2019) la población se refiere a todo aquel aspecto de análisis, a la unidad que se usará como análisis conformado por un grupo de participantes, los cuales tienen ciertas características en común. La población estuvo compuesta por aquellos documentos ingresados al área de gestión administrativa de la entidad, en un periodo de 3 meses. La población en este estudio, la conformaron 1000 documentos.

3.4. MUESTRA

Acorde con Carrasco (2019) es un pequeño fragmento de la población, los cuales ayudarán a la investigación. La muestra será no probabilística por conveniencia la cual estuvo compuesta por aquellos documentos ingresados al área de gestión administrativa de la entidad. La muestra en este estudio, la constituyeron 350 documentos.

3.5. INSTRUMENTOS

- **Data Set:** Para poder trabajar la información, se elaboró un *data set* con los documentos entrantes de la oficina de gestión administrativa de la entidad, por el lapso de 3 meses, para tener almacenados las terminologías a utilizar en el árbol de decisión J48. Esta información fue clasificada y verificada por el árbol de decisión. (Carrasco, 2019).

- **Listas:** Se utilizó para medir los resultados a un conjunto de listas que estuvo compuesto por los documentos y las oficinas estos serán debidamente ordenados mediante el árbol de decisión J48 (Carrasco, 2019).
- **Validez/confiabilidad:** Se aplicó para validar el instrumento a fin de que se utilice métodos estadísticos como: Chi-cuadrado, CATPCA y coeficiente de correlación. Los mismos se realizaron en el software SPSS V.26 a fin de que se depure la data y se presente exhaustiva la información con el que se garantiza la fiabilidad del mismo.
- **Procedimiento del experimento**: Hernández y Mendoza (2018) señalan que para la investigación se requiere realizar los siguientes pasos:
- Se definieron los instrumentos que se utilizaron para obtener la información.
  - Se realizaron todos los procesos requeridos para la entrada de datos.
  - Se configuró el árbol de decisión J48.
  - Se realizó el clasificado, es por ello que se utilizó el árbol de decisión J48.
  - Se realizó el análisis de la información obtenida.
- **Plan de procesamiento y análisis de datos:** Para la obtención requerida de la información para esta investigación, se procedió a realizar el análisis de los documentos y sintetizar las partes más importantes. Para esto, se emplearon las tablas de información, por ello estos resultados fueron presentados en tablas para hacer el respectivo análisis e interpretación. En la estadística, se empleó el método descriptivo, que se manifiesta a través de porcentajes. Además, se emplearon técnicas propias del árbol de decisión para interpretar la información.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. DESCRIBIR EL PROCESO DE CLASIFICACIÓN DE DOCUMENTOS ACTUAL, MEDIANTE EL DIAGRAMA BPMN, EN LA UGEL CHUCUITO, JULI 2023.

Para Akujuobi y Zhang (2017) es un conjunto de datos que tomará para su tratamiento un proceso sistematizado, jerarquizado y ordenado que seguirá los siguientes pasos:

4.1.1. Recopilación de datos

Después de verificar los documentos del área de gestión administrativa durante los 3 meses, se consigue simplificar la información en un pequeño data set. Los documentos revisados fueron 1000 documentos, representando una muestra 350 documentos bajo la forma probabilístico aleatorio simple por conveniencia.

Las respuestas posibles que se obtuvieron del análisis organizacional del área en el estudio presentan lo siguiente:

- Código de identificación del documento
- Asunto principal del documento
- Datos adicionales
- Nombre del usuario
- Fecha de presentación del documento

Cuando se obtenga la información relevante, se realizará una data set provisional. Detallada en la Tabla 1.

Tabla 1

Datos conseguidos de la entidad

	Código del documento	Asunto del documento	Datos adicionales	Nombre del usuario	Fecha en que se presenta documento	Oficina que recepciona el documento
Descripción de la información obtenida	Una vez el documento ingrese a la UGEL Chucuito este será asignada un código para un seguimiento del caso	Es el tema que contiene el documento	Son ideas o razones complementarias que explican el asunto del documento.	Es para reconocer que ciudadano presenta el documento	Es el identificado por el cual se reconoce en qué fecha se presentó el documento	Es la oficina que recibe el asunto y para ello origina una respuesta al destinatario.

4.1.2. Normalización de la información obtenida

La información obtenida de los archivos irregulares no contiene ningún patrón o secuencia que un árbol de decisión pueda reconocer o utilizar. Por este motivo, es necesario normalizar estos documentos. Con el objetivo de estandarizar los procedimientos de tratamiento de datos, se utilizan dos técnicas.

4.1.3. Reducción de la información (Eliminar información que es importante)

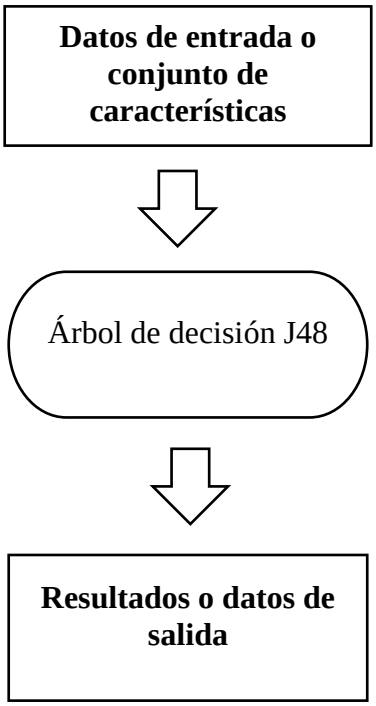
De la información obtenida en la data se limita, porque esta información no brinda soporte sobre la intención o asunto del documento.

4.1.4. Separación de la información

El árbol de decisión J48 recibe parámetros de data en ingreso-egreso. Extraídos por la data set. Detallados en la Figura 9

Figura 9

Diagrama simplificado de funcionamiento del árbol de decisión J48



4.1.4.1. Información de Entrada

En un árbol de decisiones, las propiedades del archivo se agrupan para hacer posible su clasificación y evaluación posterior. A tal efecto, se tendrá en cuenta de la data útiles para el proceso de clasificación.

- Data para conformar la información de ingreso
- Asunto del documento e información suplementaria: Con base en esta información, son clasificados y enviados al área que corresponda.

¿Cómo conseguir información a fin de convertirlo en información de ingreso en J48-árbol para decisiones?

Se procesa la información de ingreso a fin de que el proceso de normalización comprende los siguientes pasos:

- Eliminar o reemplazar la data inconsistente
- Identificar palabras claves
- Reducir palabras claves de especificaciones

Para ilustrar el proceso de normalización de la data extraída de la documentación y convertido en datos ingreso en el árbol, se creó un diseño simple que se basó en la documentación conjunto con la data obtenida en la figura 7, dentro del Diagrama BPMN del proceso de clasificación de documentos actual. Dado que previamente se eliminaron del documento, en este se extrajo información requerida.

4.1.4.2. Diagrama BPMN

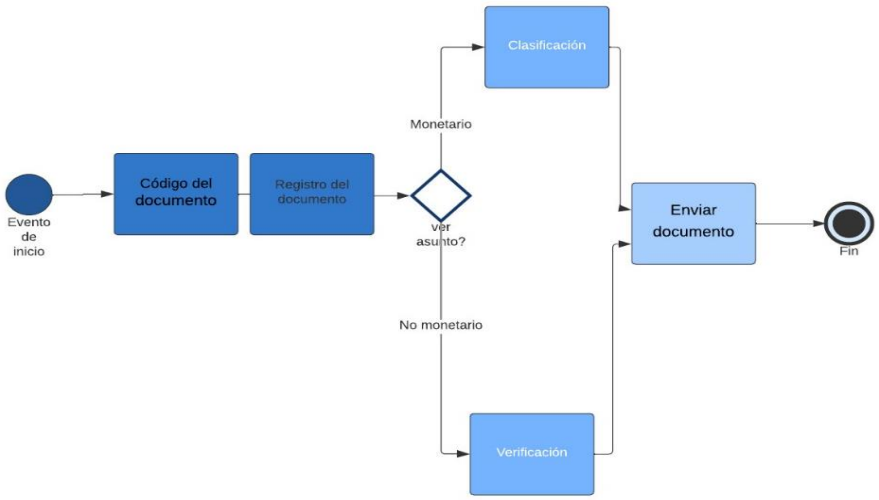
- Asunto del documento: solicito material educativo que falta con carácter urgencia para el inicio de clases en la UGEL Chucuito
- Información adicional: El personal responsable no elaboró un informe de cumplimiento cuando se recibió el material.

La información mostrada como frases a corresponder a la data importante de las actas, la mismas fueron “Asunto del Documento” y “Data Adicional”. Estos datos fueron procesados de forma manual, procesando los datos de cada documento que ingresaba al campo de control administrativo de la UGEL Chucuito.

A continuación, se presenta en la Figura 10 el diagrama del proceso de clasificación de documentos actual:

Figura 10

Diagrama BPMN del proceso de clasificación de documentos actual



4.2. PREPARACIÓN DEL DATASET, EN EL ÁRBOL DE DECISIÓN J48

Este procedimiento integral sirve como una orientación fundamental para la preparación del conjunto de datos y la aplicación del árbol de decisión J48. Su objetivo es garantizar que los datos estén enriquecidos semánticamente y preparados de manera efectiva para llevar a cabo la clasificación de palabras clave con éxito:

4.2.1. Reemplazar/Eliminar información irrelevante

Tabla 2

Reemplazar/Eliminar

Identificador	Texto original	Texto trabajado
Caso 1	Solicito materiales faltantes para el inicio de clases en la UGEL Chucuito. El personal a cargo no hizo el informe de conformidad cuando se recibieron los materiales.	Solicito materiales faltantes para el inicio de clases de la escuela. El personal a cargo no hizo el informe de conformidad cuando se recibieron los materiales.

Característica del desarrollo

El desarrollo permite eliminar y/o reemplazar a fin de simplificar los datos de entrada y formar un conjunto de datos leídos por el J48-árbol para decidir. Así como de automatizar el desarrollo, donde se creó el detalle por Python como algoritmo utilizando Anaconda IDE, que se enfoca en identificar conjuntos de letras utilizadas repetidamente que pueden reemplazarse por oración única o eliminarse donde no aflija el desarrollo en el estándar de la data. Se realizó por el algoritmo con el que completa el desarrollo a continuación:

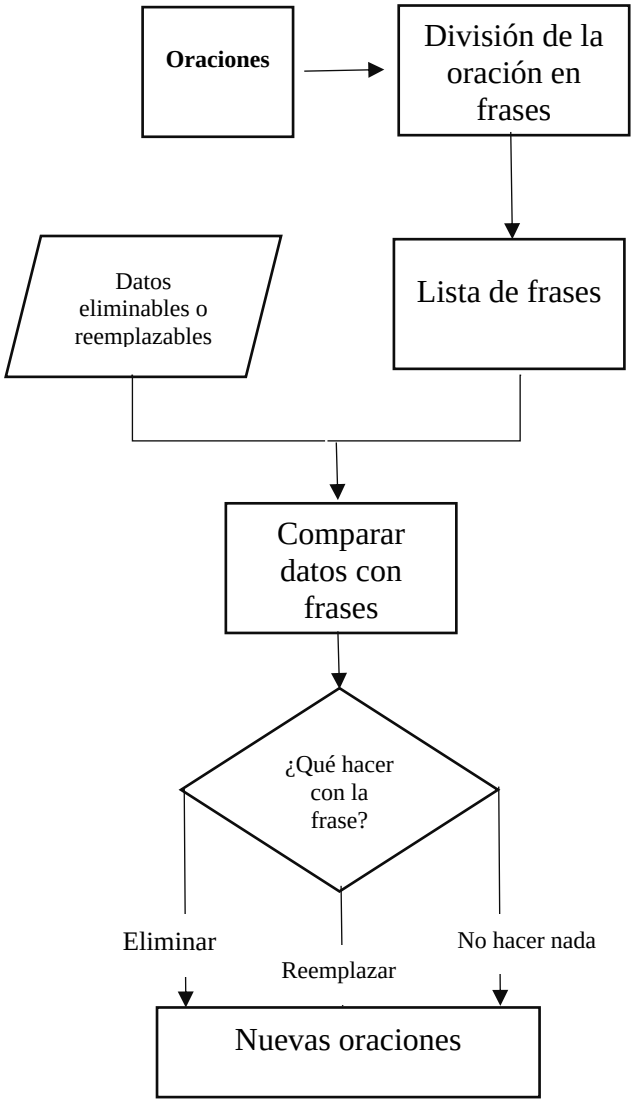
a. Dividir la oración en frases: El ingreso de este paso fue el asunto del documento y la Información adicional, de donde se obtuvo una lista de oraciones u oraciones abreviadas, que utiliza el algoritmo a continuación.

b. Compare data a declaración: Este proceso se realizó para evaluar si las declaraciones contenían datos que pudieran eliminarse y/o reemplazarse de forma segura.

c. Comparar los datos con la oración, el algoritmo dio una respuesta, las mismas permiten respuesta posible fueron: el recorte de la oración, sustitución de la oración y/o no accionar en la oración.

Figura 11

Diagrama proceso de eliminar/reemplazar la información irrelevante



Esta presenta Figura 11 se observa el diagrama en muestra por ciclo de vida en el algoritmo de reemplazos o/y eliminación innecesaria de entrada en el proceso de salida estándar.

4.2.2. Identificación de palabras clave

El proceso de enriquecimiento semántico y clasificación de palabras clave implica una serie de pasos detallados para mejorar la comprensión y relevancia de los datos, lo que contribuirá a un mejor rendimiento en tareas como la

implementación de árboles de decisión J48 u otros modelos de aprendizaje automático. A continuación, una explicación más detallada:

4.2.2.1. Análisis Semántico:

Utiliza técnicas de procesamiento de lenguaje natural (PLN) para realizar un análisis semántico de las palabras clave. Esto implica comprender el significado contextual de las palabras, considerando sinónimos, antónimos y relaciones semánticas.

4.2.2.2. Uso de Modelos de Embeddings:

Este modelo de palabras como Word2VEC, gloVE O fASTtext permite expresar letras vectoriales en un espacio semántico. Además, que captura la relación entre palabras y mejora de compresión contextual.

4.2.2.3. Clasificación de Palabras Clave:

Implementa algoritmos de clasificación para asignar categorías o etiquetas a las palabras clave. Puedes utilizar métodos supervisados, como clasificadores basados en aprendizaje profundo o algoritmos de clasificación tradicionales, para categorizar las palabras en función de su significado y contexto.

4.2.2.4. Validación y Refinamiento:

Valida la calidad de la clasificación mediante la revisión manual de algunas instancias. Refina las etiquetas asignadas según sea necesario y asegúrate de que la clasificación refleje de manera precisa el contexto semántico de las palabras clave.

4.2.2.5. Integración con Datos Existente:

Integra la información semántica y las etiquetas clasificadas con el conjunto de datos existente. Asegúrate de que estas mejoras se apliquen de manera coherente y homogénea en todo el dataset.

4.2.2.6. Visualización (Opcional):

Si es posible, realiza visualizaciones para comprender mejor la distribución de las palabras clave en el espacio semántico. Esto puede ayudar a identificar patrones y relaciones semánticas que podrían no ser evidentes de otra manera.

4.2.2.7. Iteración y Mejora Continua:

El proceso de enriquecimiento semántico y clasificación de palabras clave es iterativo. Realiza ajustes y mejoras continuas en función de la retroalimentación, resultados de evaluación y nuevas incorporaciones de datos.

Establece una forma en base al ingreso tratado anterior con letras claves a ser analizadas posteriormente.

Tabla 3

Proceso de Identificación de palabras claves

Identificador	Texto original	Lista palabras clave
Caso 1	Solicito materiales faltantes para el inicio de clases de la escuela. El personal a cargo no hizo el informe de conformidad cuando se recibieron los materiales	a. Solicitud b. Material en stock c. Material Faltante d. Colegio e. Personal f. Conformidad g. Informe h. Recepción

En la tabla se observa la primera columna contiene el ID del documento, en este ejercicio en la columna dos presenta el valor del desarrollo reemplazar/eliminar de datos irrelevantes; finalizando en la columna tres es el resultado del detalle para identificar las letras relevantes.

4.2.2.8. Procesar información

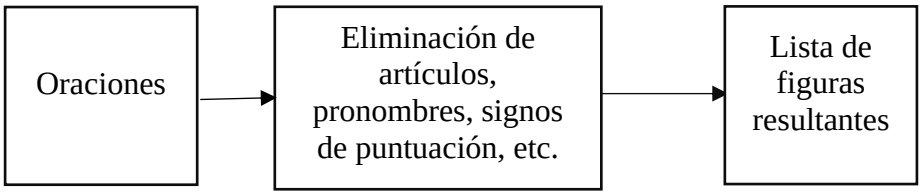
La detección de palabras clave es el resultado del proceso de eliminar información innecesaria de una oración, como artículos, pronombres, conjunciones, etc., que no está relacionada con el grupo de data ingreso del J48-árbol para decisiones. El desarrollo realiza un algoritmo minúsculo realizado en Python que ocupa IDE-Anaconda.

El trabajo u operación de este algoritmo de proceso fue sencillo. Elimine las palabras o caracteres innecesarios para obtener solo las palabras clave necesarias para formar los datos de entrada.

La funcionalidad del algoritmo en el desarrollo es fácil. La eliminación de letra o frase innecesaria, permite lograr palabras claves solamente con el fin de necesitar la forma de data por ingreso.

Figura 12

Diagrama simplificado, proceso de identificación de palabras clave



En la Figura 12 se presenta el diagrama en muestra por ciclo de vida en el algoritmo de reemplazos o/y eliminación innecesaria de entrada en el proceso de salida estándar.

4.2.3. Minimización de letras claves por detalles

El desarrollo realiza la evaluación de cada palabra clave y determinar si debe continuar como entrada o si debe eliminarse. Para hacer esto, cada palabra clave se compara con el grupo de detalles en forma de data ingreso de J48-árbol por decisiones. Uno necesita saber el detalle que componen la data de ingreso. Es así que, se ha creado una lista de los datos más importantes y se puede utilizar para crear una clasificación de árbol de decisión J48.

4.2.4. Lista de características

- Moneda de cambio
- Afinidad y su relación personal
- Desarrollo del acta
- Entidad que se involucra y su tipo
- Característica procesal

- Estado

Los detalles seleccionados expresan las cualidades en que se dividen la data contenida en las actas, a fin de concluir posterior en su análisis de entre 300 a más actas, durante el cual se recopiló una lista de atributos o características importantes para en el detalle de clasificar por el J48-árbol para decisiones. El valor se ejecuta para recibir por función se determinaron agrupando los conceptos relacionados con cada función y los expertos de la UGEL.

4.2.4.1. Conceptualización de los detalles

Moneda de cambio

Esta característica como su nombre lo indica, explora el aspecto económico ligado al documento.

Tabla 4

Característica moneda de cambio

Posibles Valores	Descripción
Saldo	Cuando la razón del documento involucra el desenvolvimiento de dinero por parte de la UGEL
Materiales	Cuando la razón del documento se relaciona con materiales de la UGEL o administrados por la misma
No monetario	Cuando no forma parte de ninguno de los dos posibles valores anteriores

Afinidad y su relación personal

La cualidad permite la correlación del cliente en el acta con el involucrada en la misma.

Tabla 5

Característica afinidad y su relación personal

Posibles Valores	Descripción
Titular	La persona a la que afecta o involucra el documento es la misma que la presentó.
Familiar	Muchas veces los familiares, ya sean los padres o hermanos, presentan un documento en nombre de su familiar para realizar algún trámite
No especificado	Cuando no forma parte de ninguno de los dos posibles valores anteriores.

Desarrollo del acta

Esta característica vislumbra cual será el trato que recibirá el documento una vez sea enviado a la oficina correcta.

Tabla 6

Característica Desarrollo del acta

Posibles Valores	Descripción
Contra respuesta	Cuando el documento presentado requiere una contestación en forma de otro documento oficial
Respuesta	Cuando el documento presentado no requiere una contestación escrita o en forma de documento
Lectura	Cuando la función del documento es únicamente informar o poner en conocimiento asunto que corresponde
No especifica	Cuando el documento no forma parte de ninguna de las tres posibles respuestas

Entidad que se involucra y su tipo

La cualidad logra establecer que entidad se involucra en el acta, debido a que en la UGEL hay varios departamentos que atienden a áreas designadas en alguna acta, por parte de organizaciones y personal.

Tabla 7

Característica entidad que se involucra y su tipo

Posibles Valores	Descripción
Colegio	Cuando la institución educativa involucrada con el documento es un Colegio
CETPRO	Cuando la institución educativa involucrada con el documento es un CETPRO
No especificado	Cuando el documento no forma parte de ninguna de las dos posibles respuestas

Característica procesal

Esta cualidad permite establecer el modelo procesal esperar a la realización de las personas en el acta; por ende, esta es especial a fin de necesitar el entendimiento del desarrollo que existe y/o existirá así de como surgen.

Tabla 8

Característica procesal

Posibles Valores	Descripción
Contratación	Proceso por el cual las instituciones educativas pertenecientes a la jurisdicción de la UGEL contratan docentes. Dicho proceso se da por medio de concurso
Adjudicación	Proceso por el cual el docente ganador de un concurso público toma posesión de la plaza ganada
Visación	Proceso por el cual un certificado o documento es validado por la UGEL
No especifica	Cuando el documento no forma parte de ninguna de las tres posibles respuestas

Estado

Esta cualidad permite establecer tiempos por evento del acta que hace referencia a otros posteriores en pasado, presente y futuro. Lo que culmina en casi todas las oficinas de la entidad UGEL, por ende, es importante que se tome el control en estos eventos de inicio a fin para que las demás oficinas lo generen igual.

Tabla 9

Característica Estado

Posibles Valores	Descripción
Normal	Cuando el documento no está sujeto a ninguna convocatoria o proceso
Próxima	Cuando el documento menciona un proceso que aún no ha comenzado.
En Curso	Cuando el documento menciona un proceso que se está realizando en la actualidad.
Finalizado	Cuando el documento menciona un proceso que ha finalizado.

4.2.4.2. Característica del desarrollo

La formación de data inicial resulta del proceso comparativo como de la suma. A fin de lograr este objetivo, primero se debe tener una lista de propiedades o cualidades y sus resultados posibles. De allí se recibe eso, sigamos adelante, utiliza otro algoritmo programado en Python usando Anaconda IDE para que contribuya en letras claves en atributos. Se eligió el valor propio más adecuado con un valor de "No específico" entre los valores posibles para determinar lo que contienen los datos de entrada.

Tabla 10

Esquema de datos de entrada para la data set

Valor tipo monetario	Valor relación afinidad con la persona	Valor tratamiento del documento	Valor tipo entidad involucrada	Valor tipo proceso	Valor estado

En la tabla 10 de algoritmo presenta los detalles a continuación:

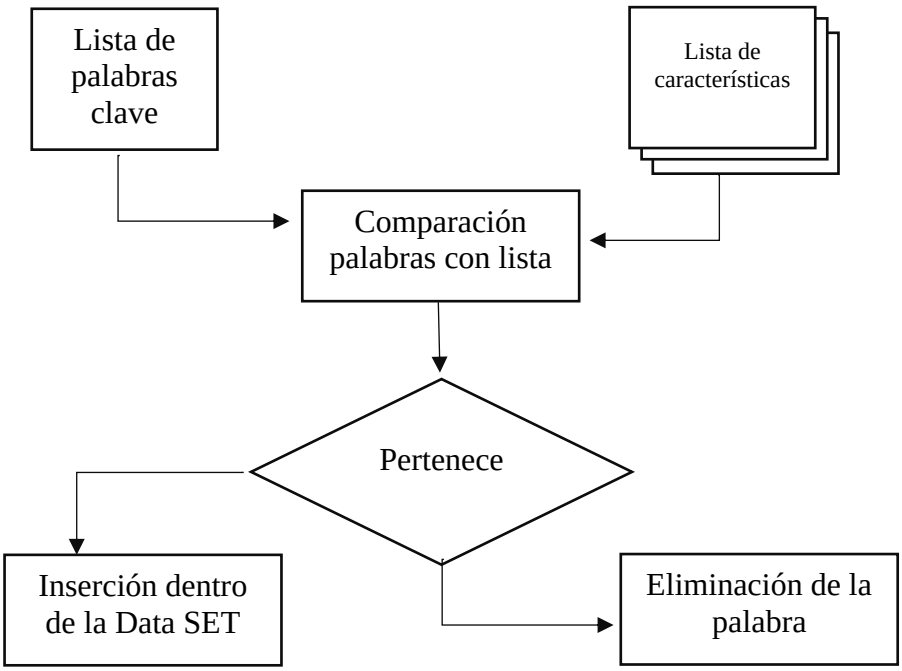
i. Comparativa de frases claves por cualidad, dado que estas no siempre se corresponden a las cualidades o valor posible, el proceso de comparación tuvo que utilizar la función de sinónimo y el uso de la etiqueta de contexto, función creada para esta investigación.

ii. Si una palabra clave no recibía un valor para todas las características, se eliminaba o se ignoraba cuando se agregaba al conjunto de datos como atributo de datos por ingreso

iii. En sí, la letra clave se asemeja a la cualidad de una de ellas, se aplica el conjunto datos como cualidad de datos por ingreso.

Figura 13

Detalle de la creación data de ingreso en data-set



La figura a continuación presenta el diagrama de cómo es el proceso de obtener datos de entrada y simultáneamente incluirlos en el conjunto de datos en función de letras claves que se obtuvieron en el desarrollo de identificar las mismas.

Tabla 11

Data obtenida de ingreso en data-set

Identificador	**Lista palabra clave**	**Información de entrada resultante**	
		Característica	**Valor**
Caso 1	Solicitud Materiales en stock	Tipo monetario Relación afinidad con la persona	Materiales Titular
	Materiales faltantes	Tratamiento del documento	Respuesta
	Colegio	Tipo de entidad	Colegio
	Personal	Tipo de persona	No especifica
	Conformidad Informe Recepción	Estado	finalizado

En la Tabla 11 se observa en la columna uno contiene como identificación de acta, en este ejercicio de caso 1, posterior la columna 2 presenta la resultante del detalle de identificar las letras claves. Para finalizar en la columna tres el resultado del detalle de reducir las letras claves por cualidad.

4.2.4.3. Datos de salida

El conjunto de datos obtenido inicialmente de las actas se utiliza para generar los data por egreso que forman parte del conjunto de datos finales. Debido a que se indica la adecuada diversificación de las actas; es así que al utilizar parte de la ejecución por el J48-árbol para decisiones. Detallada a continuación en la Tabla

Tabla 12

Resultados datos de salida de la data set

Identificador del documento	Oficina receptora (Resultado clasificación)
Caso 1	Almacén

4.2.4.4. El conjunto de datos no se analizó

Si se conocen los datos de entrada y salida de una posible evaluación individual y el identificador del documento, está listo un conjunto de datos no analizados; es decir, un conjunto de datos formado a partir de la información de los documentos. El conjunto de datos se construyó de la siguiente manera.

Tabla 13

Configuración de la data set no analizada

Identificador							**Resultado**
	Tipo monetario	Relación afinidad con las personas	Tratamiento del documento	Tipo de entidad involucrada	Tipo de proceso	Estado	
Caso 1	Materiales	Titular	Respuesta	Colegio	No especifica	Finalizado	Almacén

De la tabla 13 podemos decir que después de procesar los datos de todos los documentos, se obtuvo un conjunto de datos de 350 ítems.

4.2.4.5. CATPCA-analizar los elementos idóneos categóricos

A fin de la reducción de las dimensiones de la data se necesita reducir pertinente la factibilidad, por ende, la reducción de la dimensión se

analiza en relación de variables donde puede eliminar la información. Por lo tanto, se analizó las características en correlación.

Tabla 14

Coeficientes de correlación de las características

Variable 1	Variable 2	Coeficientes de correlación
Tipo monetario	Tratamiento del documento	-0.287
Tipo monetario	Tipo entidad involucrada	0.453
Tipo monetario	Tipo proceso	-0.037
Tipo monetario	Estado	-0.047
Tratamiento del documento	Tipo entidad involucrada	-0.085
Tratamiento del documento	Tipo proceso	0.174
Tratamiento del documento	Estado	0.015
Tipo entidad involucrada	Tipo proceso	0.045
Tipo entidad involucrada	Estado	-0.071
Tipo proceso	Estado	-0.093

Análisis

En la tabla podemos observar el análisis de correlación entre las variables proporcionadas muestra diversas relaciones, algunas de las cuales son más fuertes, mientras que otras son débiles o inexistentes. A continuación, se presenta un análisis detallado de cada par de variables, considerando los coeficientes de correlación:

Tipo monetario y Tratamiento del documento (-0.287)

Existe una correlación negativa débil entre estas dos variables. Esto sugiere que, a medida que una de las variables aumenta, la otra tiende a disminuir,

pero esta relación no es lo suficientemente fuerte como para ser considerada significativa en muchos contextos.

Tipo monetario y Tipo entidad involucrada (0.453)

Aquí se observa una correlación positiva moderada. A medida que una de las variables aumenta, la otra tiende a aumentar también. Esto indica una relación positiva entre el tipo monetario y el tipo de entidad involucrada, pero la fuerza de esta relación aún no es fuerte.

Tipo monetario y Tipo proceso (-0.037)

La correlación es prácticamente nula, lo que significa que no existe una relación significativa entre estas dos variables. La variabilidad de una variable no tiene un impacto claro sobre la otra.

Tipo monetario y Estado (-0.047)

Similar al caso anterior, la correlación es muy débil, indicando que no hay una relación significativa entre el tipo monetario y el estado.

Tratamiento del documento y Tipo entidad involucrada (-0.085)

Existe una correlación negativa débil, lo que sugiere que no hay una relación fuerte entre el tratamiento del documento y el tipo de entidad involucrada. Las variables parecen estar poco relacionadas.

Tratamiento del documento y Tipo proceso (0.174)

La correlación es débilmente positiva. Esto sugiere que hay una relación ligera entre el tratamiento del documento y el tipo de proceso, pero no es suficientemente fuerte como para poder hacer predicciones claras entre ellas.

Tratamiento del documento y Estado (0.015)

La correlación es prácticamente nula, indicando que no hay una relación significativa entre el tratamiento del documento y el estado.

Tipo entidad involucrada y Tipo proceso (0.045)

La correlación es muy débil y positiva, lo que indica que no existe una relación fuerte ni clara entre estas dos variables.

Tipo entidad involucrada y Estado (-0.071)

Esta correlación también es débil y negativa, lo que sugiere una relación muy tenue entre el tipo de entidad involucrada y el estado, sin una conexión significativa.

Tipo proceso y Estado (-0.093)

La correlación es también débil y negativa, lo que indica que no hay una relación fuerte entre estas dos variables.

En general, los coeficientes de correlación muestran que la mayoría de las relaciones entre las variables son débiles o prácticamente inexistentes. Las correlaciones más notables se encuentran entre Tipo monetario y Tipo entidad involucrada (0.453) y Tipo monetario y Tratamiento del documento (-0.287). Sin embargo, en su conjunto, las relaciones entre las variables no parecen ser lo suficientemente fuertes como para realizar predicciones claras o conclusiones significativas sin un análisis más profundo y detallado.

4.2.4.6. Final data-set

Finaliza con el análisis de Chi-cuadrada y CATPCA, donde se logró la modificación final data-set, a ser usada por el J48-árbol para decisiones. El final data-set tiene 350 componentes.

Tabla 15

Final data-set

Identificador	**Características**					**Resultado**
	Tipo monetario	Tratamiento del documento	Tipo de entidad involucrada	Tipo de proceso	Estado	
Caso 1	Materiales	Respuesta	Colegio	No especifica	Finalizado	Almacén

En la tabla 14 podemos identificar en la primera columna presenta el identificador del documento, para proteger la confidencialidad de los documentos, se les asignó un identificador ajeno al que usa la UGEL para darles seguimiento. De la segunda hasta la séptima columna representan los datos de entrada que recibirá el árbol de decisión J48 para realizar el proceso de clasificación. Para finalizar, en la última columna se encuentran los datos de salida que recibirá el árbol de decisión J48 para realizar el proceso de entrenamiento y posterior verificación de clasificación. Esta data set puede ser leído perfectamente por el árbol desarrollado, ya que fue creado y personalizado expresamente para esta investigación.

4.2.4.7. Eventos encontrados

- Recopilar información de documentos manualmente requiere mucha mano de obra, en este caso fue particularmente difícil porque los

documentos no llegaron al área administrativa en el tiempo especificado.

- El proceso de obtención de datos de los documentos se puede agilizar haciendo uso de otras herramientas de I. A., tales como reconocedores de texto y/o voz para reducir el tiempo necesario para realizar esta tarea.
- Antes de crear la data set final y las características que estarían presentes en los datos de entrada, se debe hacer un proceso de análisis de los datos obtenidos de los documentos. Dicho análisis debe ser realizado con la ayuda de expertos de la UGEL, a fin de crear variables o características representativas de los documentos.

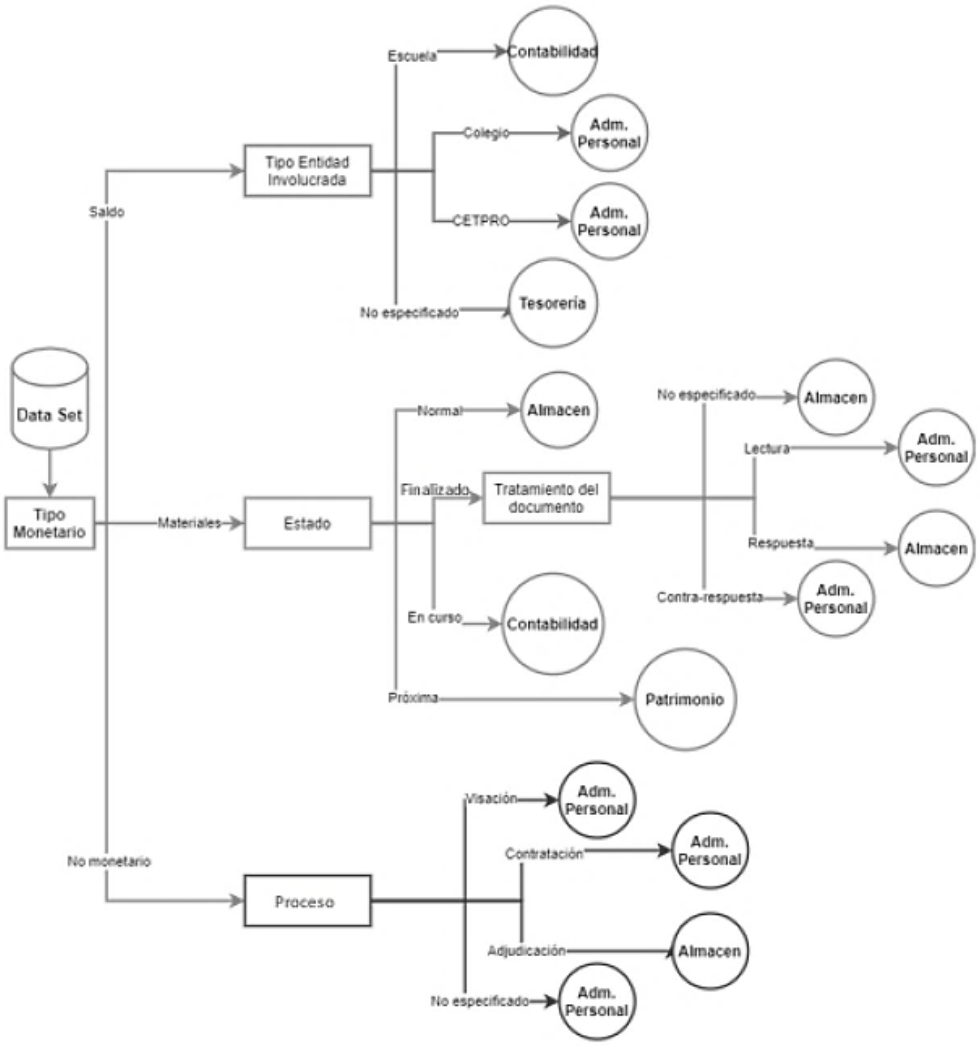
4.3. CONSTRUIR J48-ÁRBOL PARA DECISIONES CON WEKA COMO HERRAMIENTA EN LA UGEL CHUCUITO JULI, EN EL AÑO 2023

4.3.1. Configuración J48-árbol para decisiones

Conceptualmente crea en base a una data previa estandarizada. Este estudio utiliza la herramienta WEKA, que incluye algoritmos de J48-árbol para decisiones y crea un conjunto de datos que logró encontrar la configuración óptima para análisis de confiabilidad del J48-árbol en la diversificación correcta de actas. La configuración del J48-árbol para decisiones explica en detalle a continuación con un ejercicio cómo funciona el árbol generado. Utilizando el software y el conjunto de datos WEKA, así se obtuvo J48-árbol para decisiones.

Figura 14

Configuración árbol de decisión J48 obtenido



Nota: Obtenido de (Postobón, 2023)

En la Figura 15 se visualiza algunas consideraciones tomadas en cuenta para la obtención del árbol J48 fue establecer el número de objetos mínimos que contendría cada hoja del árbol, pues si bien se podía mejorar ligeramente el resultado generando hojas con 1 solo documento clasificado, el árbol resultante de ese proceso no sería representativo para solucionar el problema inicial, ya que dicho árbol hubiera sido creado con *overfitting*; y, por definición un árbol con *overfitting* no sirve para predecir nuevos datos ajenos al data set original.

4.3.1.1. Detalle en figuras

Figura 15

Detalle en figuras obtenidas de J48-árbol para decisiones

Figura	Descripción
	Es la data set ya normalizada que entrara en el árbol de decisión.
	Característica por la cual será clasificada en ese nodo.
	Valor de la característica por la cual fue dividida o clasificada en ese nivel.
	Resultado de la clasificación u oficina a la cual sería enviada un documento de acuerdo a la clasificación del árbol de decisión J48.

Nota: Obtenido de (Postobón, 2023)

4.3.1.2. Ejemplo de funcionamiento

Caso 1

Recepción de los datos de entrada. En este caso son los valores de las características.

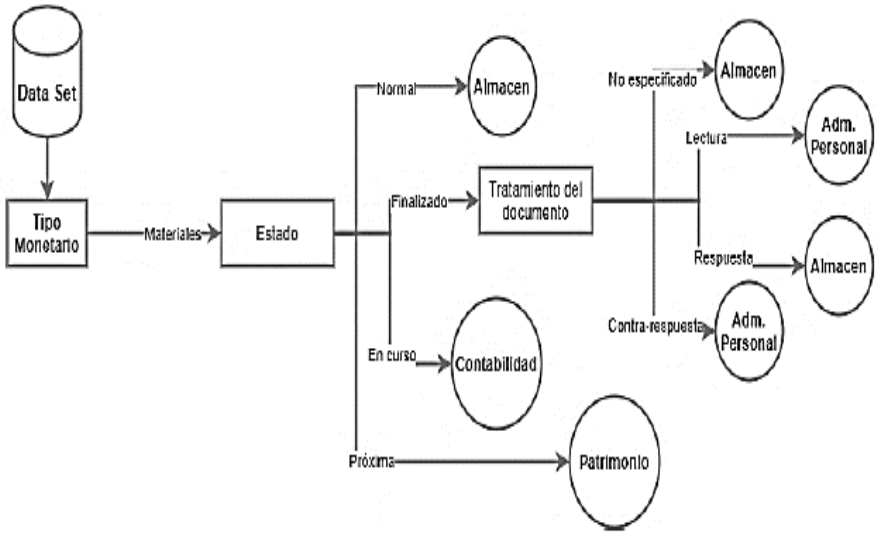
- Materiales
- Titular
- Respuesta
- Colegio
- No especifica
- Finalizado

Paso 1: Nodo de entrada, clasificación.

En cada nodo se da el proceso de clasificación con base en el valor de la característica, eligiendo así una rama u hoja del árbol de decisión J48. Para este paso, la característica del nodo es: Tipo monetario y el valor para esa característica de la entrada es Materiales". Dicho esto, se eliminan las otras ramas salientes de Tipo Monetario a excepción de "Materiales".

Figura 16

Representación del paso 1. Almacén



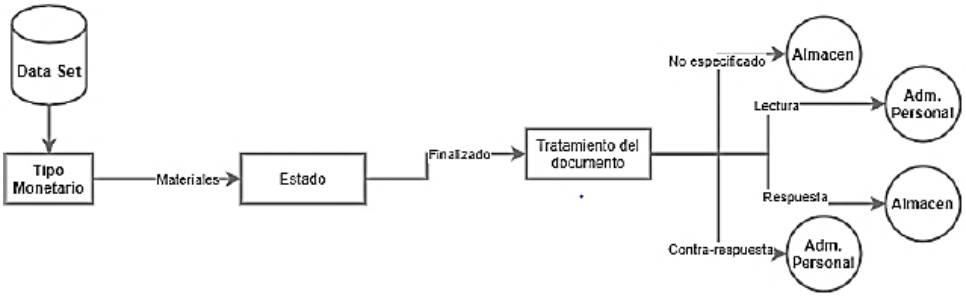
Nota: Obtenido de (Postobón, 2023)

Paso 2: Repetir la etapa anterior, para llegar a la culminación.

Ahora, la característica a evaluar es Estado y su valor en la entrada es "Finalizado". Por tanto, se eliminan las otras ramas salientes de Estado a excepción de "Finalizado".

Figura 17

Representación del paso 2, iteración primera. Almacén

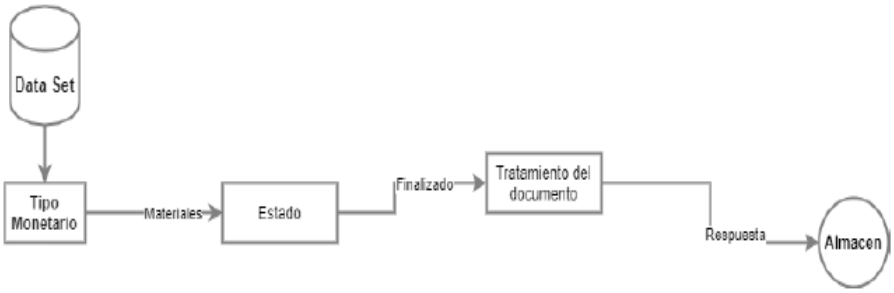


Nota: Obtenido de (Postobón, 2023)

Ahora, la característica a evaluar es Tratamiento del documento y su valor en la entrada es “Respuesta”. Por lo tanto, se eliminan las otras ramas salientes de la característica Tratamiento del documento a excepción de “Respuesta”.

Figura 18

Representación del paso 2, segunda iteración Almacen



Nota: Obtenido de (Postobón, 2023)

Paso 3: Interpretación del resultado.

Cuando ya no hay más características que evaluar, el árbol muestra la respuesta para la clasificación del documento. En este caso la respuesta es “Almacén”.

Figura 19

Representación del paso 3. Almacen

Nota: Obtenido de (Postobón, 2023)

4.3.1.3. Ocurrencias encontradas

Por la forma en que fue configurado el árbol J48, algunas características no son usadas en el proceso de generación del árbol, lo cual indica que se puede mejorar la data set mediante un mejor análisis de los datos de entrada mejorando así el árbol J48 final.

Características no usadas:

Afinidad y relación personal

4.3.2. Resultados de fiabilidad del árbol de decisión J48

Para determinar la confiabilidad de la diversificación correcta de las actas del J48-árbol, el desarrollo limita la aclaración del cuadro de inconsistencia de resultados y uso de ecuaciones de calcula la perfección. Así como la validez correlacionada que mantiene la viabilidad/confianza que se logró en el ensamblaje del árbol con los datos nuevo e ingresos.

4.3.2.1. Matriz de confusión resultante

Esta matriz de resultados se logró por medio de WEKA como herramienta y posterior clasificación como método del j48-árbol para decisiones en un grupo de data creada en el estudio.

Tabla 16

Matriz de confusión obtenida

		Valores predichos				
		Almacén	Recursos Humanos	Abastecimiento	Contabilidad	Tesorería
Valores reales	Almacén	71	4	3	0	3
	Recursos Humanos	4	29	1	0	2
	Abastecimiento	3	5	68	3	2
	Contabilidad	1	4	1	61	2
	Tesorería	5	5	4	3	66

A continuación, en la tabla 15 se presentan la data resultante del desarrollo de la diversificación de las actas que uso J48-árbol para decisiones (Matriz de confusión). Los mismos se presentan:

- 71 actas diversifican al Almacén idóneamente.
- 29 actas diversifican a Recursos Humano idóneamente
- 68 actas diversifican a Abastecimientos idóneamente
- 61 actas diversifican a contabilidad idóneamente
- 66 actas diversifican tesorería idóneamente

Interpretación

A fin de determinar la confiabilidad de la diversificación idónea de las actas de J48-árbol, el desarrollo limita la interpretación de la matriz de resultandos y la utilización de ecuaciones que permitan calcular en precisión. Además, la validación cruzada mantiene la confiabilidad o precisión lograda en el proceso de construcción del árbol incluso con nuevos datos o entradas.

Ecuaciones

Ecuación: Cálculo de exactitud o fiabilidad en la clasificación

$$Exactitud = \frac{Documentos\ clasificados\ de\ manera\ correcta}{Documentos\ clasificados\ de\ manera\ incorrecta + Documentos\ clasificados\ de\ manera\ correcta}$$

Ecuación: Cálculo de tasa de error en la clasificación

$$Tasa\ de\ error = \frac{Documentos\ clasificados\ de\ manera\ incorrecta}{Documentos\ clasificados\ de\ manera\ incorrecta + Documentos\ clasificados\ de\ manera\ correcta}$$

Ecuación: total documentos clasificados de manera correcta.

$$Documentos\ clasificados\ de\ manera\ correcta = \sum Documentos\ clasificados\ de\ manera\ correcta\ por\ la\ oficina$$

Ecuación: total documentos clasificados de manera incorrecta.

$$Documentos\ clasificados\ de\ manera\ incorrecta = \sum Documentos\ clasificados\ de\ manera\ incorrecta\ por\ la\ oficina$$

4.3.2.2. Cálculo de exactitud

Este cálculo permite conocer a detalle y confiabilidad el J48-árbol para decidir correctamente la diversificación de las actas.

$Documentos\ Clasificados\ de\ manera\ correcta$ = 71 + 29 + 68 + 61 + 66

$Documentos\ Clasificados\ de\ manera\ correcta$ = 295

$Documentos\ Clasificados\ Incorrectamente$ = 10 + 7 + 13 + 8 + 17

$Documentos\ Clasificados\ Incorrectamente$ = 55

Una vez calculados el total de documentos clasificados correcta e incorrectamente, procedemos a calcular la exactitud y tasa de error en la clasificación de documentos haciendo uso de las ecuaciones (2) y (3).

$$Exactitud = \frac{295}{295 + 55}$$

$$Exactitud = 84.29\%$$

$$Tasa\ de\ error = \frac{55}{295 + 55}$$

$$Tasa\ de\ error = 15.71$$

- Exactitud de la clasificación de documentos usando el árbol de decisión J48 = 84.29%
- Tasa de error en la clasificación de documentos usando el árbol de decisión J48 = 15.71%

4.3.2.3. Especificación de la interpretación de resultados

La interpretación de la data resultante con respecto a la diversificación de las actas por J48-árbol para decisiones en oficina, presenta un cálculo.

$$Exactitud\ por\ oficina = \left(\frac{Total\ documentos\ clasificados\ a\ la\ oficina}{Total\ documentos\ dirigidos\ a\ la\ oficina}\right) * 100$$

$Exactitud\ (Almacen)$ = (60/71)

$Exactitud\ (Almacen)$ = 84.51

Tabla 17

Resumen de exactitud árbol J48 por Oficina

Oficina	Total de documentos	Clasificados correctamente	Exactitud por oficina
Almacén	71	60	84.51
Recursos Humanos	29	21	72.41
Abastecimientos	68	62	91.18
Contabilidad	61	55	90.16
Tesorería	66	58	87.88

En la tabla 16 podemos observar

1. **Almacén (84.51%)**

La oficina de Almacén tiene una exactitud del 84.51%, lo que indica que, de los 71 documentos procesados, 60 fueron clasificados correctamente. Esta exactitud es bastante buena y sugiere que la mayoría de los documentos fueron clasificados adecuadamente, aunque aún hay margen de mejora, ya que alrededor del 15.49% de los documentos fueron clasificados incorrectamente. Este porcentaje de error puede ser relevante si el volumen de documentos sigue creciendo.

2. Recursos Humanos (72.41%)

La oficina de Recursos Humanos tiene la exactitud más baja con un 72.41%. Esto significa que, de los 29 documentos procesados, solo 21 fueron clasificados correctamente, lo que resulta en un 27.59% de error. Esto sugiere que la oficina enfrenta dificultades con la clasificación de documentos y que se podría trabajar en mejorar la precisión del proceso de

clasificación, posiblemente mediante la revisión de los criterios de clasificación o el entrenamiento del personal encargado.

3. Abastecimientos (91.18%)

La oficina de Abastecimientos tiene una exactitud muy alta del 91.18%, lo que refleja una clasificación eficiente, con 62 de los 68 documentos clasificados correctamente. Este alto porcentaje sugiere que la oficina está operando de manera efectiva en cuanto a la clasificación de documentos. Aunque siempre hay espacio para la mejora, la oficina parece estar manejando bien su flujo de trabajo.

4. Contabilidad (90.16%)

La oficina de Contabilidad tiene una excelente exactitud del 90.16%, con 55 de los 61 documentos clasificados correctamente. Esto indica que la clasificación en esta oficina es muy precisa, con solo un 9.84% de error. Este rendimiento sugiere que los procedimientos de clasificación son eficientes, aunque siempre es posible hacer ajustes menores para maximizar la precisión.

5. Tesorería (87.88%)

Finalmente, la oficina de Tesorería tiene una exactitud de 87.88%, con 58 de los 66 documentos clasificados correctamente. Este es un buen resultado, pero implica que alrededor del 12.12% de los documentos fueron clasificados incorrectamente. Como con las demás oficinas, este resultado sugiere que se puede mejorar la precisión de clasificación, pero el desempeño es bastante sólido en comparación con otras oficinas.

4.4. ANÁLISIS DESCRIPTIVO DE LAS HIPÓTESIS

Se inicia el análisis comparativo entre el estado previo (pre) y el posterior (post) a la implementación del sistema. Se recopiló información crucial de ambos casos para llevar a cabo un análisis estadístico detallado que pueda examinar a fondo la hipótesis planteada.

Indicador 1: Índice de gestión administrativa

En la siguiente tabla se muestran los resultados descriptivos:

Tabla 18

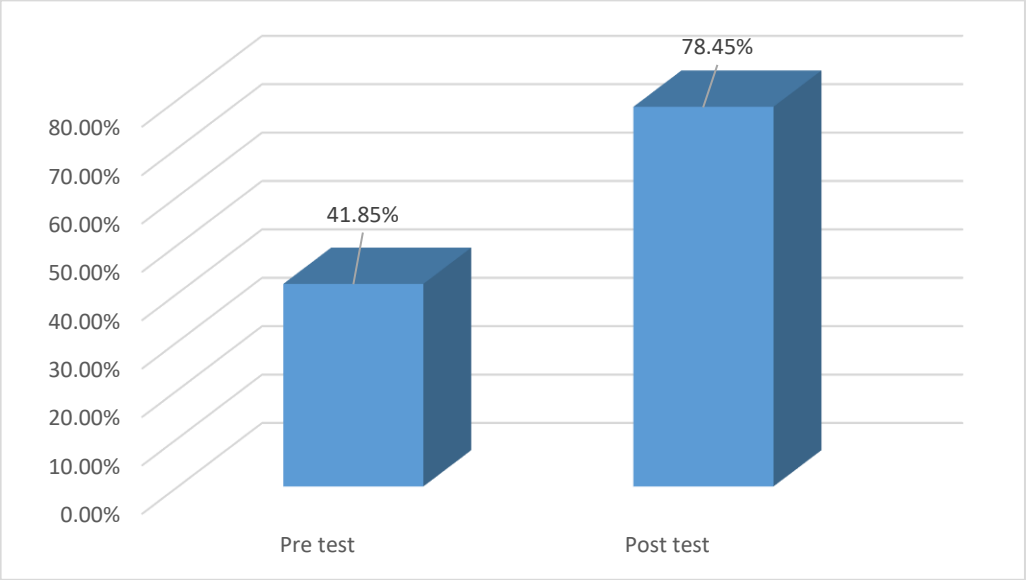
Análisis descriptivo antes y después del "árbol de decisiones" – "Índice de gestión administrativa"

Estadística descriptiva					
	N	Mínimo	Máximo	Media	Desviación
Índice de gestión administrativa pre	15	32.00	56.00	41.85	.683
Índice de gestión administrativa post	15	61.00	91.00	78.45	.844
N valido por lista	15				

Se puede observar que el índice de gestión administrativa, el "valor promedio" era 41, 85 %; y, ahora el "valor promedio" es 78, 45 % de los datos. El índice de gestión administrativa ha aumentado en forma gigantesca; además, el valor mínimo de la prueba anterior es 32 %, el valor máximo es 56 %, el valor mínimo de la post prueba es 61 % y el máximo es 91 %. En cuanto a la dispersión del índice de gestión administrativa, hay un cambio del 6 % en el pre test y del 8 % en el post test.

Figura 20

Índice de rotación de stock antes y después del sistema web



Análisis

En la Figura 21 se observan los resultados de un pre test con un 41.85% y un post test con un 78.45% reflejan una mejora significativa en el desempeño o conocimiento de los participantes entre la evaluación inicial y la final. Este cambio indica que las estrategias, intervenciones o métodos implementados durante el período entre ambos test han tenido un impacto positivo en los resultados.

4.5. CONTRASTACIÓN DE HIPÓTESIS

4.5.1. Hipótesis general

Hipótesis alterna: La construcción de un árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, mejora a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.

Hipótesis nula: La construcción de un árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, **NO** mejora a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.

Tabla 19

Prueba de Chi cuadrado de Pearson del objetivo general

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	292,056[a]	2	<.001
Razón de verosimilitud	311,456	2	<.001
Asociación lineal por lineal	287,159	1	<.001
N de casos	350		

Nota. Obtenido de SPSS versión 27

Análisis

En La tabla 18observamos la prueba de Chi-cuadrado de Pearson nos da un valor de la prueba es 292,056 con 2 grados de libertad (gl), y la significación asintótica es <.000 que es mucho menor que la significancia de 0.05. Esto sugiere que se acepte la hipótesis alterna y rechace la hipótesis nula, y se concluye que la implementación del árbol de decisiones J48 mejora significativamente la gestión documentaria.

4.5.2. Hipótesis especifica 1

Hipótesis alterna: El proceso de clasificación de documentos actual, mejora mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.

Hipótesis nula: El proceso de clasificación de documentos actual, **NO** mejora mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.

Tabla 20

Prueba de Chi cuadrado de Pearson el objetivo específico 1

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	291,899[a]	2	<.001
Razón de verosimilitud	328,440	2	<.001
Asociación lineal por lineal	280,554	1	<.001
N de casos	350		

Nota. Obtenido de SPSS versión 27

Análisis

En la Tabla 19 observamos la prueba de Chi-cuadrado de Pearson nos da un valor de la prueba es 291,899[a] con 2 grados de libertad (gl), y la significación asintótica es <.001 es mucho menor que la significancia de 0.05. Esto sugiere que se acepte la hipótesis alterna y rechace la hipótesis nula, y se concluye que el diagrama BPMN del árbol de decisiones J48 mejora significativamente el proceso de clasificación de documentos actual.

4.5.3. Hipótesis especifica 2

Hipótesis alterna: La preparación del dataset, para el árbol de decisión J48, mejora mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023.

Hipótesis nula: La preparación del dataset, para el árbol de decisión J48, **NO** mejora mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023.

Tabla 21

Prueba de Chi cuadrado de Pearson del objetivo específico 2

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	301,687[a]	2	<.001
Razón de verosimilitud	326,382	2	<.001
Asociación lineal por lineal	288,970	1	<.001
N de casos	350		

Nota. Obtenido de SPSS versión 27

Análisis

En la Tabla 20 observamos la prueba de Chi-cuadrado de Pearson nos da un valor de la prueba es 301,687 con 2 grados de libertad (gl), y la significación asintótica es <.001 es mucho menor que la significancia de 0.05. Esto sugiere que se acepte la hipótesis alterna y rechace la hipótesis nula, y se concluye que la preparación del dataset, para el árbol de decisión J48 mejora significativamente el enriquecimiento de la semántica y clasificación de las palabras claves.

5.2.2. Hipótesis especifica 3

Hipótesis alterna: La construcción del árbol de decisión J48 mejora, utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.

Hipótesis nula: La construcción del árbol de decisión J48 **NO** mejora, utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.

Tabla 22

Prueba de Chi cuadrado de Pearson del objetivo específico 3

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	311,042[a]	2	<.001
Razón de verosimilitud	339,719	2	<.001
Asociación lineal por lineal	305,854	1	<.001
N de casos	350		

Nota. Obtenido de SPSS versión 27

Análisis

En la Tabla 21 observamos la prueba de Chi-cuadrado de Pearson nos da un valor de la prueba es 311,042 con 2 grados de libertad (gl), y la significación asintótica es <.001 es mucho menor que la significancia de 0.05. Esto sugiere que se acepte la hipótesis alterna y rechace la hipótesis nula, y se concluye que la construcción del árbol de decisión J48 mejora significativamente utilizando la herramienta WEKA.

4.6. DISCUSIÓN

En este estudio, de los resultados de la implementación del árbol de decisiones J48 en la gestión documentaria de la UGEL Chucuito, usando un dataset normalizado en WEKA, ha mostrado buenos resultados. Con una exactitud del 84.29 %, el modelo es efectivo en clasificar documentos. Sin embargo, la tasa de error de 15.71 % sugiere que se pueden realizar ajustes para mejorar la precisión. En general, el modelo es exitoso y optimizable para una mejor clasificación. De los resultados podemos contrastar con el estudio de Quenta (2021) el cálculo del % exacto se clasifica en 86, 63 % con tasa de error 17, 37 %, además se puede decir que el pre test es 80, 24 % y el post test de 82,63 % mejorando la gestión documentaria en la entidad. Pero, estos estudios coinciden con el

estudio de Zambrana (2020) quien dice que el cálculo % exacto se clasifica en 78, 45 % y tasa de error 19, 80 %, además se puede decir que el pre test es 44, 5 % y el post test es de 80, 15% mejorando notablemente la gestión documentaria en la entidad. Es por ello que se puede decir que la hipótesis general es válida.

De los resultados de la implementación de la data set, a partir del árbol J48 es un proceso a largo plazo y difícil que se realiza a través de estandarizar la data y brindar soporte al personal en cargo de diversificación de actas por la entidad estudiada, sin utilización de herramienta alguna. La data-set generada se adecuó para el uso de 48-árbol como idónea. De la misma forma se contrasto con (Quenta, 2021) quien mensiona que la generación de la data set necesaria para el árbol J48 evaluado en estetrabajo de investigación fue el proceso más largo y complicado, siendo necesarios múltiples técnicas de normalización de datos y ayuda por parte de las personas a cargo de clasificar los documentos de la UGEL sin el uso de ninguna herramienta. La data set generada tuvo que ser adecuada para su uso por parte del árbol J48.

De los resultados, al realizar la modificación del J48-árbol para decisiones se necesita la data-set estandarizada y con la aplicación de WEKA como herramienta. A fin es necesario la utilización de validez cruzada en el proceso de capacitación por el J48-árbol para decisiones donde no se presente problemas de overfitting. También se contrasto con (Quenta, 2021) quien afirma que para la correcta configuración del árbol de decisión J48 fue necesario una data set normalizada y conocimiento previo del proceso de configuración de un árbol de decisión en la herramienta WEKA. Adicionalmente, fue necesario el uso de la validación cruzada durante el proceso de entrenamiento para generar un árbol de decisión J48 que no tenga problemas de Overfitting.

De los resultados de la construcción del árbol de decisión J48 utilizando WEKA en la UGEL Chucuito ha sido exitosa, alcanzando una exactitud del 84.29% en la clasificación de documentos. El modelo ha demostrado ser una herramienta efectiva para mejorar la gestión documentaria, mostrando un buen desempeño en la clasificación y proporcionando una base sólida para optimizar procesos futuros en la gestión de documentos dentro de la institución. Finalmente se discute con (Quenta, 2021) quien dice que la determinar la fiabilidad o exactitud del árbol de decisión J48 en la correcta clasificación de documentos fue simple, pues el proceso se limitó a interpretar la matriz de confusión resultante de la clasificación de documentos usando el árbol de decisión J48, obteniendo así una fiabilidad del 82.63%.

V. CONCLUSIONES

PRIMERA: Se concluye que la implementación del árbol de decisiones J48 en la gestión documentaria de la UGEL Chucuito, usando un dataset normalizado en WEKA, ha mostrado buenos resultados. Con una exactitud del 84.29 %, el modelo es efectivo en clasificar documentos. Sin embargo, la tasa de error de 15.71 % sugiere que se pueden realizar ajustes para mejorar la precisión. En general, el modelo es exitoso y optimizable para una mejor clasificación.

SEGUNDA: Se concluye que la implementación de la data set, a partir del árbol J48 es un proceso a largo plazo y difícil que se realiza a través de estandarizar la data y brindar soporte al personal en cargo de diversificación de actas por la entidad estudiada, sin utilización de herramienta alguna. La data-set generada se adecuó para el uso de 48-árbol como idónea.

TERCERA: Se concluye que, al realizar la modificación del J48-árbol para decisiones se necesita la data-set estandarizada y con la aplicación de WEKA como herramienta. A fin es necesario la utilización de validez cruzada en el proceso de capacitación por el J48-árbol para decisiones donde no se presente problemas de overfitting.

CUARTA: Se concluye que la construcción del árbol de decisión J48 utilizando WEKA en la UGEL Chucuito ha sido exitosa, alcanzando una exactitud del 84.29% en la clasificación de documentos. El modelo ha demostrado ser una herramienta efectiva para mejorar la gestión documentaria, mostrando un buen desempeño en la clasificación y proporcionando una

base sólida para optimizar procesos futuros en la gestión de documentos dentro de la institución.

VI. RECOMENDACIONES

PRIMERA: Se recomienda realizar un seguimiento continuo del modelo J48, evaluando regularmente su exactitud y ajustando los parámetros según sea necesario. Además, se sugiere explorar métodos para reducir la tasa de error, como la inclusión de más datos o la mejora del preprocesamiento.

SEGUNDA: Es fundamental estandarizar el dataset de manera adecuada y proporcionar capacitación constante al personal encargado de la diversificación de actas. Además, se debe considerar la implementación de herramientas tecnológicas que faciliten la gestión de documentos y la correcta clasificación de la data.

TERCERA: Se recomienda utilizar la técnica de validación cruzada para evitar problemas de sobreajuste (overfitting) en el árbol J48, asegurando que el modelo sea robusto y generalizable. Además, se debe mejorar la capacitación del personal para garantizar la correcta aplicación del modelo.

CUARTA: Es importante seguir optimizando el árbol de decisión J48, experimentando con diferentes configuraciones y ajustando su rendimiento en función de los resultados obtenidos. Asimismo, se debe seguir monitoreando su efectividad para asegurar que se mantenga útil y eficiente en el largo plazo.

VII. REFERENCIAS BIBLIOGRÁFICAS

Agenjo, X., & Hernández, F. (2022). “Sobre el enriquecimiento semántico de materias en un entorno Linked Open Data”. *Anuario ThinkEPI, 16*(e16a017), 14. doi:https://doi.org/10.3145/thinkepi.2022.e16a17/10.3145/thinkepi.2022.e

Akujuobi, U. y Zhang, X. (2017). Delve: a dataset-driven scholarly search and analysis system. *ACM SIGKDD Explorations Newsletter, 19*(2), 36–46. doi:10.1145/3166054.3166059

Alania Ricaldi, P. F. (2019). *Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión.* [Tesis de licenciatura, Universidad Nacional Daniel Alcides Carrión]. Obtenido de http://repositorio.undac.edu.pe/handle/undac/829

Anaya, T., Montalvo, J., Ignacio, A., & Arispe, C. (2021). Escuelas rurales en el Perú: factores que acentúan las brechas digitales en tiempos de pandemia (COVID-19) y recomendaciones para reducirlas*. *Educación*.

Balbuena, J. (2022). *Modelos de Detección de Emociones en Texto y Rostros para Agentes Conversacionales Multimodales* .

Barrientos, E., Coronel, L., Cuesta, F., & Dwwar, R. (2019). Sistema de administración de ventas tienda a tienda:Aplicando técnicas de inteligencia artificial. *RISTI*. Obtenido de https://www.researchgate.net/profile/Dewar-Rico-Bautista/publication/339227416_Sistema_de_administracion_de_ventas_tienda_a_tienda_Aplicando_tecnicas_de_inteligencia_artificial/links/5e9fd671a6fdcc20bb360b63/Sistema-de-administracion-de-ventas-tienda-a-ti

Carrasco, S. (2019). *Metodología de la investigación científica. Pautas metodológicas para diseñar y elaborar el proyecto de investigación.* Lima: Editorial San Marcos E.I.R.L. LTDA. Obtenido de https://www.academia.edu/26909781/Metodologia_de_La_Investigacion_Cientifica_Carrasco_Diaz_1_

Choque, E. (2023). *Sistema IoT para el control medio ambiental de la ciudad de Puno, 2023.* Lima: Universidad César Vallejo. Obtenido de https://repositorio.ucv.edu.pe/handle/20.500.12692/118230

Cordero, E., Erazo, A., & Cordero, D. (2020). Soluciones corporativas de inteligencia de negocios en las pequeñas y medianas empresas. *Revista Arbitrada Interdisciplinaria Koinonía*. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=7439114

Espino Quinones, L. y Garcia Torres, M. E. (2018). *Aplicación de minería de datos basado en árboles de decisión para predecir el riesgo de morosidad de los clientes en la empresa de seguros Oncosalud S.A.C. 2018.* [Tesis de licenciatura, Universidad Autónoma del Perú]. Obtenido de https://hdl.handle.net/20.500.13067/700

Garces Eslava, D. M. (2020). *Método de procesamiento de lenguaje natural y técnicas de minería de datos aplicadas a la clasificación de incidentes informáticos.* [Tesis de licenciatura, Universidad de Lima]. Obtenido de https://repositorio.ulima.edu.pe/handle/20.500.12724/11703

Henriquez, K. (2022). Aplicación de las tecnologías de la información y la comunicación (TIC) en el derecho: ¿amenaza u oportunidad? *Saber y Justicia*. Obtenido de https://saberyjusticia.enj.org/index.php/SJ/article/view/140

Hernandez, R. y Mendoza, C. (2018). *Metodología de la investigación.* México: Mc Graw Hill. Obtenido de https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf

Ilyas, N., Shahzad, A., & Kim, K. (2020). Convolutional-Neural Network-Based Image Crowd Counting: Review, Categorization, Analysis, and Performance Evaluation. *MDPI*. doi:https://doi.org/10.3390/s20010043

Jiménez, M. (2020). *Un modelo para la obtención de interacciones medicamentos mediante aprendizaje profundo sobre el corpus de extracción 2013.*

Karabadji, N. E., Seridi, H., Bousetouane, F., Dhifli, W. y Aridhi, S. (2017). An evolutionary scheme for decision tree construction. *Knowledge-Based Systems, 119*, 166-177. doi:10.1016/j.knosys.2016.12.011

Kastrati, Z., Imran, A. y Yayilgan, S. (2019). El impacto del aprendizaje profundo en la clasificación de documentos usando representaciones semánticamente ricas. *Information Processing & Management*, *56*(5), 1618-1632. doi:10.1016/j.ipm.2019.05.003

Muñoz, Y., & Moreno, L. (2020). *Implementación de un algoritmo para la clasificación automática de lenguaje de señas colombiano en video usando aprendizaje profundo.*

Quenta Banegas, J. (2021). *Análisis de fiabilidad del árbol de decisión J48 en la correcta clasificación de documentos en la Unidad de Gestión Educativa Local El Collao-Ilave.* [Tesis de licenciatura, Universidad Nacional del Altiplano]. Obtenido de http://repositorio.unap.edu.pe/handle/UNAP/16338

Ripley, B. D. (2014). *Pattern Recognition and Neural Networks.* Oxford: Cambridge University Press. doi:10.1017/CBO9780511812651

Ru, Z., Hua, D., Wenbin, X., Sha, W., Lu, G., Qian, X. y Jinjiao, L. (2022). El modelo de árbol de decisiones predice el nacimiento vivo después de la cirugía para las adherencias intrauterinas de moderadas a graves. *BMC Pregnancy Childbirth*(78). doi:10.1186/s12884-022-04375-x

Ruelas, R. (2023). *Sistema inteligente para el Control de aforo en Institución de Salud Pública, Puno 2023.* Lima: Universidad César Vallejo. Obtenido de https://repositorio.ucv.edu.pe/handle/20.500.12692/113063

Simón, A., García, M., Puebla, M., Sánchez, N., Perea, J., & Comas, R. (2022). Enfoques semánticos basados en ontologías para la recuperación de información en Sistemas de Información Geográfica. *Anales de la Academia de Ciencias de Cuba*, *12*(1), 9. doi:chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://scielo.sld.cu/pdf/aacc/v12n1/2304-0106-aacc-12-01-e1124.pdf

Valero, C. (2021). Derecho e Inteligencia Artificial en el mundo de hoy: escenarios internacionales y los desafíos que representan para el Perú. *THEMIS*. doi:https://doi.org/10.18800/themis.202101.017

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series,, 1168*(2). doi:10.1088/1742-6596/1168/2/022022

Zambrana Cárdenas, J. A. (2020). *Automatización de flujos de aprobación de gestión en procesos de negocios mediante árboles de decisiones/. Bolivia: . Obtenido de.* [Tesis de maestría, Universidad Mayor de San Andres]. Obtenido de https://repositorio.umsa.bo/handle/123456789/27698

ANEXOS

Anexo 1 Matriz de Consistencia

Título: Análisis de la gestión documentaria mediante el árbol de decisiones J48 para una data set normalizada usando la herramienta WEKA, en la UGEL Chucuito, Juli 2023.

PROBLEMA	OBJETIVO	HIPÓTESIS	VARIABLES	METODOLOGÍA
Problema general	**Objetivo general**	**Hipótesis general**	**Variable Independiente:** Árbol de decisiones J48 **Variable Dependiente:** Herramienta WEKA	**Enfoque:** • Cuantitativo **Tipo de investigación:** • Básica **Nivel de Investigación:** • Descriptiva y explicativa **Diseño:** • Experimental **Población:** • 1000 documentos **Muestra:** • 350 documentos **Técnica e instrumentos:** **Técnica:** • Análisis documental **Instrumentos:** • Guía de análisis documental **Métodos de análisis de datos:** • J48 • SPSS
¿Cómo influye el árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023?	Implementar el árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.	La construcción de un árbol de decisiones J48 en la gestión documentaria, utilizando un dataset normalizado, mejora a través de la herramienta WEKA, en la UGEL Chucuito, Juli 2023.		
Problemas específicos	**Objetivos específicos**	**Hipótesis específicas**		
P.E.1 ¿Cómo se puede optimizar el proceso de clasificación de documentos actual, mediante el Diagrama BPMN, en la UGEL Chucuito, Juli 2023??	**O.E.1** Describir el proceso de clasificación de documentos actual, mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.	**H.E.1** El proceso de clasificación de documentos actual, mejora mediante el diagrama BPMN, en la UGEL Chucuito, Juli 2023.		
P.E.2 ¿Cómo se puede mejorar la precisión y relevancia del modelo predictivo J48 al enriquecer semánticamente un dataset de documentos clasificados Chucuito, mediante el análisis y clasificación de palabras clave en la UGEL Chucuito, Juli 2023?	**O.E.2** Preparar la dataset, para el árbol de decisión J48, mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023.	**H.E.2** La preparación del dataset, para el árbol de decisión J48, mejora mediante el enriquecimiento de la semántica y clasificación de las palabras claves en la UGEL Chucuito, Juli 2023		
P.E.3 ¿Cómo se puede construir el árbol de decisión J48 utilizando la herramienta WEKA para clasificar los documentos en la UGEL Chucuito, Juli 2023?	**O.E.3** Construir el árbol de decisión J48 utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.	**H.E.3** La construcción del árbol de decisión J48 mejora, utilizando la herramienta WEKA en la UGEL Chucuito, Juli 2023.		

Anexo 2 Solicitud de la gestión documentaria

MESA DE PARTES
N° Exp.: 12559
N° Folios:
Fecha: 23 NOV. 2022
Firma: Hora: 14:47

SOLICITA: AUTORIZACIÓN PARA REALIZAR TRABAJO DE INVESTIGACIÓN

SEÑOR DIRECTOR DE LA UGEL CHUCUITO JULI
Mg. José Gabriel Vizcarra Fajardo

Yo, Jhenery Anyela Carbajal Pari, bachiller de Ingeniería de Sistemas de la Universidad Nacional del Altiplano Puno, identificada con DNI N° 72438631, domiciliada en el Jr. Pucara N° 403, ante Ud. con el debido respeto me presento y expongo:

Que, en la actualidad me encuentro desarrolando mi Proyecto de Tesis para obtener mi Grado de Ingeniero de Sistemas, proyecto de Tesis que lleva por título "Análisis del árbol de decisiones J48 para mejorar la gestión documentaria en la UGEL Chucuito, Juli, 2023" en beneficio de todo el Magisterio de la UGEL Chucuito Juli .Por tal motivo recurro a su digna autoridad para SOLICITAR la autorización para ejecutar el trabajo de investigación implementando una estrategia para mejorar la gestión documentaria, para lo cual solicito apoyo con documentos requeridos para su ejecución.

Por lo expuesto,

Ruego a usted Señor Director atender a mi solicitud, por ser de justicia y legal.

Jhenery Anyela Carbajal Pari
DNI N° 72438631

Anexo 3 Declaración Jurada de autenticidad de Tesis

Universidad Nacional del Altiplano Puno | VRI Vicerrectorado de Investigación | Repositorio Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Jhenery Angela Carbajal Pari,
identificado con DNI 72428631 en mi condición de egresado de:

☒ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**
Ingeniería de Sistemas,
informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación** denominada:
"ANÁLISIS DE LA GESTIÓN DOCUMENTARIA MEDIANTE EL ARBOL DE DECISIONES J48 PARA UNA DATA SET NORMALIZADA USANDO LA HERRAMIENTA WEKA, EN LA UGEL CHUCUITO JULI 2023"

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 11 de diciembre del 2024

FIRMA (obligatoria) Huella

Universidad Nacional del Altiplano Puno | VRI Vicerrectorado de Investigación | Repositorio Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Roger Wilber Tapia Barrios,
identificado con DNI 72251137 en mi condición de egresado de:

☑ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

_____,

informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación** denominada:
" Análisis de la Gestión Documentaria Mediante el árbol de Decisiones J48 Para una Data Set Normalizada Usando la Herramienta Weka, en la UGEL Chucuito, Juli 2023 "

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 11 de diciembre del 2024

FIRMA (obligatoria) Huella

Anexo 4 Autorización para el depósito de tesis en el repositorio institucional

Universidad Nacional del Altiplano Puno | VRI Vicerrectorado de Investigación | Repositorio Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Jhenery Angela Carbajal Pari,
identificado con DNI 72438631 en mi condición de egresado de:

☑ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ingeniería de Sistemas,

informo que he elaborado el/la ☑ **Tesis** o ☐ **Trabajo de Investigación** denominada:

"ANÁLISIS DE LA GESTIÓN DOCUMENTARIA MEDIANTE EL ÁRBOL DE DECISIONES J48 PARA UNA DATA SET NORMALIZADA USANDO LA HERRAMIENTA WEKA EN LA UGEL CHUCUITO JULI 2023"

para la obtención de ☐**Grado,** ☒ **Título Profesional** o ☐ **Segunda Especialidad.**

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: https://creativecommons.org/licenses/by-nc-sa/4.0/

En señal de conformidad, suscribo el presente documento.

Puno 11 de diciembre del 2024

FIRMA (obligatoria)

Huella

Universidad Nacional del Altiplano Puno | VRI Vicerrectorado de Investigación | Repositorio Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Roger Wilber Tapia Barrios,
identificado con DNI 72251137 en mi condición de egresado de:

☒ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ingeniería de Sistemas,

informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación** denominada:

" Análisis de la Gestión Documentaria mediante el árbol de Decisiones J48 Para una Data Set Normalizada usando la Herramienta Weka, en la UGEL Chucuito, Juli 2023 "

para la obtención de ☐**Grado,** ☒ **Título Profesional** o ☐ **Segunda Especialidad.**

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: https://creativecommons.org/licenses/by-nc-sa/4.0/

En señal de conformidad, suscribo el presente documento.

Puno 11 de diciembre del 2024

FIRMA (obligatoria)

Huella