



UNIVERSIDAD CÉSAR VALLEJO

ESCUELA DE POSGRADO

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

Procesamiento del lenguaje natural y productividad en los expedientes
judiciales en un distrito judicial Lima Norte, 2023

TESIS PARA OBTENER EL GRADO ACADÉMICO DE:

Maestro en Ingeniería de Sistemas con mención en Tecnología de la Información

AUTOR:

Lucero Carrillo, Erick David (orcid.org/0000-0002-3416-6995)

ASESORES:

Mg. Poletti Gaitán, Eduardo Humberto (orcid.org/0000-0002-2143-4444)

Dr. Pereyra Acosta, Manuel Antonio (orcid.org/0000-0002-2593-5772)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2023

DEDICATORIA

Agradezco a Dios por la salud, fortaleza y sabiduría que me ha otorgado. Expreso mi profundo agradecimiento a mis amados padres, abuelos y familiares, quienes me han enseñado la invaluable lección de la perseverancia. Dedico también mi gratitud a mi compañera de vida, tíos, hermanos y seres queridos, quienes han iluminado mi camino con su amor y apoyo incondicional.

AGRADECIMIENTO

Agradezco a Dios por permitirme cumplir mi sueño de culminar mi grado de magíster, también a la casa estudio por la formación académica. A los magníficos docentes (asesor y coasesor). También a mis compañeros de aula. Y por supuesto a la institución pública la cual me apoyo para continuar con la investigación.



ESCUELA DE POSGRADO

MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN

Declaratoria de Autenticidad del Asesor

Yo, POLETTI GAITAN EDUARDO HUMBERTO, docente de la ESCUELA DE POSGRADO MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Procesamiento del lenguaje natural y productividad en los expedientes judiciales en un distrito judicial Lima Norte, 2023", cuyo autor es LUCERO CARRILLO ERICK DAVID, constato que la investigación tiene un índice de similitud de 12.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 04 de Enero del 2024

Apellidos y Nombres del Asesor:	Firma
POLETTI GAITAN EDUARDO HUMBERTO DNI: 18073124 ORCID: 0000-0002-2143-4444	Firmado electrónicamente por: EPOLETTIG el 07-01- 2024 17:25:09

Código documento Trilce: TRI - 0719728



ESCUELA DE POSGRADO

MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN

Declaratoria de Originalidad del Autor

Yo, LUCERO CARRILLO ERICK DAVID estudiante de la ESCUELA DE POSGRADO del programa de MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaro bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Procesamiento del lenguaje natural y productividad en los expedientes judiciales en un distrito judicial Lima Norte, 2023", es de mi autoría, por lo tanto, declaro que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. He mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
LUCERO CARRILLO ERICK DAVID DNI: 71574421 ORCID: 0000-0002-3416-6995	Firmado electrónicamente por: ELUCEROC el 08-01- 2024 08:16:48

Código documento Trilce: INV - 1461237

ÍNDICE DE CONTENIDOS

DEDICATORIA	ii
AGRADECIMIENTO	iii
DECLARATORIA DE AUTENTICIDAD DEL ASESOR	iv
DECLARATORIA DE ORIGINALIDAD DEL AUTOR	v
ÍNDICE DE TABLAS	vii
RESUMEN	viii
ABSTRACT	ix
I.INTRODUCCIÓN	11
II.MARCO TEÓRICO	15
III.METODOLOGÍA	22
3.1. Tipo y diseño de investigación	22
3.2. Variables y operacionalización	23
3.3. Población, muestra y muestreo	24
3.3.1.Población	24
3.3.2.Muestra	24
3.3.3.Muestreo	24
3.3.4 Unidad de análisis	25
3.4. Técnicas e instrumentos de recolección de datos	25
3.5. Procedimientos	25
3.6. Método de análisis de datos	26
3.7. Aspectos éticos	26
IV.RESULTADOS	27
V.DISCUSIÓN	36
VI.CONCLUSIONES	43
VII.RECOMENDACIONES	44
REFERENCIAS	45
ANEXOS	52

ÍNDICE DE TABLAS

Tabla 1. Alfa de Cronbach	25
Tabla 2. Resultados de prueba de normalidad	28
Tabla 3. Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural- dimensión preprocesamiento	28
Tabla 4. Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural- dimensión extracción de datos	29
Tabla 5. Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural- dimensión tiempo de registro	31
Tabla 6. Prueba de correlación de la dimensión preprocesamiento de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.	32
Tabla 7. Prueba de correlación de la dimensión extracción de datos de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.	34
Tabla 8. Prueba de correlación de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.	35

RESUMEN

La tesis considera determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023. Siendo la metodología que parte de los cuantitativo, de tipo aplicada, con diseño no experimental de tipo descriptivo – correlacional y de corte transversal y usando el instrumento de guía de observación documental para identificar la relación entre las variables descritas acorde a los objetivos de estudio.

Los resultados han mostrado que las dimensiones de preprocesamiento y extracción de datos tienen una asociación positiva y significativa con la variable productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023, usando el análisis inferencial.

Se concluyó la relevancia de una relación positiva entre las variables de la investigación, evidenciando un coeficiente Rho de Spearman entre procesamiento del lenguaje natural y la productividad de 0.962; lo que nos brinda un importante efecto del procesamiento del lenguaje natural que nos ayudaría a identificar y extraer información clave de documentos legales de manera automática, facilitando la revisión y comprensión de los documentos, ahorrando tiempo al personal jurisdiccional.

Palabras claves: procesamiento del lenguaje natural, productividad, influencia

ABSTRACT

The thesis considers determining the influence of natural language processing on the productivity of judicial files in a judicial district of Lima Norte, 2023. Being the methodology that starts from the quantitative, basic type, with non-experimental design of descriptive - correlational and cross-sectional type and using the instrument of documentary observation guide to identify the relationship between the variables described according to the objectives of the study.

The results have shown that the dimensions of preprocessing and data extraction have a positive and significant association with the variable productivity of judicial files in a judicial district Lima Norte, 2023, using inferential analysis.

We concluded the relevance of a positive relationship between the research variables, evidencing a Spearman's Rho coefficient between natural language processing and productivity of 0.962; which gives us an important effect of natural language processing that would help us to identify and extract key information from legal documents automatically, facilitating the review and understanding of documents, saving time to jurisdictional staff.

Keywords: natural language processing, productivity, influence

I. INTRODUCCIÓN

En la tendencia de la inteligencia artificial se considera a las computadoras a interactuar con el lenguaje humano. Este es un nuevo método para procesar grandes cantidades de datos, conocido como caja negra en la lógica del procesamiento de información. Dado que los algoritmos efectivos ayudan a tomar decisiones en una variedad de procesos complejos, el PLN ha ayudado a brindar soluciones en las últimas décadas en diferentes aspectos. (Fernández y Esteves, 2022).

El procesamiento del lenguaje natural considera un aspecto crucial en el desempeño de expedientes judiciales, ya que permite la extracción de información valiosa y la automatización de tareas relacionadas con el análisis de documentos legales, siendo expresado el problema que los expedientes judiciales enfrenta desafíos significativos en términos de tiempo debido al volumen de documentos, la complejidad del contenido legal, la diversidad de idiomas, la necesidad de mantener la privacidad y seguridad, y la búsqueda eficiente de información relevante. Superar estos desafíos requiere una planificación adecuada y ejecución de análisis informáticos avanzados como el NLP para la priorización de tiempo y los recursos dedicados a la gestión y análisis de expedientes legales (Barsha y Munshi, 2023; Kulkarni & Shivananda, 2021).

Ramalho (2017) considera que los actores de justicia utilizan el lenguaje natural para tomar decisiones en el campo del derecho; sin embargo, aún existen vacíos argumentativos y se ha evidenciado problemas en los operadores jurídicos al conceder información confusa; siendo en el contexto del desarrollo socio tecnológico de vital importancia para el aporte desde la ciencia del derecho.

Se ha demostrado que el PLN es capaz de desarrollar algoritmos que contribuyan a un cambio mental para lograr modelos de argumentación, ordenamiento de textos, clasificar fuentes de datos, creación y priorización del aspecto legislativo, así como los sistemas de PLN pueden ayudar al personal jurisdiccional a buscar y recuperar información de manera más eficiente en grandes volúmenes de documentos legales. Esto puede acelerar la revisión de expedientes y facilitar la identificación de precedentes legales relevantes. (Barsha y Munshi, 2023).

Los sistemas de PLN pueden analizar documentos legales para resumirlos o extraer información importante, como fechas clave, nombres de las partes involucradas, argumentos legales y decisiones judiciales. Esto permite a los abogados obtener rápidamente una visión general de un expediente. De esta forma se puede analizar documentos legales para resumirlos o extraer información importante, como fechas clave, nombres de las partes involucradas, argumentos legales y decisiones judiciales. Priorizando obtener rápidamente una información rápida de un expediente (Baca et al., 2023).

En el avance de la investigación y desarrollo sobre la innovación en los gestores públicos; así como la modernización del sistema de justicia que involucra que los servicios públicos orientados al proceso de justicia funcionan con sistemas de Inteligencia Artificial, que incluyen el aspecto de reconocimiento facial, recopilación de datos a través de los teléfonos inteligentes para cuantificar el comportamiento línea para la búsqueda de eficiencia y transparencia en el sistema judicial peruano, para explorar posibilidades que permita extraer información importante (Farfán et al.,2023).

En el Perú, el sistema judicial parte de enfrentar desafíos similares, donde automatizar tareas repetitivas, como la revisión de documentos estándar, la generación de documentos legales básicos y la elaboración de informes, lo que ahorra tiempo y reduce errores humanos. Originando la búsqueda de eficiencia y transparencia en el sistema judicial peruano ha llevado a explorar las posibilidades que ofrece la IA y el PLN en esta área específica de extracción de información (Quispe, 2021).

No obstante, la aplicación de estas tecnologías no es homogénea en todo el país. Lima, como el procesamiento del lenguaje natural que puede aumentar la productividad en el ámbito legal al automatizar tareas, facilitar la búsqueda y organización de información, y proporcionar análisis más profundos de documentos legales. Esto no solo ahorra tiempo y recursos, sino que también puede generar discusiones y tomar aspectos relevantes en el sistema judicial. Sin embargo, es importante recordar que la aplicación de tecnología NLP en el ámbito legal debe cumplir con las regulaciones y estándares éticos para garantizar la confidencialidad y la equidad en los procedimientos judiciales.

En esta investigación se propone explorar cómo el NLP influye en la productividad de los casos judiciales en un distrito judicial Lima Norte, considerando que optimice tiempos y necesidades, de esta forma contribuir a la precisión en el aspecto normativo. Además, se examinarán los desafíos y oportunidades que estas tecnologías presentan en el sistema judicial peruano, con el objetivo de aportar para una administración de justicia con rapidez y ahorro de tiempo.

De esta forma la problemática que se estudiará será: ¿Cómo el procesamiento del lenguaje natural influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023? siendo los problemas específicos: ¿Cómo el preprocesamiento influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023?; ¿Cómo la extracción de datos influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023?.

Desde una perspectiva teórica, la relevancia del tema parte de analizar debido a la necesidad de la celeridad de los expedientes judiciales para lograr ser atendidos de forma oportuna. Para lo cual el procesamiento del lenguaje natural ayuda a sistematizar los sistemas de justicia y aumentar la productividad de los expedientes judiciales. El aspecto metodológico se aborda un análisis correlacional para determinar la asociación de las variables estudiadas y sus respectivas hipótesis.

La justificación práctica en el contexto para la transformación digital del estado, donde la relevancia del estudio parte de acelerar la identificación y extracción de información clave, lo que ahorra tiempo y recursos valiosos de abogados, jueces y personal judicial; considerando que los sistemas de NLP contribuyen de forma acelerado y concisa de precedentes legales, decisiones judiciales y detalles importantes, lo que genera un marco de diseño en las decisiones.

La resolución de casos judiciales en un tiempo razonable es fundamental para garantizar los aspectos relevantes de las partes involucradas. El uso del PLN puede acelerar el procesamiento de los expedientes judiciales, reduciendo la carga de trabajo de los funcionarios judiciales y mejorando la productividad del sistema judicial.

El objetivo general parte de: Determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Los objetivos específicos relacionados a: OE1: Analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023 y OE2: Analizar la influencia de la extracción de datos en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Para el alcance de dichos objetivos se consideran la hipótesis general que a continuación se detalla: El procesamiento del lenguaje natural influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Siendo las hipótesis específicas planteadas las siguientes: HE1: El preprocesamiento del influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023 y la HE2: La extracción de datos influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

II.MARCO TEÓRICO

En este acápite se abordan consideraciones que conducen a tener como referencia los aspectos orientadores y relevantes de la investigación para argumentar de forma sólida la investigación.

En el análisis internacional Chamorro et al. (2022) donde aporta el objetivo de identificar la viabilidad de una propuesta para identificar el nivel de percepción en la satisfacción de los estudiantes a partir de la ciencia de datos y el procesamiento del lenguaje natural sostenido en la red social twitter y la recolección de datos. Considera el método cuantitativo con un aporte descriptivo y correlacional, que conlleva a demostrar el grado de percepción de la satisfacción de los estudiantes mediante el uso de la ciencia de datos y el PLN. Los resultados han evidenciado que la propuesta identificar un nivel de satisfacción de los estudiantes de 57.27% y cuyo punto de referencia se encuentra relacionado a la satisfacción universitaria con indicadores definidos y aplicados por semestre. Se concluye que es posible obtener un marco de referencia donde usando los enfoques de minería de datos y procesamiento de lenguaje natural, contribuyen de forma significativa a la satisfacción del aprendizaje virtual.

Lee et al. (2023) parte de identificar la importancia del reconocimiento de voz soportado en el PNL para reducir la carga de trabajo clínico. La metodología utilizada es de 21 enfermeros participaron en la evaluación del sistema de RS. El tiempo de documentación y la tasa de error de reconocimiento se evaluaron en paralelo entre los registros generados por SR y la entrada por teclado durante 4 sesiones. Los resultados evidenciaron que el ahorro de tiempo en la carga usando NPL es de 87% y se han procesado 32,456 expedientes, contribuyendo a mejorar en cada sesión. Concluyó que al usar el lenguaje natural se ha logrado integrar los registros de enfermería que ha conllevado a mejorar la productividad con diferentes precisiones en diferentes especialidades.

Leadbetter (2023) tiene como objetivo evaluar la efectividad del PNL para la generación de trabajos de fin de carrera. La metodología parte de un sistema de programación usando Phyton y luego utiliza la API de OpenAI para desplegar modelos de texto para la generación y evaluación de texto, como GPT-3 y Davinci-003. Los resultados han mostrado proporcionar a estudiantes e investigadores una herramienta que les permita elaborar rápidamente de alta calidad sobre una

variedad de temas, sin tener que pasar meses investigando y escribiendo. Concluyó el uso de la potente herramienta de aprovecha la potencia de la PNL y el aprendizaje profundo sin tener que implementar todos los complejos algoritmos y técnicas desde cero.

Delaney et al.(2018) parte del objetivo de vincular la extracción de información usando el PNL con un sustento semántico y la vinculación terminológica en 400 proyectos de desarrollo de software. La metodología parte de un sistema de programación usando Phyton y luego utiliza la API de OpenAI para desplegar modelos de texto para la generación y evaluación de texto. Concluyó que el procesamiento del lenguaje natural conducirá a una metodología novedosa para el desarrollo comparativo de perfiles usando métricas de productividad para el desarrollo de softwares.

Lage (2022) considera el objetivo de precisar las decisiones judiciales brasileñas considerando el procesamiento del lenguaje natural. La metodología utiliza varios algoritmos de constante conocimiento, incluidos clasificadores y modelos de diseño minucioso de última generación. Los resultados muestran que el prototipo funcional de rendimiento usando el lenguaje natural conlleva a un sistema de puntuación para preceder los resultados en las decisiones judiciales brasileñas. Concluyó que la complejidad del sistema legal junto con un enorme volumen de litigios hace que este problema sea aún más difícil, para lo cual se hace uso del PNL

En el ámbito nacional encontramos la investigación de Uscamayta (2023) donde el objetivo es determinar cómo el PNL ayuda a la calidad y celeridad de decisiones judiciales en el proceso civil peruano. La metodología incluye un análisis descriptivo-propositivo y el uso de entrevistas. Según los resultados, el procesamiento del lenguaje natural puede generar calidad y celeridad de decisiones judiciales en los procesos civiles en Perú. Concluyó que se llevó a cabo un diagnóstico para identificar la toma de decisiones jurisdiccionales y poner énfasis en admitir demandas o casos particulares.

Garces (2020) busca analizar las técnicas de PNL y minería de datos para clasificar incidentes informáticos. La metodología es un caso de estudio que analiza incidentes informáticos operativos en un centro educativo con base en un área de mesa de servicios que soporta las actividades de los empleados. Los resultados

indican que el requerimiento se responde con mayor rapidez y ahorro de tiempo, lo que resulta en una métrica de precisión y precisión del 85 %. Concluyó que han ayudado a predecir variables en el análisis considerado.

Mientras Quispe (2021) que parte de analizar la forma en que los sistemas inteligentes desarrollados en el campo del Derechos tanto interno y externo. La metodología es descriptiva y con un análisis documental, que conlleva a la reflexión de que en el sistema de justicia los jueces sean reemplazados por los robots, que implican abordar instancias superiores producto del avance tecnológico. Los hallazgos sugieren que el progreso tecnológico ha dado lugar a soluciones para la ciencia del derecho que permiten crear modelos para argumentar y tomar decisiones, clasificar textos legales, extraer información y crear y planificar un sistema legislativo. Concluyó que la inteligencia artificial ha ayudado a crear un proceso del sistema inteligente PROMETEA que puede generar dictámenes jurídicos, ejecuciones fiscales y analizar la documentación en tiempos de reducción de carga procesal.

Para la variable procesamiento del lenguaje natural Peña et al. (2022) considera la variable procesamiento del lenguaje natural como una de las herramientas del aprendizaje automático que implica el uso de máquinas que aprenden a leer, descifrar e interpretar el lenguaje humano, así como la creación de soluciones de software que respondan a los aspectos sintácticos, gramaticales y semánticos. Esto crea una conexión con los modelos relevantes de lingüística aplicada, computacional y lenguaje humano y automatización.

Los estudios sobre los modelos de extracción de relaciones (ER) y los aspectos de forma rápida se derivan de la base teórica del PNL.

De ahí que la ER es considerada una extracción de información (EI) que conduce a reconocer las relaciones semánticas en las entidades (Pawar et al.,2017).

Téllez (2005) sustenta el aspecto teórico computacional en construir sistemas computacionales para lograr mejoras automáticas mediante la experiencia, al desarrollar un programa o sistema capaz de proporcionar una experiencia para ser el soporte del algoritmo de aprendizaje.

La necesidad de una clasificación automática del texto conlleva a clasificar de forma automática un conjunto de documentos dentro de un conjunto

determinado para la clasificación de texto que combina el aspecto tecnológico con la recuperación de información y aprendizaje automático (Peña et al.,2022).

El enfoque del proceso inductivo conduce a construir un clasificador por observar las características de un conjunto de documentos clasificados en línea. De esta forma el método inductivo es considerado un proceso indexado de clasificación para presentar los textos que implica un conjunto de entrenamiento para desarrollar catálogos para la presentación de nuevos documentos.

Dhamija y Bag (2020) consideran una cascada de módulos como la arquitectura general de los sistemas de extracción de información.

Desde la concepción de los sistemas para las perspectivas de procesos se requiere un aspecto de procesos en base a características relevantes (Sarabia, 1995).

En el aspecto diferenciado de procesos requiere de un modelo enmarcado en una caja negra, que implica una estructura interna con procesos de entorno y efectos en los procesos que implica modelizar un sistema que parte de un aspecto activo y el entorno para llevar a cabo el procesamiento (Sarabia, 1995).

Para la dimensión preprocesamiento es una etapa crítica en el flujo de trabajo de NLP que involucra la preparación y limpieza de los datos de texto antes de que se puedan aplicar técnicas de análisis o modelado de lenguaje natural. El preprocesamiento es esencial para garantizar que los datos de texto estén en un formato adecuado y sean susceptibles de ser analizados de manera efectiva (Peña et al.,2022).

De la dimensión preprocesamiento se considerada a los indicadores de limpieza de texto, tokenización y eliminación de palabras vacías.

En la limpieza de texto se explica cómo manipular datos de texto con expresiones regulares. Cuando se trabaja con contenido web sin formato que contiene etiquetas HTML, texto largo y texto repetido, se requieren expresiones regulares. Durante el desarrollo y la salida de la aplicación, estos datos no son necesarios. Para limpiar una variedad de datos básicos y adicionales, puede usar expresiones regulares (Kulkarni y Shivananda, 2021).

La tokenización consiste en dividir el texto en partes pequeñas y significativas. Hay marcadores de frase y de palabra. Verá un token de una sola palabra en esta sección. Este es un paso previo al preprocesamiento de texto

necesario para todos los tipos de análisis. Muchas bibliotecas, como NLTK, spaCy y TextBlock, pueden realizar tokenización. Las siguientes son algunas formas de hacerlo (Kulkarni y Shivananda, 2021).

La eliminación de palabras vacías proporciona instrucciones sobre cómo eliminar palabras vacías. Las palabras vacías son palabras simples y comunes que no tienen más significado que otras palabras clave. Puede concentrarse en las palabras clave importantes eliminando las palabras que se utilizan con menos frecuencia (Kulkarni y Shivananda, 2021).

Los motores de búsqueda pueden concentrarse en recuperar páginas con palabras clave como desarrollar, chatbot y python, lo que resulta en más páginas de interés real. Otras palabras comunes y raras también pueden eliminarse (Kulkarni y Shivananda, 2021).

Mientras la dimensión extracción de datos se enfoca en la obtención de información relevante y estructurada a partir de los textos procesados. En esta etapa, se busca identificar patrones, características o elementos de interés en los datos textuales para su posterior análisis o aplicación en diversas tareas de NLP (Miranda y Guzman, 2014).

De esta forma la extracción de información se ocupa de estructurar los contenidos dentro de los textos relevantes del estudio, que conlleva a sistematizarlo y extraer la información requerida para ser asociado a identificar aquellos patrones donde se localiza el enfoque de construcción de forma automática de los planes.

La dimensión extracción de datos presenta dos indicadores relacionados a el etiquetado de texto y extracción de palabras claves.

El etiquetado del texto parte de la identificación de elementos de archivo que se detalla, buscando palabras clave o temas específicos que describan qué capítulo debería pertenecer a este libro y luego lo ajustas de acuerdo. El modelado de temas es muy útil en este tipo de situaciones. Esto se conoce como etiquetar y organizar documentos (Miranda y Guzman, 2014).

La extracción de palabras clave es una tarea esencial en el procesamiento del lenguaje natural (PLN) que implica identificar las palabras o frases más relevantes en un texto. Estas palabras clave representan la esencia del contenido y son útiles en diversas aplicaciones, como resúmenes automáticos, indexación de documentos y recuperación de información. (Miranda y Guzman, 2014).

Para la variable de productividad Porter (2017) sobre la productividad se basa en el concepto de ventaja competitiva y la importancia de la eficiencia en la empresa para lograr un desempeño superior en el mercado. La productividad se vincula estrechamente con la eficiencia en la cadena de valor. Porter enfatiza la importancia de reducir los costos y optimizar los procesos en todas las actividades de la cadena de valor. Una mayor eficiencia en la producción, la distribución, el marketing y otras áreas puede generar una ventaja competitiva, ya que permite a la empresa ofrecer productos o servicios de alta calidad a un costo menor.

Porter sugiere que la productividad no solo se trata de reducir costos, sino también de crear valor. La diferenciación de productos o servicios puede aumentar la productividad al permitir a una empresa cobrar precios más altos. Esto puede lograrse a través de la innovación, la calidad superior, la marca y otros factores que hacen que los productos o servicios sean únicos y valiosos para los clientes.

Por otro lado, la teoría de Porter (2017) considera la productividad como un factor crítico para el desarrollo en el mercado. La eficiencia en la cadena de valor, la diferenciación y la alineación estratégica son elementos clave en esta perspectiva. Al mejorar la productividad en sus operaciones, las empresas pueden fortalecer su posición en el mercado y lograr un mejor desempeño financiero.

Medina (2020) considera que, en la aproximación teórica de la productividad laboral, lo referente a la cultura organizacional se desempeña a partir del rendimiento y realizando esfuerzos en torno a mejorar los recursos humanos que conduzcan a herramientas desde ámbito organizacional para incorporar estrategias que sean más beneficiosas para las organizaciones, construyendo desde la teoría moderna de la gestión que existe un aporte exhaustivo que conlleve a mejorar las ventas y utilidades.

Herrera y Madril (2022) considera que la productividad es considerada una variable de desempeño de las empresas que termina en eficiencia, competitividad o rentabilidad explicada por factores internos y externos que conduce a medir la eficiencia de las economías. Siendo la medición la que se visualiza como aquella relación producto de la división de las horas estándar trabajadas con las horas de reloj en el trabajo.

Siendo la eficiencia del trabajador lo que refleja el periodo que demora el trabajador en obtener la producción y la eficacia es considerada el resultado del

proceso de organización, planificación y proyección para alcanzar los objetivos (Herrera y Madril, 2022).

La dimensión de tiempo de registro de productividad se refiere a la recopilación y el seguimiento del tiempo dedicado a tareas específicas con el fin de evaluar y mejorar la productividad en un entorno laboral o personal. Esta dimensión es crucial para comprender cómo se utiliza el tiempo, identificar áreas de mejora y optimizar la eficiencia en la realización de tareas y proyectos (Herrera y Madril, 2022)

III.METODOLOGÍA

3.1. Tipo y diseño de investigación

3.1.1. Tipo de investigación

El estudio descrito parte de un enfoque cuantitativo, de tipo aplicada. Según Ñaupas, Palacios, Romero & Valdivia (2018), consiste en aplicar los conocimientos adquiridos a partir de la investigación aplicada en brindar aporte a los problemas reales. El propósito de la tesis es contribuir al desarrollo y bienestar social en diversos ámbitos. Conforme a Castro, Gómez & Camargo (2023), precisan que se trata de investigaciones originales que tienen como objetivo adquirir nuevos conocimientos enfocados en resolver un problema específico.

3.1.2. Diseño de investigación

Debido a que las variables no se han modificado o manipulado, la investigación se considera de diseño no experimental y, por lo tanto, se verificó tal como se inició en un contexto de investigación específico. Debido a que los análisis se realizaron en un momento específico, es de alcance correlacional (Hernández et al., 2014).

En la investigación correlacional, no distingue entre variables dependientes e independientes; en cambio, busca identificar las relaciones entre las variables existentes. En cambio, se analizan las variables de estudio y cómo se relacionan entre sí.

El esquema de análisis es el siguiente:

Figura 1

Esquema



Nota. Obtenido de Sánchez y Reyes (1984)

Dónde:

Y1: Variables del procesamiento del Lenguaje Natural

Y2: Productividad

X: Muestra

3.2. Variables y operacionalización

Variable 1: Procesamiento del Lenguaje Natural, que es una variable cuantitativa

- **Definición conceptual**

El procesamiento del lenguaje natural (NLP) es considerado un análisis producto de la IA para generar algoritmos, técnicas y modelos computacionales que permiten a las computadoras comprender, analizar y generar efectivamente lenguaje humano (Baca et al.,2023).

- **Definición operacional**

Se utilizará el instrumento de la ficha de observación. Con una población muestral de 378 expedientes judiciales, donde se consideró las dimensiones de preprocesamiento y extracción de datos.

- **Dimensión**

Preprocesamiento

- **Indicadores**

- Limpieza de texto
- Tokenización
- Eliminación de palabras vacías

- **Dimensión**

Extracción de datos

- **Indicadores**

- Etiquetado de texto
- Extracción de palabras claves

- **Escala de medición:** la variable del presente estudio es medida por la escala razón.

Variable 2: Productividad, que es una variable cuantitativa

- **Definición conceptual**

La productividad es un concepto fundamental en la economía y la gestión empresarial que se aborda por el uso de recursos utilizados y la cantidad de producción o resultados obtenidos (Gutierrez, 2010).

- **Definición operacional**

Para medir la variable se utilizará el instrumento de la ficha de observación. Con una población muestral de 378 expedientes judiciales, donde se consideró la dimensión de tiempo de registro.

- **Dimensión**

Tiempo de registro

- **Indicadores**

- Datos de salida
- Productividad General

3.3. Población, muestra y muestreo

3.3.1.Población

Es considerado un conglomerado de individuos u otro conjunto que comparta una característica o propiedad común y que es el objeto de estudio en un análisis particular (Hernández et. al., 2014). La población se encuentra constituida por 21,521 expedientes judiciales de Lima Norte del periodo 2023 entre enero y setiembre.

Criterios de inclusión: Se considera el criterio de especialización donde la referencia por áreas hace referencia a los Juzgados Especializados, donde a partir de la Sub Especialidad de Violencia contra la mujer y los Integrantes del grupo familiar, donde el total de expedientes asciende a 21,521 expedientes del periodo de Enero-Setiembre 2023.

Criterios de exclusión: Para la investigación se excluye a los otros juzgados que no intervienen en el estudio de investigación.

3.3.2.Muestra

La muestra equivale a 378 expedientes de los Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los Integrantes del grupo familiar de Lima Norte del periodo de Enero-Setiembre 2023.

3.3.3.Muestreo

El muestreo que se realiza es probabilístico simple y asciende a 378 expedientes judiciales del área antes descrita, que se muestra con la siguiente fórmula.

$$n = \frac{Z_{\alpha}^2 Npq}{e^2(N - 1) + Z_{\alpha}^2 pq}$$

$$n = \frac{(1.95) * (1.95) * (21,521) * (0.5) * (0.5)}{(0.05) * (0.05) * (21,521 - 1) + (1.95) * (1.95) * (0.5) * (0.5)} = 378$$

3.3.4 Unidad de análisis

Datos pertenecientes a los expedientes de los Juzgados Especializados de la Sub Especialidad de la Violencia contra la mujer y los integrantes del grupo familiar.

3.4. Técnicas e instrumentos de recolección de datos

Se considera la técnica de análisis documental, usando como instrumento la ficha de observación, cuya parte del procedimiento encaja en juntar datos de manera sistemática y estructurada durante una observación o estudio.

Confiabilidad:

El alfa de Cronbach realizado asciende a 0.838. En este caso, un valor de 0,838 sugiere una buena consistencia interna, lo que implica que las preguntas están correlacionadas de manera positiva entre sí, y además que el instrumento usado tiene una fuerte confiabilidad, es decir que el instrumento de recopilación de datos ha permitido realzar mediciones estables y consistentes.

Tabla 1

Alfa de Cronbach

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
,838	6

Nota: Elaboración propia del investigador

3.5. Procedimientos

El método utilizado es recopilar datos del sistema oficial, lo que implica en un primer paso buscar datos en sus sistemas para construir las variables. En segundo paso bajar la información y luego proceder a realizar su análisis y

tratamiento que permita constituir un logro en la evidencia científica para generar datos de evidencia significativa.

Luego se hace uso del SPSS V.26 para la estimación de la relación con el Rho de Spearman.

3.6. Método de análisis de datos

Los métodos cuantitativos son un enfoque de investigación y análisis que se considera en el análisis inferencial numérico y estadísticos con el fin de comprender y explicar fenómenos, resolver problemas y tomar decisiones (Vilalta, 2016) Gamra (2013) afirma que Microsoft Office Excel se usa para el diseño de tablas Y gráficos.

Posteriormente se han transferido la recopilación de datos de Excel a SPSS para su procesamiento tanto descriptivo e inferencial, logrando generar tablas, gráficos, y el análisis correlacional para manejar grandes cantidades de datos u otros formatos.

El coeficiente de Rho de Spearman también facilita la identificación de la conexión presente entre las variables en estudio, lo que permite obtener pruebas científicas.

3.7. Aspectos éticos

El suscrito mantuvo en todo momento los principios éticos de investigación y su vez respetando las normativas vigentes de la casa de estudios. En relación al presente documento, específicamente en la redacción se tomó como protocolo seguir las pautas de la guía de elaboración de proyecto de tesis, además se utilizó la Guía APA en su séptima edición para realizar la citación y referencias bibliográficas; también se utilizó la herramienta Turnitin en sus dos facetas Draft Coach y en la plataforma Clementina para validar que el contenido detallado es original del suscrito. Por último, hacemos mención a los documentos vigentes de la casa de estudio: RVI N°062-2023-VI-UCV, RCU N°0200-2018/UCV y RCU N°0470-2022/UCV.

IV.RESULTADOS

El aporte descriptivo e inferencial y la técnica del coeficiente de Rho de Spearman para el análisis de correlación al ser consideradas variables cuantitativas de razón.

Por otro lado, para identificar la normalidad de los datos se considera la prueba de normalidad del Kolmogorov-Smirnov considerando que son 378 datos para mostrar la normalidad de los datos.

Tabla 2

Resultados de prueba de normalidad

	Pruebas de normalidad					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Procesamiento del lenguaje natural						
Limpieza de texto	,360	378	,000	,635	378	,000
Tokenización	,360	378	,000	,635	378	,000
Eliminación de palabras vacías	,360	378	,000	,635	378	,000
Extracción de datos						
Etiquetado de texto	,364	378	,000	,634	378	,000
Extracción de palabras claves	,364	378	,000	,634	378	,000
Productividad						
Datos de salida	,282	378	,000	,797	378	,000

a. Corrección de significación de Lilliefors

Nota: Elaboración propia del investigador

Interpretación

Los resultados de pruebas de Normalidad, se evidencia que los indicadores de las variables independientes procesamiento del lenguaje natural, además de los indicadores de la variable dependiente productividad tienen un valor p-valor = 0; y contrastándolo con los criterios de decisión de la Tabla 2, el p-valor (Sig<0.05) es

menor a 0.05; se rechaza la hipótesis nula, concluyendo que no hay una distribución normal de los datos.

4.1. Descripción de los resultados a modo descriptivo

Las estadísticas descriptivas se detallan a continuación.

Tabla 3

Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural-dimensión preprocesamiento

		Preprocesamiento
N	Válido	378
	Perdidos	0
Media		8,61
Mediana		10,00
Moda		10
Desv. Desviación		1,498
Varianza		2,244
Asimetría		-,149
Error estándar de asimetría		,125
Curtosis		-1,988
Error estándar de curtosis		,250
Rango		3
Mínimo		7
Máximo		10
Percentiles	25	7,00
	50	10,00
	75	10,00

Nota: Elaboración propia del investigador

Interpretación

La interpretación de los resultados de la tabla 3 son:

- **Media (Promedio):** La media es 8.61, lo que indica el valor típico de la variable. En este caso, representa un nivel moderado de la variable.
- **Error Estándar:** El error estándar es 0.125. Este valor indica la variabilidad esperada en la media de las muestras. Cuanto menor sea el error estándar, más precisa será la estimación de la media.

- Mediana: La mediana es 10, que es similar a la media. Indica que la distribución de los datos no está sesgada de manera significativa.
- Varianza y Desviación Estándar: La varianza es 2.24 y la desviación estándar es 1.49.
- Mínimo y Máximo: El valor mínimo es 7.00 y el máximo es 10.00. Esto proporciona información sobre el rango total de la variable, que es 3.00.
- Rango: El rango es la diferencia entre el máximo y el mínimo ($10.00 - 7.00 = 3.00$), mostrando la extensión total de los datos.
- Rango intercuartil (IQR): Es la resta del tercer cuartil (Q3) y el primer cuartil (Q1). En este caso, es 3.00, lo que indica la dispersión central de los datos, excluyendo los valores extremos.
- Asimetría: La asimetría es -0.149, lo que sugiere una ligera asimetría negativa. Los datos están ligeramente sesgados hacia la izquierda, pero no de manera significativa.
- Curtosis: La curtosis es -1.988, lo que indica una ligera curva en la distribución. La curtosis negativa sugiere colas más ligeras y una distribución más aplanada en comparación con una distribución normal.

Estos estadísticos han proporcionado una descripción detallada de la distribución y la tendencia central de la dimensión preprocesamiento de la variable independiente procesamiento del lenguaje natural. Con ello, se ha podido determinar la dispersión de los datos.

Tabla 4

Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural-dimensión extracción de datos

		Extracción de datos
N	Válido	378
	Perdidos	0
Media		4,54
Mediana		5,00
Moda		5
Desv. Desviación		,499
Varianza		,249
Asimetría		-,181
Error estándar de asimetría		,125

Curtosis		-1,978
Error estándar de curtosis		,250
Rango		1
Mínimo		4
Máximo		5
Percentiles	25	4,00
	50	5,00
	75	5,00

Nota: Elaboración propia del investigador

Interpretación

La interpretación de los resultados de la tabla 4 son:

- **Media (Promedio):** La media es 4.54, lo que indica el valor típico de la variable. En este caso, representa un nivel moderado de la variable.
- **Error Estándar:** El error estándar es 0.125. Este valor indica la variabilidad esperada en la media de las muestras.
- **Mediana:** La mediana es 5, que es similar a la media. Indica que la distribución de los datos no está sesgada de manera significativa.
- **Varianza y Desviación Estándar:** La varianza es 0.249 y la desviación estándar es 0.499.
- **Mínimo y Máximo:** El valor mínimo es 4.00 y el máximo es 5.00. Esto proporciona información sobre el rango total de la variable, que es 1.00.
- **Rango:** Es la resta entre el máximo y el mínimo ($5.00 - 4.00 = 3.00$), mostrando la extensión total de los datos.
- **Rango intercuartil (IQR):** Es la sustracción del tercer cuartil (Q3) y el primer cuartil (Q1). En este caso, es 1.00, lo que indica la dispersión central de los datos, excluyendo los valores extremos.
- **Asimetría:** La asimetría es -0.181, lo que sugiere una ligera asimetría negativa. Los datos están ligeramente sesgados hacia la izquierda, pero no de manera significativa.
- **Curtosis:** La curtosis es -1.978, lo que indica una ligera curva en la distribución. La curtosis negativa sugiere colas más ligeras y una distribución más aplanada en comparación con una distribución normal.

Estos estadísticos han proporcionado una descripción detallada de la distribución y la tendencia central de la dimensión preprocesamiento de la variable independiente procesamiento del lenguaje natural

Tabla 5

Cuadro de estadística descriptiva de variable procesamiento del lenguaje natural- dimensión tiempo de registro

N	Tiempo de registro	
	Válido	Perdidos
	378	0
Media		9,08
Mediana		9,00
Moda		9
Desv. Desviación		,668
Varianza		,447
Asimetría		-,095
Error estándar de asimetría		,125
Curtosis		-,761
Error estándar de curtosis		,250
Rango		2
Mínimo		8
Máximo		10
Percentiles	25	9,00
	50	9,00
	75	10,00

Nota: Elaboración propia del investigador

Interpretación

La interpretación de la tabla 5, son:

- **Media (Promedio):** La media es 9.08, lo que indica el valor típico de la variable. En este caso, representa un nivel moderado de la variable.
- **Error Estándar:** El error estándar es 0.125. Este valor indica la variabilidad esperada en la media de las muestras.
- **Mediana:** La mediana es 9, que es similar a la media. Indica que la distribución de los datos no está sesgada de manera significativa.
- **Varianza y Desviación Estándar:** La varianza es 0.447 y la desviación estándar es 0.668.

- Mínimo y Máximo: El valor mínimo es 8.00 y el máximo es 10.00. Esto proporciona información sobre el rango total de la variable, que es 2.00.
- Rango: Es la resta entre el máximo y el mínimo ($10.00 - 8.00 = 2.00$), mostrando la extensión total de los datos.
- Rango intercuartil (IQR): Es la resta del tercer cuartil (Q3) y el primer cuartil (Q1). En este caso, es 2.00, lo que indica la dispersión central de los datos, excluyendo los valores extremos.
- Asimetría: La asimetría es -0.095, lo que sugiere una ligera asimetría negativa. Los datos están ligeramente sesgados hacia la izquierda, pero no de manera significativa.
- Curtosis: La curtosis es -0.761, lo que indica una ligera curva en la distribución. La curtosis negativa sugiere colas más ligeras y una distribución más aplanada en comparación con una distribución normal.

Estos estadísticos han proporcionado una descripción detallada de la distribución y la tendencia central de la dimensión preprocesamiento de la variable independiente procesamiento del lenguaje natural.

4.2. Descripción de los resultados inferenciales

Formulación de las hipótesis específicas

La formulación de la hipótesis específica 1 de investigación es la siguiente:

Ho: No existe influencia significativa del preprocesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Ha: Existe influencia significativa del preprocesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Tabla 6

Prueba de correlación de la dimensión preprocesamiento de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.

Correlaciones	Preprocesamiento
	o

Rho de Spearman	Productividad	Coeficiente de correlación	,894**
		Sig. (bilateral)	,000
		N	378

** . La correlación es significativa en el nivel 0,01 (bilateral).

Nota: Elaboración propia del investigador

Interpretación

En la tabla 6 la interpretación es:

- Coeficiente de Rho de Spearman entre preprocesamiento y la productividad es de 0.894.
- Significancia Estadística: La correlación es significativa a un nivel de significancia del 0.01 (bilateral).
- El p-valor es menor a 0.05, y conforme la Tabla 6, se rechaza la hipótesis nula, de que existe influencia significativa del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023
- La correlación de 0.894 sugiere una fuerte relación positiva entre el preprocesamiento y la productividad en la muestra analizada. La significancia estadística respalda que las variables están relacionadas. Por lo tanto, existe una asociación significativa y positiva entre el preprocesamiento y la productividad en la muestra analizada.

Para realizar las pruebas de correlaciones de la dimensión extracción de datos de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad, de la hipótesis específica 2, se formuló las hipótesis respectivas en la tabla siguiente:

Ho: No existe influencia significativa de la extracción de datos en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Ha: Existe influencia significativa de la extracción de datos en la productividad de los expedientes judiciales en un distrito judicial Lima Norte

Tabla 7

Prueba de correlación de la dimensión extracción de datos de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.

Correlaciones		Extracción de datos
Rho de Spearman	Productividad	Coefficiente de correlación ,353**
		Sig. (bilateral) ,000
		N 378

** . La correlación es significativa en el nivel 0,01 (bilateral).

Nota: Elaboración propia del investigador

Interpretación

La interpretación de la Tabla 7 es:

- Coeficiente de Rho de Spearman entre extracción de datos y la productividad es de 0.353.
- Significancia Estadística: La correlación es significativa a un nivel de significancia del 0.01 (bilateral).
- El p-valor es menor a 0.05, y conforme la Tabla 7, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, de que existe influencia significativa del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023
- La correlación de 0.353 sugiere una fuerte relación positiva entre la extracción de datos y la productividad en la muestra analizada. La significancia estadística respalda que las variables están relacionadas. Por lo tanto, existe una asociación significativa y positiva entre la extracción de datos y la productividad en la muestra analizada.

Formulación de la hipótesis general

Finalmente, para realizar las pruebas de correlaciones de la variable independiente procesamiento del lenguaje natural y la variable dependiente

productividad, de la hipótesis general, se formuló las hipótesis respectivas en la tabla siguiente.

Tabla 8

Prueba de correlación de la variable independiente procesamiento del lenguaje natural y la variable dependiente productividad.

Correlaciones		Procesamiento del lenguaje natural
Rho de Spearman	Productividad	Coeficiente de correlación
		,962**
		Sig. (bilateral)
		,000
		N
		378

** . La correlación es significativa en el nivel 0,01 (bilateral).

Nota: Elaboración propia del investigador

Interpretación

La interpretación de los resultados de la Tabla 8 es:

- Coeficiente Rho de Spearman entre procesamiento del lenguaje natural y la productividad es de 0.962.
- Significancia Estadística: La asociación es significativa a un nivel de significancia del 0.01 (bilateral).
- El p-valor es menor a 0.05, y conforme la Tabla 8, se rechaza la hipótesis nula, de que existe influencia significativa del preprocesamiento del lenguaje natural y la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023
- La asociación de 0.962 sugiere una fuerte relación positiva entre el procesamiento del lenguaje natural y la productividad en la muestra analizada. La significancia estadística respalda que las variables están relacionadas. Por lo tanto, existe una asociación significativa y positiva entre el procesamiento del lenguaje natural y la productividad en la muestra analizada.

V.DISCUSIÓN

El procesamiento del lenguaje natural considera un aspecto crucial en el desempeño de expedientes judiciales, ya que permite la extracción de información valiosa y la automatización de tareas relacionadas con el análisis de documentos legales, siendo expresado el problema que los expedientes judiciales enfrenta desafíos significativos en términos de tiempo debido al volumen de documentos, la complejidad del contenido legal, la diversidad de idiomas, la necesidad de mantener la privacidad y seguridad, y la búsqueda eficiente de información relevante. Superar estos desafíos requiere una planificación adecuada y la aplicación de tecnologías avanzadas de PLN para dinamizar el tiempo y los recursos dedicados a la gestión y análisis de expedientes legales (Barsha y Munshi, 2023).

El objetivo general usando la prueba estadística Rho de Spearman, prueba no paramétrica se evidenció que existe una influencia positiva y significativa del procesamiento del lenguaje natural y la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023, aceptando la hipótesis alterna planteada y rechazando la hipótesis nula. Asimismo, el coeficiente de Rho de Spearman arrojó un coeficiente de 0.956, evidenciando dicha asociación positiva y fuerte entre las variables analizadas. Dichos resultados coinciden con Leadbetter (2023) donde han señalado los resultados proporcionar a los estudiantes e investigadores una herramienta que les permita elaborar rápidamente de alta calidad sobre una variedad de temas, sin tener que pasar meses investigando y escribiendo. Así mismo con Uscamayta (2023) cuyos resultados evidencian que el procesamiento del lenguaje natural puede optimizar el tiempo en los procesos judiciales.

Garces (2020) busca analizar las técnicas de PNL y minería de datos para clasificar incidentes informáticos cuyos resultados indican que el requerimiento se responde con mayor rapidez y ahorro de tiempo, lo que resulta en una métrica de precisión y precisión del 85 %.

Evidenciando de esta forma que la variable procesamiento del lenguaje natural es una de las herramientas del aprendizaje automático que implica el uso de máquinas que aprenden a leer, descifrar e interpretar el lenguaje humano, así como la creación de soluciones de software que respondan a los aspectos sintácticos, gramaticales y semánticos. Esto crea una conexión con los modelos

relevantes de lingüística aplicada, computacional y lenguaje humano y automatización (Appadoo et al.,2020; Peña et al.,2022).

El PLN permite que las máquinas comprendan y respondan al lenguaje humano de manera natural. Esto es esencial para desarrollar sistemas de inteligencia artificial (IA) que puedan interactuar con los usuarios de una manera que se asemeje a la comunicación entre humanos (Sun et al.,2023).

De esta forma las computadoras analizan grandes cantidades de texto de manera eficiente. Esto es fundamental para tareas como la extracción de información, la clasificación de documentos, los resúmenes automáticos y el análisis de sentimientos, entre otras aplicaciones (Chai, 2023; Yan et al., 2020).

El PLN es esencial en motores de búsqueda avanzados, donde las máquinas deben entender la intención del usuario expresada en lenguaje natural para devolver resultados relevantes y precisos (Boone et al.,2023).

Los expedientes judiciales suelen contener grandes cantidades de información en forma de documentos legales, testimonios, transcripciones y otros registros. El PLN permite analizar rápidamente estos documentos para extraer información relevante, identificar patrones y resumir el contenido, facilitando la revisión por parte de abogados y jueces (Kolthoff et al.,2023).

El PLN facilita la búsqueda eficiente de información específica en expedientes judiciales. Puede ayudar a los profesionales del derecho a encontrar rápidamente precedentes, referencias legales, antecedentes de casos y otra información relevante, ahorrando tiempo y recursos (Rehman et al.,2023).

De esta forma se puede clasificar automáticamente documentos legales en categorías específicas, lo que simplifica la organización y gestión de los expedientes judiciales. Por ejemplo, puede ayudar a identificar la naturaleza de un caso, el tipo de documentos asociados y la jurisprudencia aplicable (D’Orazio et al.,2023).

El PLN puede extraer información clave de documentos complejos, como nombres de partes involucradas, fechas importantes, eventos relevantes y términos legales específicos. Esto facilita la comprensión rápida y precisa de los detalles esenciales de un caso (Goldrick et al.,2023; Diwan, 2023).

El PLN puede utilizarse para analizar el tono emocional y las opiniones expresadas en documentos legales, como argumentos legales, testimonios y

declaraciones. Esto puede ser útil para evaluar la persuasión emocional en un caso o identificar posibles sesgos (Kobayashi et al.,2022; Zhou y Cen, 2023).

El primer objetivo específico referente a analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023, aceptando la hipótesis alterna planteada y rechazando la hipótesis nula, dado que existe una asociación significativa y positiva entre el preprocesamiento y la productividad en la muestra analizada. Los resultados se asocian con Chamorro et al. (2022) el cual evidencia que la propuesta de identificar un nivel de satisfacción de los estudiantes de 57.27% y cuyo punto de referencia se encuentra relacionado a la satisfacción universitaria con indicadores definidos y aplicados por semestre.

Por otro lado, se relacionan con Leadbetter (2023) cuyos resultados han mostrado proporcionar a estudiantes e investigadores una herramienta que les permita elaborar rápidamente de alta calidad sobre una variedad de temas, sin tener que pasar meses investigando y escribiendo

Delaney et al.(2018) parte del objetivo de vincular la extracción de información usando el PLN para el análisis semántico y la vinculación terminológica en 400 proyectos de desarrollo de software. Concluyó que el procesamiento del lenguaje natural conducirá a una metodología novedosa para el desarrollo comparativo de perfiles usando métricas de productividad para el desarrollo de softwares.

Lage (2022) los resultados muestran que el prototipo funcional de rendimiento usando el lenguaje natural conlleva a un sistema de puntuación para preceder los resultados en las decisiones judiciales brasileñas. Concluyó que la complejidad del sistema legal junto con un enorme volumen de litigios hace que este problema sea aún más difícil, para lo cual se hace uso del procesamiento del lenguaje natural.

De esta manera el preprocesamiento de expedientes judiciales es una fase crucial en el sistema judicial que puede influir significativamente en la productividad y eficiencia del proceso legal.

El preprocesamiento implica la organización y clasificación de la gran cantidad de información presente en los expedientes judiciales. Esto facilita el

acceso rápido a la información relevante durante las fases posteriores del proceso legal, permitiendo al personal jurisdiccional encontrar y optimizar recursos.

Los expedientes judiciales a menudo contienen información redundante o duplicada. El preprocesamiento implica la identificación y eliminación de dicha redundancia, lo que no solo ahorra espacio, sino que también facilita la revisión y la toma de decisiones al evitar confusiones causadas por información repetida.

El preprocesamiento incluye la indexación adecuada de los expedientes judiciales. Una buena indexación facilita la búsqueda y recuperación de documentos específicos, lo que ahorra tiempo y esfuerzo al personal jurisdiccional y contribuye a una mayor productividad.

El preprocesamiento también puede incluir la revisión de documentos para garantizar el cumplimiento de normativas y procedimientos legales. Esto es crucial para evitar posibles problemas legales derivados de errores o incumplimientos en la gestión de expedientes judiciales.

Los expedientes judiciales suelen contener una gran cantidad de datos no estructurados y ruidosos. El preprocesamiento implica la limpieza de estos datos, eliminando errores tipográficos, formatos inconsistentes y otra información redundante. Esto mejora la calidad de los datos y evita problemas derivados de información incorrecta o mal formateada.

El preprocesamiento incluye la normalización del texto, que implica la estandarización de mayúsculas y minúsculas, la eliminación de caracteres especiales y la reducción de palabras a sus formas base (lematización). Esto facilita la comparación y búsqueda de información, mejorando la consistencia en el análisis de expedientes.

El preprocesamiento puede incluir estrategias para manejar sinónimos y variaciones lingüísticas, como la reducción de palabras a sus raíces (stemming). Esto ayuda a agrupar términos relacionados y garantiza que la búsqueda y el análisis sean más comprensivos.

En expedientes judiciales extensos, la reducción de dimensionalidad durante el preprocesamiento puede ser útil para manejar grandes conjuntos de datos. Esto mejora la eficiencia en el almacenamiento y procesamiento de información, facilitando un análisis más rápido.

De esta manera el preprocesamiento de expedientes judiciales es fundamental para garantizar la calidad de los datos, facilitar análisis eficientes y mejorar la productividad en la gestión de la información legal. Un preprocesamiento cuidadoso contribuye a la precisión y eficiencia en la toma de decisiones judiciales y legales.

El aporte de Peña et al. (2022) considera que la variable procesamiento del lenguaje natural como una de las herramientas del aprendizaje automático que implica el uso de máquinas que aprenden a leer, descifrar e interpretar el lenguaje humano, así como la creación de soluciones de software que respondan a los aspectos sintácticos, gramaticales y semánticos. Esto crea una conexión con los modelos relevantes de lingüística aplicada, computacional y lenguaje humano y automatización.

Los estudios sobre los modelos de extracción de relaciones (ER) y los servicios web utilizados en la traducción automática se derivan de la base teórica del PNL.

De ahí que la ER es considerada una extracción de información (EI) que conduce a reconocer las relaciones semánticas en las entidades descritas (Pawar et al.,2017).

El segundo objetivo específico que busca la influencia de la extracción de datos en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023, cuyo coeficiente de Rho de Spearman entre extracción de datos y la productividad es de 0.353, siendo la fuerza y dirección de la relación entre las dos variables. De esta forma se rechaza la hipótesis nula, de que existe influencia significativa del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023. Dichos resultados se relacionan con Uscamayta (2023) donde el objetivo es determinar cómo el procesamiento del lenguaje natural ayuda a identificar la optimización en la toma de decisiones judiciales en el proceso civil peruano. La metodología incluye un análisis descriptivo-propositivo y el uso de entrevistas. Según los resultados, el procesamiento del lenguaje natural puede involucrar calidad y decisiones en los procesos civiles en Perú. Concluyó que se llevó a cabo un diagnóstico para mejorar la calidad y la rapidez de la toma de decisiones jurisdiccionales y poner énfasis en admitir demandas o casos particulares. Así mismo con Garces (2020) que busca

analizar las técnicas de PNL y minería de datos para ordenar los aspectos informáticos. La metodología es un caso de estudio que analiza con base en un área de mesa de servicios que soporta las actividades de los empleados. Los resultados indican que el requerimiento se responde con mayor rapidez y ahorro de tiempo, lo que resulta en una métrica de precisión y precisión del 85 %. Concluyó que la preparación de datos, la selección de métodos y la exploración de datos han ayudado a predecir variables en el análisis de incidentes de una institución educativa.

Siendo el aporte teórico de la dimensión extracción de datos que se enfoca en la obtención de información relevante y estructurada a partir de los textos procesados. En esta etapa, se busca identificar patrones, características o elementos de interés en los datos textuales para su posterior análisis o aplicación en diversas tareas de NLP (Miranda y Guzman, 2014). De esta forma la extracción de información se ocupa de estructurar los contenidos dentro de los textos relevantes del estudio, que conlleva a sistematizarlo y extraer la información requerida para ser asociado a identificar aquellos patrones donde se localiza el enfoque de construcción de forma automática de los planes.

La extracción de datos automatiza la tarea de recopilar información relevante de los expedientes judiciales. En lugar de depender totalmente de la revisión manual, la extracción automatizada acelera el proceso y reduce la carga de trabajo manual, mejorando la eficiencia y la productividad.

Los expedientes judiciales suelen contener una gran cantidad de documentos, muchos de los cuales pueden no ser pertinentes para un caso específico. La extracción de datos permite identificar y extraer rápidamente la información relevante, acelerando el proceso de revisión y análisis.

La automatización de la extracción de datos reduce la probabilidad de errores humanos asociados con la lectura y la interpretación manual de documentos. Esto es crítico en el ámbito legal, donde la precisión en la información es fundamental para tomar decisiones informadas.

La extracción de datos estructura la información de los expedientes de una manera que facilita el análisis. Puede ayudar a organizar y presentar la información de manera clara, permitiendo una comprensión rápida y profunda de los detalles del caso. Siendo esencial cuando se manejan grandes volúmenes de expedientes

judiciales. Permitiendo procesar rápidamente grandes cantidades de información, identificando patrones, tendencias y relaciones que podrían pasar desapercibidos en una revisión manual.

VI.CONCLUSIONES

- Primero: Con respecto a los resultados obtenidos en la presente tesis, bajo la implementación del marco teórico estudiado se verificó una influencia positiva fuerte entre las variables de la investigación, evidenciando un coeficiente de Rho de Spearman entre extracción de datos y la productividad de 0.962; lo que nos brinda un importante efecto del procesamiento del lenguaje natural que nos ayudaría a identificar y extraer información clave de documentos legales de manera automática, facilitando la revisión y comprensión de los documentos, ahorrando tiempo al personal jurisdiccional.
- Segundo: Con respecto a los resultados obtenidos en la presente tesis, se verificó una influencia positiva fuerte del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, evidenciado en el coeficiente de Rho de Spearman que asciende a 0.894, ello debido a que el preprocesamiento permite la clasificación automática de documentos en categorías específicas. Esto facilita la organización de los expedientes judiciales, permitiendo un acceso más rápido a la información relevante y mejorando la eficiencia en la gestión de casos.
- Tercero: Con respecto a los resultados obtenidos en la presente investigación, se identificó una influencia positiva débil de la extracción de los datos en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, evidenciando un coeficiente de Rho de Spearman que asciende a 0.353; considerando que los expedientes judiciales a menudo contienen documentos legales complejos con estructuras y lenguaje técnico. La extracción de datos puede enfrentar desafíos al tratar con esta complejidad, especialmente si los documentos no siguen un formato estándar. Esto puede limitar la capacidad de extracción precisa de información clave.

VII.RECOMENDACIONES

En cuanto al objetivo general para las oficinas especializadas de informática se debe utilizar el PNL como aquel recurso que permita incidir en optimizar recursos y tiempos que conlleva a una reforma completa en la gobernanza de datos del sistema judicial y contribuir a brindar resultados óptimos y lograr las metas establecidas.

En cuanto al primero objetivo específico al facilitar el acceso oportuno de información, se recomienda a los responsables de las diferentes áreas de los Juzgados Especializados implementar un marco de trabajo innovador para lo cual se establece al área de tecnología de información implementar el preprocesamiento para la clasificación de forma rápida de los expedientes judiciales y un acceso oportuno a la gestión de información para lograr atender de forma oportuna los expedientes judiciales.

En cuanto al segundo objetivo específico para el desarrollo de plantillas de extracción se recomienda a los responsables de las diferentes áreas de los Juzgados Especializados agilizar la extracción de datos al enfocarse en la información clave que se encuentra en ubicaciones predecibles en los documentos para integrar las soluciones de extracción de datos con sistemas de gestión documental para una transición más suave de la información entre diferentes plataformas para mejorar la capacidad de adaptación del sistema a nuevos tipos de documentos o formatos.

REFERENCIAS

- Alanoca, H. A., Chahuares, B. A., Caceres, K. A., & Saire, J. C. (2023). Neural Machine Translation for Aymara to Spanish. En K. Arai (Ed.), *Intelligent Systems and Applications* (pp. 290-298). Springer International Publishing. https://doi.org/10.1007/978-3-031-16075-2_19
- Arceda, V. E. M. (2023). Neoantigen Detection Using Transformers and Transfer Learning in the Cancer Immunology Context. En M. Rocha, F. Fdez-Riverola, M. S. Mohamad, & A. B. Gil-González (Eds.), *Practical Applications of Computational Biology and Bioinformatics, 17th International Conference (PACBB 2023)* (pp. 97-102). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-38079-2_10
- Appadoo, K., Soonnoo, M. B., & Mungloo-Dilmohamud, Z. (2020). Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1-6. <https://doi.org/10.1109/CSDE50874.2020.9411584>
- Aguilar, G. A. (2012). Competencias digitales y docencia: Una experiencia desde la práctica universitaria. *Innovación Educativa*, 12, 21
- Bender, E. M., & Lascarides, A. (2022). *Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics* (1.^a ed., Vol. 1). Springer Nature. https://books.google.es/books?hl=es&lr=&id=_4hyEAAAQBAJ&oi=fnd&pg=PP1&dq=natural+language+processing+and+productivity&ots=FkusVXdi4A&sig=RrWXnwcV3CG7Te57UVKks1_8Np4#v=onepage&q=natural%20language%20processing%20and%20productivity&f=false
- Baca, L., Ardiles, N., Cruz, J., Mamani, W., & Capcha, J. (2023). Deep Learning Model Based on a Transformers Network for Sentiment Analysis Using NLP in Sports Worldwide. En M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, & T. Ören (Eds.), *Advances in Computing and Data Sciences* (pp. 328-339). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37940-6_27
- Bankar, S., & Shukla, K. (2023). Performance Management and Artificial Intelligence: A Futuristic Conceptual Framework. En S. Grima, K. Sood, & E. Özen (Eds.), *Contemporary Studies of Risks in Emerging Technology, Part*

- B (pp. 341-360). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-80455-566-820231019>
- Barsha, S., & Munshi, S. A. (2023). Implementing artificial intelligence in library services: A review of current prospects and challenges of developing countries. *Library Hi Tech News, ahead-of-print*(ahead-of-print). <https://doi.org/10.1108/LHTN-07-2023-0126>
- Bhatt, P., & Muduli, A. (2022). Artificial intelligence in learning and development: A systematic literature review. *European Journal of Training and Development, 47*(7/8), 677-694. <https://doi.org/10.1108/EJTD-09-2021-0143>
- Cárdenas, E. R., & Molano, V. M. (2021). Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales. *Revista Direito GV, 17*(1), e2101. <https://doi.org/10.1590/2317-6172202101>
- Castellano, P. S. (2021). Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justitia? *IDP. Revista de Internet, Derecho y Política, 33*, Article 33. <https://doi.org/10.7238/idp.v0i33.373817>
- Chamorro, O., Arce, D., Morales, G., León, C., Ramos, P., Auqui, E., & Levano, M. (2022). Sentiment analysis through twitter as a mechanism for assessing university satisfaction. *Indonesian Journal of Electrical Engineering and Computer Science, 28*(1), Article 1. <https://doi.org/10.11591/ijeecs.v28.i1.pp430-440>
- Chu, C., & Wang, H. (1988). The Use of Artificial Intelligence in Process Planning. *International Journal of Operations & Production Management, 8*(1), 5-18. <https://doi.org/10.1108/eb054810>
- Diwan, S. A. (2023). Implementation Patterns of Natural Language Processing Using Pre-Trained Deep Learning Models. *International Journal of Intelligent Systems and Applications in Engineering, 11*(1), Article 1.
- D'Orazio, M., Bernardini, G., & Di Giuseppe, E. (2023). Automated Priority Assignment of Building Maintenance Tasks Using Natural Language Processing and Machine Learning. *Journal of Architectural Engineering, 29*(3), 04023027. <https://doi.org/10.1061/JAEIED.AEENG-1516>
- Dhamija, P., & Bag, S. (2020). Role of artificial intelligence in operations environment: A review and bibliometric analysis. *The TQM Journal, 32*(4), 869-896. <https://doi.org/10.1108/TQM-10-2019-0243>

- Delaney, S., Chan, C. C. K., & Smith, D. (2018). Natural Language Processing for Productivity Metrics for Software Development Profiling in Enterprise Applications. *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*, 83-87. <https://doi.org/10.1145/3299819.3299830>
- Farfán, J. L., Farfán, J. A., Farfán, B., & Núñez, J. P. (2023). Inteligencia artificial y Derecho: ¿La justicia en manos de la IA? *Frónesis: Revista de filosofía jurídica, social y política*, 30(2), 173-197.
- Fernández, J. R. T. (2019). *Programa de gestión estratégica para la ejecución de obras mediante la modalidad de núcleo ejecutor en el Programa Nacional de Saneamiento Rural-PNSR*. [Tesis de posgrado, Universidad Nacional Pedro Ruiz Gallo]. <http://repositorio.unprg.edu.pe/handle/20.500.12893/5905>
- Goldrick, S., Alosert, H., Lovelady, C., Bond, N. J., Senussi, T., Hatton, D., Klein, J., Cheeks, M., Turner, R., Savery, J., & Farid, S. S. (2023). Next-generation cell line selection methodology leveraging data lakes, natural language generation and advanced data analytics. *Frontiers in Bioengineering and Biotechnology*, 11. <https://www.frontiersin.org/articles/10.3389/fbioe.2023.1160223>
- Garcés, D. M. (2020). *Método de procesamiento de lenguaje natural y técnicas de minería de datos aplicadas a la clasificación de incidentes informáticos* [Tesis de pregrado, Universidad de Lima]. <https://repositorio.ulima.edu.pe/handle/20.500.12724/11703>
- Han, S., Zhang, R. F., Shi, L., Richie, R., Liu, H., Tseng, A., Quan, W., Ryan, N., Brent, D., & Tsui, F. R. (2022). Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *Journal of Biomedical Informatics*, 127, 103984. <https://doi.org/10.1016/j.jbi.2021.103984>
- Hidalgo, P., Luna, A., & Rivas, L. (2022). Social and Scientific Relevance of Peruvian Undergraduate Theses in Computer and Systems Engineering Schools. *2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, 229-234. <https://doi.org/10.1109/TALE54877.2022.00045>
- Herrera, J. D., & Madril, Kevin M. (2022). *Estandarización del sistema productivo para el mejoramiento de la productividad de la empresa «Rectificadora*

- Cotopaxi*» [Tesis de pregrado, Latacunga: Universidad Técnica de Cotopaxi (UTC)]. <http://repositorio.utc.edu.ec/handle/27000/9564>
- Heyman, G., Huyssegems, R., Justen, P., & Van Cutsem, T. (2021). Natural language-guided programming. *Proceedings of the 2021 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, 39-55. <https://doi.org/10.1145/3486607.3486749>
- Kobayashi, D., Yasuda, S., Iuti, T., & Ito, S. (2022). Application of Natural Language Processing in Semiconductor Manufacturing. *2022 International Symposium on Semiconductor Manufacturing (ISSM)*, 1-3. <https://doi.org/10.1109/ISSM55802.2022.10026893>
- Kolthoff, K., Bartelt, C., & Ponzetto, S. P. (2023). Data-driven prototyping via natural-language-based GUI retrieval. *Automated Software Engineering*, 30(1), 13. <https://doi.org/10.1007/s10515-023-00377-x>
- Kannan, D., Mamatha, T., & Kausar, F. (2023). Latest Trends in Language Processing to Make Semantic Search More Semantic. En *Data Science with Semantic Technologies: Deployment and Exploration* (pp. 29-48). Scopus. <https://doi.org/10.1201/9781003310792-3>
- Kumar, V. S., Alemran, A., Karras, D. A., Kant Gupta, S., Kumar Dixit, C., & Haralayya, B. (2022). Natural Language Processing using Graph Neural Network for Text Classification. *2022 International Conference on Knowledge Engineering and Communication Systems (ICKES)*, 1-5. <https://doi.org/10.1109/ICKES56523.2022.10060655>
- Kulkarni, A., & Shivananda, A. (2021). *Natural Language Processing Recipes* (2.^a ed., Vol. 2). Apress. <https://doi.org/10.1007/978-1-4842-7351-7>
- Lage, A., Allende, H., Santana, O., & Oliveira, L. (2022). Predicting Brazilian Court Decisions. *PeerJ Computer Science*, 8, e904. <https://doi.org/10.7717/peerj-cs.904>
- Leadbetter, L. (2023). *Maximizing Productivity and Quality in Senior Thesis Writing with Artificial Intelligence and Natural Language Processing Driven Tools* [Tesis de posgrado, CLAREMONT MCKENNA COLLEGE]. https://scholarship.claremont.edu/cmc_theses/3285

- Lee, T.-Y., Li, C.-C., Chou, K.-R., Chung, M.-H., Hsiao, S.-T., Guo, S.-L., Hung, L.-Y., & Wu, H.-T. (2023). Machine learning-based speech recognition system for nursing documentation – A pilot study. *International Journal of Medical Informatics*, 178. Scopus. <https://doi.org/10.1016/j.ijmedinf.2023.105213>
- Manoharan, G., Durai, S., Rajesh, G. A., & Ashtikar, S. P. (2023). A Study on the Application of Natural Language Processing Used in Business Analytics for Better Management Decisions A Literature Review. En *Artificial Intelligence and Knowledge Processing: Improved Decision-Making and Prediction* (pp. 249-261). Scopus. <https://doi.org/10.1201/9781003328414-21>
- Medina, M. R., Costanzo, S. B., Cuadrado, M. V., & Carmignani, M. G. (2022). *Enfoques teóricos para elaborar modelos de interpretación: El tipo urbano en América Latina*. IX Encuentro de Docentes e Investigadores de Historia de la Arquitectura, el Diseño y la Ciudad «Jorge Ramos de Dios» (La Plata y Buenos Aires, 17 al 19 de agosto de 2022). <http://sedici.unlp.edu.ar/handle/10915/150030>
- Masakowski, Y. R. (2020). Artificial Intelligence and the Future Global Security Environment. En Y. R. Masakowski (Ed.), *Artificial Intelligence and Global Security* (pp. 1-34). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78973-811-720201001>
- Mascarenhas, S. J. FR. O. A. J., & Mascarenhas, S. J. FR. O. A. J. (2018). Artificial Intelligence and the Emergent Turbulent Markets: New Challenges to Corporate Ethics Today. En *Corporate Ethics for Turbulent Markets* (pp. 215-242). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-187-820181008>
- Porter, M. E. (2017). *Ser competitivo: Edición actualizada y aumentada* (1.ª ed., Vol. 1). Deusto. https://www.google.com.pe/books/edition/Ser_competitivo/8wE2DwAAQBAJ?hl=es&gbpv=0
- Quispe, J. (2020). *Propuesta de un modelo de gestión para el adecuado manejo, caracterización y almacenamiento de residuos peligrosos en los laboratorios de la Universidad Nacional De Ucayali, provincia de Coronel Portillo—Ucayali, Perú, 2020* [Tesis de pregrado, Universidad Nacional de Ucayali].

- http://repositorio.unu.edu.pe/bitstream/handle/UNU/4319/UNU_AMBIENTAL_2020_T_JULIAN-QUISPE-PILCO.pdf?sequence=1&isAllowed=y
- Rehman, M., Razzaq, A., Baig, I. A., Jabeen, J., Tahir, M. H. N., Ahmed, U. I., Altaf, A., & Abbas, T. (2022). Semantics Analysis of Agricultural Experts' Opinions for Crop Productivity through Machine Learning. *Applied Artificial Intelligence*, 36(1), 2012055. <https://doi.org/10.1080/08839514.2021.2012055>
- Ramalho, A. (2017). *Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems* (SSRN Scholarly Paper 2987757). <https://doi.org/10.2139/ssrn.2987757>
- Samant, R. M., Bachute, M. R., Gite, S., & Kotecha, K. (2022). Framework for Deep Learning-Based Language Models Using Multi-Task Learning in Natural Language Understanding: A Systematic Literature Review and Future Directions. *IEEE Access*, 10, 17078-17097. <https://doi.org/10.1109/ACCESS.2022.3149798>
- Sun, C., Hahn, C., & Trippel, C. (2023, junio 2). *Towards Improving Verification Productivity with Circuit-Aware Translation of Natural Language to SystemVerilog Assertions*. First International Workshop on Deep Learning-aided Verification. <https://openreview.net/forum?id=FKH8qCuM44>
- Segura, R. E. (2023). Inteligencia artificial y administración de justicia: Desafíos derivados del contexto latinoamericano. *Revista de Bioética y Derecho*, 58, 45-72. <https://doi.org/10.1344/rbd2023.58.40601>
- Torres, M. L., & Iregui, P. M. (Eds.). (2020). *Diálogos interdisciplinarios sobre los problemas de la justicia en Colombia*. Universidad del Rosario. <https://doi.org/10.12804/urosario9789587845617>
- Vega, J. E. L., Becerra, V., Guzmán, M. de los Á., & Sandovalaa, J. K. L. (2023). Inteligencia artificial en la justicia colombiana: ¿La solución a la congestión judicial? *Revista Lecciones Vitales*, lv0101-lv0101. <https://doi.org/10.18046/rlv.2023.5655>
- Walters, R., & Novak, M. (2021). Artificial Intelligence and Law. En R. Walters & M. Novak (Eds.), *Cyber Security, Artificial Intelligence, Data Protection & the Law* (pp. 39-69). Springer. https://doi.org/10.1007/978-981-16-1665-5_3

- Wamba, S.-L., Fosso, S., Kala, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Yan, S., Yu, H., Chen, Y., Shen, B., & Jiang, L. (2020). Are the Code Snippets What We Are Searching for? A Benchmark and an Empirical Study on Code Search with Natural-Language Queries. *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 344-354. <https://doi.org/10.1109/SANER48275.2020.9054840>
- Zhou, J., & Cen, W. (2023). Design and Application Research of a Digital Human Resource Management Platform based on ChatGPT. *Journal of Theory and Practice of Social Science*, 3(7), Article 7. [https://doi.org/10.53469/jtpss.2023.03\(07\).10](https://doi.org/10.53469/jtpss.2023.03(07).10)

ANEXOS

Anexo 1

Matriz de consistencia

Título: Procesamiento del lenguaje natural y su productividad en los expedientes judiciales en un distrito judicial Lima Norte, 2023 Autor: Erick David Lucero Carrillo								
Problema general	Objetivo general	Hipótesis general	Variable	Dimensiones	Indicadores	Ítem	Instrumento	Escala
¿Cómo el procesamiento del lenguaje natural influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023?	Determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023	El procesamiento del lenguaje natural influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023	Variable 1 Procesamiento del Lenguaje Natural (NLP)	Preprocesamiento	<ul style="list-style-type: none"> Limpieza de texto Tokenización Eliminación de palabras 		Fichaje	Cuantitativo-Razón
				Extracción de datos	<ul style="list-style-type: none"> Etiquetado de texto Extracción de palabras claves 			
Objetivos específicos	Objetivos específicos	Hipótesis específicas						
¿Cómo el preprocesamiento influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023?	Analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023	El preprocesamiento del influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023	Variable 2 Productividad		<ul style="list-style-type: none"> Datos de salida Error General = $(\text{Error} + \text{Extrac}) / \text{Datos de entrada}$ 		Fichaje	Cuantitativo Razón
				Tiempo de Registro				
¿Cómo la extracción de datos influye en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023?	Analizar la influencia de la extracción de datos en la productividad de los expedientes judiciales	La extracción de datos influye de forma positiva en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023						

	en un distrito judicial Lima Norte, 2023							
Metodología								
Método y Diseño		Población y muestra		Técnicas e instrumentos			Método de análisis de datos	
Enfoque: cuantitativo Tipo: Aplicada Diseño: No experimental		Población: 21,521 expedientes judiciales de Lima Norte del periodo Enero-Septiembre 2023 Muestra: 378 expedientes judiciales de Lima Norte. Muestreo: Probabilístico		Técnicas: Análisis documental Instrumentos: Fichaje			Descriptiva: Para el análisis descriptivo se utilizará Microsoft Excel y paquete estadístico SPSS, para poder recolectar la data y se pueda generar de manera descriptiva y gráfica para analizar e interpretar los resultados. Inferencial: Se realizará la prueba de Kolmogórov-Smirnov; al no encontrar normalidad se utilizará la prueba Rho de Pearson.	

Anexo 02

Guía observacional documental

Ficha de análisis documental (Fichaje)

"Procesamiento del lenguaje natural y su productividad en los expedientes judiciales en un distrito judicial Lima Norte, 2023"

Este instrumento de recolección de datos permitirá sistematizar los datos que se encuentran disponibles en el sistema institucional del juzgado de Lima Norte, es considerado un instrumento práctico y orientador.

Por lo cual emplearemos los datos sistematizados de la siguiente manera:

Variable: Procesamiento del lenguaje natural

	Dimensión 1: Preprocesamiento
Expediente	Tiempo de limpieza de texto (segundos) Tiempo de ejecución de preprocesamiento (segundos)
1	
2	
3	
....	
13	

Fuente: Instituto Nacional de Estadística e Informática, 2002-2022.

Variable: Productividad

	Dimensión 1: Tiempo de Registro
Expediente	Productividad. General = Tiempo de procesamiento total / Datos de entrada
1	
2	
3	
....	
13	

Fuente: Instituto Nacional de Estadística e Informática, 2002-2022.

Anexo 03

Algoritmos del procesamiento en Python

Preprocesamiento

```
import time, pytesseract, os, csv, re, shutil, subprocess
import pandas as pd
from PIL import Image
from pdf2image import convert_from_path
from pdf2image import convert_from_bytes
from os import remove
from pypdf import PdfReader

tiempo_inicio = time.time()

def mover_archivos_pdf(ruta_archivo, ruta_destino):
    shutil.move(ruta_archivo, ruta_destino)

def eliminar_archivos(carpeta_padre, archivosinextension,
v_extension):
    for carpeta in os.walk(carpeta_padre):
        carpeta_nueva = carpeta[0].replace('\\', '/') + "/"
        for ruta in carpeta[2]:
            if ruta.endswith(v_extension):
                todaruta = carpeta_nueva+ruta
                os.remove(todaruta)

def crear_matriz(destino_procesado_csv, archivosinextension):
    t = open(destino_procesado_csv+archivosinextension+'.csv', 'w',
encoding="utf8")
    lineat = "cod_exp" + ";" + "contenido" + ";" + "pag" + ";" +
"etiqueta" + "\n"
    t.write(lineat)
    # t.close()

def convierteMilisegundos(segundos):
    horas = int(segundos / 60 / 60)
    segundos -= horas*60*60
    minutos = int(segundos/60)
    segundos -= minutos*60
    print("Tiempo de ejecución: ",round(horas),"horas
:",round(minutos),"minutos :", round(segundos) ,"segundos")

def guarda_paginas_csv(archivosinextension, texto_completo, numpagina,
destino_procesado_csv):
    x_pag = 0
    with open(destino_procesado_csv+archivosinextension+".csv", 'a',
newline='', encoding='utf-8') as archivo_csv:
```

```

writer = csv.writer(archivo_csv, delimiter=';')

for dato in texto_completo:
    #print(dato)
    x_pag = x_pag + 1
    if (x_pag <= numpagina):
        #dato2 = dato.replace("Ã³", "---")
        writer.writerow([archivosinextension, dato, x_pag])

def procesar(carpeta_padre, ruta, archivosinextension,
destino_procesado_csv):
    contador = 0
    reader = PdfReader(ruta)
    texto_completo = []
    txt = ""
    for page in reader.pages:
        contador = contador + 1
        texto = page.extract_text()
        if len(texto)>0:
            #print("pagina=",contador,len(texto), "texto")
            txt = texto
        else:
            #print("pagina=",contador,len(texto), "imagen")
            count = 0
            try:
                for image_file_object in page.images:
                    with open(os.path.join(carpeta_padre, str(count) +
image_file_object.name), "wb") as fp:
                        fp.write(image_file_object.data)
                        image_path = os.path.join(carpeta_padre,
str(count) + image_file_object.name)
                        image = Image.open(image_path)
                        texto = pytesseract.image_to_string(image,
lang='spa')

                        txt = texto
                        count +=
1

            except:
                txt = "Observado"
                continue

        texto_completo.append(txt)

    guarda_paginas_csv(archivosinextension, texto_completo, contador,
destino_procesado_csv)

# Inicio

```

```

carpeta_padre = "D:/ftp_amauta_pro_csjln/demandas/2022/"
destino_pdf_procesado = "01_1_PROCESADO/PDF/"
destino_procesado_csv = "01_1_PROCESADO/CSV/"
lista_titulo=[]
lista_numpag=[]
lista_con=[]
matriz = []

# recorrer carpetas y busca archivos PDF
for carpeta in os.walk(carpeta_padre):
    carpeta_nueva = carpeta[0].replace('\\', '/')+"/"
    for ruta in carpeta[2]:
        if ruta.endswith('.pdf'):
            todaruta = carpeta_nueva+ruta
            archivosinextension = os.path.splitext(ruta)[0]
            crear_matriz(destino_procesado_csv,archivosinextension)

            procesar(carpeta_padre, todaruta, archivosinextension,
destino_procesado_csv)
            # mover_archivos_pdf(todaruta, destino_pdf_procesado)
            eliminar_archivos(carpeta_padre, archivosinextension, '.png')
            eliminar_archivos(carpeta_padre, archivosinextension,
'.jpg')

tiempo_fin = time.time()

# print(tiempo_fin - tiempo_inicio)
# caluclar tiempo de ejecucion
convierteMilisegundos(tiempo_fin - tiempo_inicio) # caluclar tiempo de
ejecucion

# subprocess.call(['python', '02_etiqueta.py'])

```

Extracción de Datos

```

import time
import pandas as pd
import numpy as np
from PyPDF2 import PdfReader
from datetime import datetime
import re
import os
import shutil
import random

def crear_encabezado(ruta_csv_destino):
    # Crear el encabezado

```

```
encabezados = ['expediente', 'org. jurisdiccional', 'especialista',
'fec. inicio', 'motivo ingreso', 'proceso', 'materia',
                'fec. presentacion', 'folios', 'depo. judicial',
'arancel', 'parte. agresora', 'parte. victima', 'presentado por',
                'casilla', 'cod. digitalizacion', 'tipo denuncia',
'fec. registro', 'fec. hecho', 'tificacion', 'lugar hecho',
                'denunciante', 'denunciado', 'puntaje', 'riesgo',
'fundamentos', 'decision']
```

```
df_csv = pd.DataFrame.from_dict(encabezados)
df_csv.to_csv(ruta_csv_destino, header=True,
              index=False, sep=';', encoding='utf-8')
```

```
def documento_caratula_01(texto_modificado):
    resultados = []
    try:
        uno = texto_modificado.split('expediente')
        if (len(uno) > 0):
            dos = uno[1].split('org. jurisdiccional ')
            cod_exp = re.sub(r'\n', '', dos[0])
            cod_exp = cod_exp.upper()
            cod_exp = cod_exp.strip()
        else:
            cod_exp = "NO HAY DATO (Parte 1)"

        tres = uno[1].split('org. jurisdiccional ')
        if (len(tres) > 0):
            cuatro = tres[1].split('especialista')
            nom_juzg = re.sub(r'\n', '', cuatro[0])
            nom_juzg = nom_juzg.upper()
            nom_juzg = nom_juzg.strip()
        else:
            nom_juzg = "NO HAY DATO (Parte 1)"

        cinco = tres[1].split('especialista ')
        if (len(cinco) > 0):
            seis = cinco[1].split('fec. inicio')
            nom_espec = re.sub(r'\n', ' ', seis[0])
            nom_espec = nom_espec.upper()
            nom_espec = nom_espec.strip()
        else:
            nom_espec = "NO HAY DATO (Parte 1)"

        siete = cinco[1].split('fec. inicio ')
        if (len(siete) > 0):
            ocho = siete[1].split('motivo de ingreso')
```

```

        f_inicio = re.sub(r'\n', '', ocho[0])
        f_inicio = f_inicio.strip()
    else:
        f_inicio = "NO HAY DATO (Parte 1)"

nuevo1 = texto_modificado.split('motivo de ingreso')
if (len(nuevo1) > 0):
    nuevo2 = nuevo1[1].split('proceso')
    motivo = re.sub(r'\n', '', nuevo2[0])
    motivo = motivo.upper()
    motivo = motivo.strip()
else:
    motivo = "NO HAY DATO (Parte 1)"

nuevo3 = nuevo1[1].split('proceso')
if (len(nuevo3) > 0):
    nuevo4 = nuevo3[1].split('materia')
    proceso = re.sub(r'\n', ' ', nuevo4[0])
    patron = re.compile(r'(\D+)(\d+)')
    proceso = patron.sub(r'\1 \2', proceso)
    proceso = proceso.upper()
    proceso = proceso.strip()
else:
    proceso = "NO HAY DATO (Parte 1)"

nueve = siete[1].split('materia ')
if (len(nueve) > 0):
    diez = nueve[1].split('fecha de presentación')
    materia = re.sub(r'\n', ' ', diez[0])
    materia = materia.upper()
    materia = materia.strip()
else:
    materia = "NO HAY DATO (Parte 1)"

once = nueve[1].split('fecha de presentación ')
if (len(once) > 0):
    doce = once[1].split('folios')
    f_presentacion = doce[0][0:19]
    f_presentacion = f_presentacion.strip()
else:
    f_presentacion = "NO HAY DATO (Parte 1)"

folios = 'Folio'
depo_judicial = "Deposito"
arancel = 'Arancel'

resultados = [cod_exp, nom_juzg, nom_espec,

```

```

        f_inicio, motivo, proceso, materia, f_presentacion,
folios, depo_judicial, arancel]

    except Exception as e:
        resultados = [cod_exp, nom_juzg, nom_espec,
                      f_inicio, motivo, proceso, materia, f_presentacion,
folios, depo_judicial, arancel]

    return resultados

def documento_caratula_02(texto_modificado):
    resultados = []
    try:
        diecinueve = texto_modificado.split('presentado electrónicamente
por')
        if len(diecinueve) > 0:
            veinte = diecinueve[1].split('número de casilla')
            veintiuno = veinte[0].split(' - ')
            exp_comisaria = re.sub(r'\n', ' ', veintiuno[1])
            exp_comisaria = exp_comisaria.upper()
        else:
            exp_comisaria = "NO DATO - Pag 2"

        veintidos = diecinueve[1].split('número de casilla')
        if len(veintidos) > 0:
            veintitres = veintidos[1].split('cod. digitalización.')
            casilla = re.sub(r'\n', ' ', veintitres[0])
        else:
            casilla = "NO DATO - Pag 2"

        veinticuatro = veintidos[1].split('cod. digitalización.')
        if len(veinticuatro) > 0:
            veinticinco = veinticuatro[1].split('\n')
            cod_recep = veinticinco[0]
            cod_recep = cod_recep.upper()
        else:
            cod_recep = "NO DATO - Pag 2"

        resultados = [exp_comisaria, casilla, cod_recep]

    except Exception as e:
        resultados = [exp_comisaria, casilla, cod_recep]

    return resultados

def obtener_datos_tipificacion(parrafo_tipo):
    tipo_violencia = ''
    # Palabras clave a buscar

```

```

palabras_clave = ['economica', 'psicologica',
                  'emocional', 'fisica', 'sexual']

# palabras_clave = ['violencia economica', 'violencia psicologica',
#                   'violencia emocional', 'violencia
fisica', 'violencia sexual']
# palabras_clave = ['violencia psicologica']
patron = re.compile(
    r'\b(?:' + '|'.join(map(re.escape, palabras_clave)) + r')\b',
    re.IGNORECASE)

coincidencias = patron.findall(parrafo_tipo)
if len(coincidencias) > 0:
    for palabra in palabras_clave:
        count = coincidencias.count(palabra)
        if count > 0:
            tipo_violencia += "VIOLENCIA " + palabra + ', '
            tipo_violencia = tipo_violencia.rstrip(', ')
            tipo_violencia = tipo_violencia.upper()
        else:
            tipo_violencia = 'No se encontraron coincidencias'

return tipo_violencia

def obtener_datos_lugar_hecho(parrafo_denunciante):
    patron = r'(\w+)\s*\s*(\w+)\s*\s*(\w+)\s*\s*(.*)\s*\s*(.*)'
    lugar_hecho = ""
    resultado = re.search(patron, parrafo_denunciante)
    if resultado:
        departamento = resultado.group(1)
        provincia = resultado.group(2)
        distrito = resultado.group(3)
        direccion = resultado.group(4)
        if (departamento):
            departamento = departamento
        else:
            departamento = "NN_DEP"

        if (provincia):
            provincia = provincia
        else:
            provincia = "NN_PROV"

        if (distrito):
            distrito = distrito
        else:
            distrito = "NN_DIS"

```

```

        if (direccion):
            direccion = direccion
        else:
            direccion = "NN_DIREC"

        lugar_hecho = departamento + ', ' + provincia + ', ' + distrito +
', ' + direccion
    else:
        lugar_hecho = 'NO TIENE LUGAR HECHO'
    return lugar_hecho

def documento_pnp_00(texto_modificado):
    resultados = []
    resultados = ["DENUNCIA"]

    return resultados

def documento_pnp_01(texto_modificado):
    resultados = []
    try:
        pag_pnp_uno = texto_modificado.split('fecha y hora registro ')
        if (len(pag_pnp_uno) > 0):
            f_registro = pag_pnp_uno[1].split('hrs')
            f_registro = f_registro[0]
        else:
            f_registro = "NO EXISTE DATO"

        resultados = [f_registro]

    except Exception as e:
        f_registro = "NN-PNP-1"
        resultados = [f_registro]

    return resultados

def documento_pnp_02(texto_modificado):
    resultados = []
    try:
        pag_pnp_dos = texto_modificado.split('fecha y hora hecho ')
        if (len(pag_pnp_dos) > 0):
            f_hecho = pag_pnp_dos[1].split('hrs')
            f_hecho = f_hecho[0]
        else:
            f_hecho = "NO EXISTE DATO"

```

```

        resultados = [f_hecho]

    except Exception as e:
        f_hecho = "NN-PNP-1"
        resultados = [f_hecho]

    return resultados

def documento_pnp_03(texto_modificado):
    resultados = []
    try:
        pag_pnp_tres = texto_modificado.split('tipificacion')
        if (len(pag_pnp_tres) > 0):
            # parrafo_tipo = pag_pnp_tres[1].split('lugar del hecho')
            parrafo_tipo = pag_pnp_tres[1].split('contenido')
            parrafo_tipo = obtener_datos_tipificacion(parrafo_tipo[0])
        else:
            parrafo_tipo = "NO EXISTE DATO"

        resultados = [
            parrafo_tipo]

    except Exception as e:
        parrafo_tipo = "NN-PNP-1"
        resultados = [parrafo_tipo]

    return resultados

def documento_pnp_04(texto_modificado):
    resultados = []
    try:
        pag_pnp_cuatro = texto_modificado.split('lugar del hecho')
        if (len(pag_pnp_cuatro) > 0):
            parrafo_lugar = pag_pnp_cuatro[1].split('denunciante')
            parrafo_lugar = obtener_datos_lugar_hecho(
                parrafo_lugar[0])
        else:
            parrafo_lugar = "NO EXISTE DATO"

        resultados = [parrafo_lugar]

    except Exception as e:
        parrafo_lugar = "NN-PNP-1"
        resultados = [parrafo_lugar]

```

```

    return resultados

def documento_pnp_05(texto_modificado):
    resultados = []
    try:
        pag_pnp_cinco = texto_modificado.split('total puntaje ')
        if (len(pag_pnp_cinco) > 0):
            puntaje = pag_pnp_cinco[1].split('\n')
            puntaje = puntaje[0]
        else:
            puntaje = "NO EXISTE DATO"

        resultados = [puntaje]

    except Exception as e:
        puntaje = "NN-PNP-1"
        resultados = [puntaje]

    return resultados

def documento_pnp_06(texto_modificado):
    resultados = []
    resultados = ["DENUNCIANTE - 1", "DENUNCIADO - 2"]
    return resultados

def limpiar_texto_pagina_caratula_01(txt_concatenados):
    simbolos_pnp = ['\n']
    texto_limpio_pnp = txt_concatenados
    for simbolo in simbolos_pnp:
        if simbolo in texto_limpio_pnp:
            texto_limpio_pnp = texto_limpio_pnp.replace(simbolo, "")
    return texto_limpio_pnp

def limpiar_texto_pagina_caratula_02(txt_concatenados):
    simbolos_pnp = [':']
    texto_limpio_pnp = txt_concatenados
    for simbolo in simbolos_pnp:
        if simbolo in texto_limpio_pnp:
            texto_limpio_pnp = texto_limpio_pnp.replace(simbolo, "")
    return texto_limpio_pnp

def eliminar_lineas_texto_pagina_caratula_parte01(txt_concatenados):
    # Borra la línea que contiene "Mesa de Partes Electrónica"

```

```

lineas = txt_concatenados.split('\n')
lineas_filtradas = [
    linea for linea in lineas if "mesa de partes electrónica" not in
linea]
texto_modificado = '\n'.join(lineas_filtradas)

# Borra la línea que contiene "PARTES PROCESALES :"
lineas = texto_modificado.split('\n')
lineas_filtradas = [
    linea for linea in lineas if "partes procesales :" not in linea]
texto_modificado = '\n'.join(lineas_filtradas)

return texto_modificado

def buscar_documento_pnp_extraer_00(texto):
    resultados = {}
    lista_tipodoc = {
        'CARATULA_02': ['policia nacional del peru']
    }
    for clave, palabras in lista_tipodoc.items():
        coincidencias = []
        for palabra in palabras:
            coincidencias_palabra = re.findall(r'\b{}\b'.format(
                re.escape(palabra)), str(texto), flags=re.IGNORECASE)
            coincidencias.extend(coincidencias_palabra)
        if len(coincidencias) > 0:
            resultados[clave] = coincidencias
    return list(resultados.keys())

def buscar_documento_pnp(texto):
    resultados = {}
    lista_tipodoc = {
        'CARATULA_01': ['Cargo de Presentación de Demanda Electrónica',
'Mesa de Partes Electrónica'],
        'CARATULA_02': ['POLICIA NACIONAL DEL PERU', 'Presentado
electrónicamente por', 'casilla']
    }
    for clave, palabras in lista_tipodoc.items():
        coincidencias = []
        for palabra in palabras:
            coincidencias_palabra = re.findall(r'\b{}\b'.format(
                re.escape(palabra)), str(texto), flags=re.IGNORECASE)
            coincidencias.extend(coincidencias_palabra)
        if len(coincidencias) == 2:
            resultados[clave] = coincidencias
    return list(resultados.keys())

```

```

def celda_agrupado_todo(data_matriz):
    contenidos = {}
    for index, row in data_matriz.iterrows():
        expediente = row['cod_exp']
        contenido = str(row['contenido'])

        if expediente in contenidos:
            contenidos[expediente] += " " + contenido
        else:
            contenidos[expediente] = contenido

    df_agrupado = pd.DataFrame(contenidos.items(), columns=[
        'cod_exp', 'contenido'])

    return df_agrupado

def celda_pagina_dos_caratula(data):
    # opcion 1
    contenidos = {}
    for index, row in data.iterrows():
        expediente = row['cod_exp']
        contenido = str(row['contenido'])

        if row['pag'] == 2:
            if expediente in contenidos:
                contenidos[expediente] += " " + contenido
            else:
                contenidos[expediente] = contenido

    df_agrupado = pd.DataFrame(contenidos.items(), columns=[
        'cod_exp', 'contenido'])

    return df_agrupado

def unir_celda_caratula_principal(data):
    contenidos = {}
    for index, row in data.iterrows():
        expediente = row['cod_exp']
        contenido = str(row['contenido'])
        if row['pag'] == 1:
            contenidoA = contenido

        contenidoB = ""
        if row['pag'] == 2:
            contenidoB = contenido

        concatenado = contenidoA + " " + contenidoB

        if expediente in contenidos:

```

```

        contenidos[expediente] += " " + concatenado
    else:
        contenidos[expediente] = concatenado

df_agrupado = pd.DataFrame(contenidos.items(), columns=[
    'cod_exp', 'contenido'])
return df_agrupado

def unir_celda_pnp(data):
    contenidos = {}
    for index, row in data.iterrows():
        expediente = str(row['cod_exp'])
        contenido = str(row['contenido'])
        if expediente in contenidos:
            contenidos[expediente] += " " + contenido
        else:
            contenidos[expediente] = contenido

    df_agrupado = pd.DataFrame(contenidos.items(), columns=[
        'cod_exp', 'contenido'])
    return df_agrupado

def limpiar_caratula_principal(txt_caratula):
    contenidos = {}
    for index, row in txt_caratula.iterrows():
        expediente = str(row['cod_exp'])
        contenido = str(row['contenido']).lower()
        if expediente in contenidos:
            contenidos[expediente] += " " + contenido
        else:
            contenidos[expediente] = contenido

    df_agrupado = pd.DataFrame(contenidos.items(), columns=[
        'cod_exp', 'contenido'])
    return df_agrupado

def limpiar_pnp_minus(txt_pnp):
    contenidos = {}
    for index, row in txt_pnp.iterrows():
        expediente = row['cod_exp']
        contenido = str(row['contenido']).lower()
        if expediente in contenidos:
            contenidos[expediente] += " " + contenido
        else:
            contenidos[expediente] = contenido

```

```

df_agrupado = pd.DataFrame(contenidos.items(), columns=[
    'cod_exp', 'contenido'])
return df_agrupado

def limpiar_minuscula_pagina_dos(txt_caratula):
    contenidos = {}
    for index, row in txt_caratula.iterrows():
        expediente = str(row['cod_exp'])
        contenido = str(row['contenido']).lower()
        if expediente in contenidos:
            contenidos[expediente] += " " + contenido
        else:
            contenidos[expediente] = contenido

    df_agrupado = pd.DataFrame(contenidos.items(), columns=[
        'cod_exp', 'contenido'])
    return df_agrupado

def cambiar_simbolo_x_letra(data_pnp_cc):
    simbolos_pnp = ['o']
    texto_limpio_pnp = data_pnp_cc
    for simbolo in simbolos_pnp:
        if simbolo in texto_limpio_pnp:
            texto_limpio_pnp = texto_limpio_pnp.replace(simbolo, "o")
    return texto_limpio_pnp

def cambiar_simbolo_x_espacio_vacio(data_pnp_ec):
    simbolos_pnp = ['_', '-', '=', '?', '.', '€', '|']
    texto_limpio_pnp = data_pnp_ec
    for simbolo in simbolos_pnp:
        if simbolo in texto_limpio_pnp:
            texto_limpio_pnp = texto_limpio_pnp.replace(simbolo, "")
    return texto_limpio_pnp

def cambiar_letra_x_letra(data_pnp_ll):
    simbolos_pnp = ['hre', 'hr', 'hes', 'his', 'ht', 'hfs', 'rrs', '16;']
    texto_limpio_pnp = data_pnp_ll
    for simbolo in simbolos_pnp:
        if simbolo in texto_limpio_pnp:
            texto_limpio_pnp = texto_limpio_pnp.replace(simbolo, "hrs")
    return texto_limpio_pnp

def seleccionar_caratula(txt_caratula):
    contenidos = {}
    for index, row in txt_caratula.iterrows():
        expediente = row['cod_exp']
        contenido = row['contenido']
        data_resultado = buscar_documento_pnp(contenido)

```

```

        if (len(data_resultado) > 0):
            contenidos[expediente] = contenido

df_agrupado = pd.DataFrame(contenidos.items(), columns=[
    'cod_exp', 'contenido'])
return df_agrupado

def seleccionar_caratula_pagina_dos(data_b):
    contenidos = {}
    for index, row in data_b.iterrows():
        expediente = row['cod_exp']
        contenido = row['contenido']
        data_resultado = buscar_documento_pnp_extraer_00(contenido)
        if (len(data_resultado) > 0):
            contenidos[expediente] = contenido

df_agrupado = pd.DataFrame(contenidos.items(), columns=[
    'cod_exp', 'contenido'])
return df_agrupado

def merge_data_cruzada_policia(data_g, data_c_pnp):
    # Cruzar los datos
    df_cruzado = pd.merge(data_g, data_c_pnp, on='cod_exp')

    # Generar el nuevo DataFrame
    # Nuevo_Data = pd.DataFrame(df_cruzado['contenido'])

    return df_cruzado

def extraer_datos_parte OPCION_A(data, ruta_csv_destino):
    data_csv = {}

    # Encabezado
    #crear_encabezado(ruta_csv_destino)

    for index, row in data.iterrows():
        expediente = row['cod_exp']
        contenido = row['contenido_x']
        if expediente:

            # Parte 01
            txt_caratula = eliminar_lineas_texto_pagina_caratula_parte01(
                contenido)
            txt_caratula = limpiar_texto_pagina_caratula_01(txt_caratula)
            data_01_1 = documento_caratula_01(txt_caratula)
            # data1 = np.array(data1)

            # Parte 02_1: Partes procesales

```

```

data_02_01 = ['ABC ABC NOMBRE', 'XYZ XYZ NOMBRE']

# Parte 01_2: Datos presentado
txt_pag_02 = eliminar_lineas_texto_pagina_caratula_parte01(
    contenido)
txt_pag_02 = limpiar_texto_pagina_caratula_02(txt_pag_02)
data_02_02 = documento_caratula_02(txt_pag_02)

# Parte 03:
txt_pnp_01 = cambiar_simbolo_x_letra(contenido)
txt_pnp_01 = cambiar_simbolo_x_espacio_vacio(txt_pnp_01)
txt_pnp_01 = cambiar_letra_x_letra(txt_pnp_01)
data30 = documento_pnp_00(txt_pnp_01)
data31 = documento_pnp_01(txt_pnp_01)
data32 = documento_pnp_02(txt_pnp_01)
data33 = documento_pnp_03(txt_pnp_01)
data34 = documento_pnp_04(txt_pnp_01)
data35 = documento_pnp_05(txt_pnp_01)
data36 = documento_pnp_06(txt_pnp_01)

# Parte 4: RIESGO
data_04_1 = ["."]

# Parte 5: FUNDAMENTOS - DECISION
data_05_01 = ['fundamentos xyz', 'decision abc']

data_csv[expediente] = data_01_1 + data_02_01 + data_02_02 + \
    data30 + data31 + data32 + data33 + data34 + \
    data36 + data35 + data_04_1 + data_05_01

df_csv = pd.DataFrame.from_dict(data_csv)
df_csv.to_csv(ruta_csv_destino, header=False, index=False, sep=';',
encoding='utf-8')

def extraer_datos_parte OPCION_B(data, ruta_csv_destino):
    # Crear lista con los encabezados
    encabezados = ['expediente', 'org. jurisdiccional', 'especialista',
'fec. inicio', 'motivo ingreso', 'proceso', 'materia',
'fec. presentacion', 'folios', 'depo. judicial',
'arancel', 'parte. agresora', 'parte. victima', 'presentado por',
'casilla', 'cod. digitalizacion', 'tipo denuncia',
'fec. registro', 'fec. hecho', 'tificacion', 'lugar hecho',
'denunciante', 'denunciado', 'puntaje', 'riesgo',
'fundamentos', 'decision']

    # Crear diccionario vacío para almacenar los datos
    data_csv = {}

```

```

for encabezado in encabezados:
    data_csv[encabezado] = []

# Crear DataFrame de Pandas a partir del diccionario
df_csv = pd.DataFrame(data_csv)

# Volcar los datos en el archivo CSV a partir de la segunda fila
df_csv.iloc[1:].to_csv(ruta_csv_destino, header=True, index=False,
sep=';', encoding='utf-8')

def extraer_datos_parte_000(ruta_csv_origen, ruta_csv_destino):
    data_matriz = pd.read_csv(ruta_csv_origen, delimiter=';')
    data_csv = {}

    # Matriz completo
    txt_matriz = celda_agrupado_todo(data_matriz)
    data_g = limpiar_minuscula_pagina_dos(txt_matriz)

    #
    # Denuncia PNP: Pagina 2 (Saber que el EXPEDIENTE es de la PNO)
    txt_caratula = celda_pagina_dos_caratula(data_matriz)
    data_b = limpiar_minuscula_pagina_dos(txt_caratula)
    data_c_pnp = seleccionar_caratula_pagina_dos(data_b)

    # Merge data con la Denucia PNP
    data_d = merge_data_cruzada_policia(data_g, data_c_pnp)

    crear_encabezado(ruta_csv_destino)
    extraer_datos_parte OPCION_A(data_d, ruta_csv_destino)
    print('Terminado')

# Inicio
ruta_csv_origen =
"D:/appDesarrollo/ProyectoAmautaPro_v1/02_PROCESADO_V2/matriz_DATA_800EXP_
12716.csv"
ruta_csv_destino =
"D:/appDesarrollo/ProyectoAmautaPro_v1/02_PROCESADO_V2/matriz_DATA_800EXP_
12716_extracion_all_8.csv"

extraer_datos_parte_000(ruta_csv_origen, ruta_csv_destino)

```


Anexo 04

Validación de instrumento de recolección de datos

Experto N° 01

Evaluación por juicio de expertos

Respetado juez: Usted ha sido seleccionado para evaluar el instrumento para medir las Variables Procesamiento del Lenguaje Natural y Productividad. La evaluación del instrumento es de gran relevancia para lograr que sea válido y que los resultados obtenidos a partir de éste sean utilizados eficientemente; aportando al quehacer psicológico. Agradecemos su valiosa colaboración.

I. Datos generales del juez:

Nombre del juez:	Marlon Frank Acuña Benites		
Grado profesional:	Maestría ()	Doctor	(x)
Área de formación académica:	Clinica ()	Social	()
	Educativa ()	Organizacional	(x)
Áreas de experiencia profesional:	Educación		
Institución donde labora:	Universidad Cesar Vallejo		
Tiempo de experiencia profesional en el área:	2 a 4 años ()		
	Más de 5 años (x)		
Experiencia en Investigación <small>(si corresponde)</small>	Asesora de Tesis en Pregrado y Posgrado		

2. Propósito de la evaluación:

Validar el contenido del instrumento, por juicio de expertos.

3. Datos de la escala:

Nombre de la Prueba:	Ficha de observación
Autor:	Erick David Lucero Carrillo
Procedencia:	Lima
Administración:	Directa, virtual
Tiempo de aplicación:	25 minutos
Ámbito de aplicación:	Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los integrantes del grupo familiar de Lima Norte
Significación:	Esta guía consta de 2 fichas de observación. El objetivo es determinar la influencia del procesamiento del lenguaje natural en la productividad de los

	expedientes judiciales en un distrito judicial Lima Norte, 2023
--	---

4. Soporte teórico:

Variable 1: Procesamiento del Lenguaje Natural

Escala/ÁREA	Subescala (dimensiones)	Definición
Procesamiento del Lenguaje Natural	Preprocesamiento	Etapa crítica en el flujo de trabajo de NLP que involucra la preparación y limpieza de los datos de texto antes de que se puedan aplicar técnicas de análisis o modelado de lenguaje natural. El preprocesamiento es esencial para garantizar que los datos de texto estén en un formato adecuado y sean susceptibles de ser analizados de manera efectiva (Perla et al., 2022).
	Extracción de datos	Se enfoca en la obtención de información relevante y estructurada a partir de los textos procesados. En esta etapa, se busca identificar patrones, características o elementos de interés en los datos textuales para su posterior análisis o aplicación en diversas tareas de NLP (Miranda y Guzman, 2014).

Variable 2: Productividad

Escala/ÁREA	Subescala (dimensiones)	Definición
Productividad	Tiempo de registro	La dimensión de tiempo de registro de productividad se refiere a la recopilación y el seguimiento del tiempo dedicado a tareas específicas con el fin de evaluar y mejorar la productividad en un entorno laboral o personal. Esta dimensión es crucial para comprender cómo se utiliza el tiempo, identificar áreas de mejora y optimizar la eficiencia en la realización de tareas y proyectos (Herrera y Madrid, 2022).

Evaluación por juicio de expertos

Respetado juez: Usted ha sido seleccionado para evaluar el instrumento para medir las Variables Procesamiento del Lenguaje Natural y Productividad. La evaluación del instrumento es de gran relevancia para lograr que sea válido y que los resultados obtenidos a partir de éste sean utilizados eficientemente; aportando al quehacer psicológico. Agradecemos su valiosa colaboración.

5. Datos generales del juez:

Nombre del juez:	David Flores Zafra		
Grado profesional:	Maestría ()	Doctor	(x)
Área de formación académica:	Clinica ()	Social	()
	Educativa ()	Organizacional	(x)
Áreas de experiencia profesional:	Educación		
Institución donde labora:	Universidad Cesar Vallejo		
Tiempo de experiencia profesional en el área:	2 a 4 años ()	Más de 5 años	(x)
Experiencia en Investigación (si corresponde)	Asesora de Tesis en Pregrado y Posgrado		

6. Propósito de la evaluación:

Validar el contenido del instrumento, por juicio de expertos.

7. Datos de la escala:

Nombre de la Prueba:	Ficha de observación
Autor:	Enok David Lucero Carrillo
Procedencia:	Lima
Administración:	Directa, virtual
Tiempo de aplicación:	25 minutos
Ámbito de aplicación:	Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los integrantes del grupo familiar de Lima Norte
Significación:	Esta guía consta de 2 fichas de observación. El objetivo es determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023

8. Soporte teórico:

Variable 1: Procesamiento del Lenguaje Natural

Escala/ÁREA	Subescala (dimensiones)	Definición
Procesamiento del Lenguaje Natural	Preprocesamiento	Etapa crítica en el flujo de trabajo de NLP que involucra la preparación y limpieza de los datos de texto antes de que se puedan aplicar técnicas de análisis o modelado de lenguaje natural. El preprocesamiento es esencial para garantizar que los datos de texto estén en un formato adecuado y sean susceptibles de ser analizados de manera efectiva (Pérez et al., 2022).
	Extracción de datos	Se enfoca en la obtención de información relevante y estructurada a partir de los textos procesados. En esta etapa, se busca identificar patrones, características o elementos de interés en los datos textuales para su posterior análisis o aplicación en diversas tareas de NLP (Miranda y Guzmán, 2014).

Variable 2: Productividad

Escala/ÁREA	Subescala (dimensiones)	Definición
Productividad	Tiempo de registro	La dimensión de tiempo de registro de productividad se refiere a la recopilación y el seguimiento del tiempo dedicado a tareas específicas con el fin de evaluar y mejorar la productividad en un entorno laboral o personal. Esta dimensión es crucial para comprender cómo se utiliza el tiempo, identificar áreas de mejora y optimizar la eficiencia en la realización de tareas y proyectos (Herrera y Madrid, 2022).

Presentación de instrucciones para el juez:

A continuación, a usted le presento el cuestionario: "Ficha de observación para determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023" elaborado por Marlon Frank Acuña Benites en el año 2023, de acuerdo con los siguientes indicadores califique cada uno de los ítems según corresponda.

Categoría	Calificación	Indicador
CLARIDAD El ítem se comprende fácilmente, es decir, su sintáctica y semántica son adecuadas.	1. No cumple con el criterio	El ítem no es claro.
	2. Bajo Nivel	El ítem requiere bastantes modificaciones o una modificación muy grande en el uso de las palabras de acuerdo con su significado o por la ordenación de estas.
	3. Moderado nivel	Se requiere una modificación muy específica de algunos de los términos del ítem.
	4. Alto nivel	El ítem es claro, tiene semántica y sintaxis adecuada.
COHERENCIA El ítem tiene relación lógica con la dimensión o indicador que está midiendo.	1. totalmente en desacuerdo (no cumple con el criterio)	El ítem no tiene relación lógica con la dimensión.
	2. Desacuerdo (bajo nivel de acuerdo)	El ítem tiene una relación tangencial lejana con la dimensión.
	3. Acuerdo (moderado nivel)	El ítem tiene una relación moderada con la dimensión que se está midiendo.
	4. Totalmente de Acuerdo (alto nivel)	El ítem se encuentra está relacionado con la dimensión que está midiendo.
RELEVANCIA El ítem es esencial o importante, es decir debe ser incluido.	1. No cumple con el criterio	El ítem puede ser eliminado sin que se vea afectada la medición de la dimensión.
	2. Bajo Nivel	El ítem tiene alguna relevancia, pero otro ítem puede estar incluyendo lo que mide éste.
	3. Moderado nivel	El ítem es relativamente importante.
	4. Alto nivel	El ítem es muy relevante y debe ser incluido.

Leer con detenimiento los ítems y calificar en una escala de 1 a 4 su valoración, así como solicitamos brinde sus observaciones que considere pertinente

1. No cumple con el criterio
2. Bajo Nivel
3. Moderado nivel

4. Alto nivel

Dimensiones del instrumento: Ficha de observación para determinar la influencia del procesamiento del lenguaje natural en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023

Primera dimensión: Preprocesamiento

- **Objetivos de la Dimensión:** Analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Guía de observación N°01

N° de Expediente	Limpieza de texto	Tokenización	Eliminación de palabras vacías
1			
2			
3			
4			
5			
6			
7			
8			

378			

Nota. Obtenido de Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los Integrantes del grupo familiar de Lima Norte

Indicadores	Ítem	Claridad	Coherencia	Relevancia	Observaciones/ Recomendaciones
Limpieza de texto	Guía de observación N°1	4	4	4	
Tokenización	Guía de observación N°1	4	4	4	
Eliminación de palabras vacías	Guía de observación N°1	4	4	4	

Segunda dimensión: Extracción de datos

- **Objetivos de la Dimensión:** Analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Guía de observación N°02

N° de Expediente	Etiquetado de texto	Extracción de palabras claves
1		
2		
3		
4		
5		
6		
7		
8		
.....		
378		

Nota. Obtenido de Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los Integrantes del grupo familiar de Lima Norte

Indicadores	Ítem	Claridad	Coherencia	Relevancia	Observaciones/ Recomendaciones
Etiquetado de texto	Guía de observación N°2	4	4	4	
Extracción de palabras claves	Guía de observación N°2	4	4	4	

Tercera dimensión: Tiempo de registro

- Objetivos de la Dimensión: Analizar la influencia del preprocesamiento en la productividad de los expedientes judiciales en un distrito judicial Lima Norte, 2023.

Guía de observación N°03

N° de Expediente	Datos de entrada
1	
2	
3	
4	
5	
6	
7	
8	
.....	
378	

Nota. Obtenido de Juzgados Especializados de la Sub Especialidad de Violencia contra la mujer y los Integrantes del grupo familiar de Lima Norte

Indicadores	Ítem	Claridad	Coherencia	Relevancia	Observaciones/ Recomendaciones
Datos de entrada	Guía de observación N°2	4	4	4	

Observaciones (precisar si hay suficiencia): El instrumento presenta suficiencia _____

Opinión de aplicabilidad: Aplicable | No aplicable | Aplicable después de corregir | No aplicable

Apellidos y nombres del juez validador: David Flores Zafra

Especialidad del validador: Docente.....

- ***Pertinencia:** El ítem corresponde al concepto teórico formulado.
- ***Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo
- ***Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

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

18 de diciembre del 2023.



Firma del Experto validador

Anexo 05

Ficha de recolección de datos

N°	EXPEDIENTE	Procesamiento del Lenguaje Natural							Productividad	
		Preprocesamiento				Extracción de datos			Tiempo de Registro	
		Limpieza de texto	Tokenización	Eliminación de palabras vacías	Tiempo Total Procesamiento	Etiquetado de texto	Extracción de palabras claves	Tiempo de ejecución de extracción de datos	Datos de salida	Produ. General = (Prep + Extrac)/ Datos salida
1	EXPEDIENTE 1	1	4	2	7	1	4	5	8	1,50
2	EXPEDIENTE 2	2	5	3	10	2	5	7	10	1,70
3	EXPEDIENTE 3	2	5	3	10	2	5	7	10	1,70
4	EXPEDIENTE 4	2	5	3	10	2	5	7	10	1,70
5	EXPEDIENTE 5	2	5	3	10	1	4	5	9	1,67
6	EXPEDIENTE 6	1	4	2	7	1	4	5	8	1,50
7	EXPEDIENTE 7	1	4	2	7	2	5	7	9	1,56
8	EXPEDIENTE 8	1	4	2	7	1	4	5	8	1,50
9	EXPEDIENTE 9	1	4	2	7	1	4	5	8	1,50
10	EXPEDIENTE 10	1	4	2	7	2	5	7	9	1,56
11	EXPEDIENTE 11	2	5	3	10	1	4	5	9	1,67
12	EXPEDIENTE 12	2	5	3	10	2	5	7	10	1,70
13	EXPEDIENTE 13	1	4	2	7	1	4	5	8	1,50
14	EXPEDIENTE 14	2	5	3	10	1	4	5	9	1,67
15	EXPEDIENTE 15	1	4	2	7	1	4	5	8	1,50
16	EXPEDIENTE 16	2	5	3	10	2	5	7	10	1,70

17	EXPEDIENTE 17	2	5	3	10	1	4	5	9	1,67
18	EXPEDIENTE 18	1	4	2	7	2	5	7	9	1,56
19	EXPEDIENTE 19	2	5	3	10	1	4	5	9	1,67
20	EXPEDIENTE 20	2	5	3	10	2	5	7	10	1,70
21	EXPEDIENTE 21	2	5	3	10	2	5	7	10	1,70
22	EXPEDIENTE 22	2	5	3	10	2	5	7	10	1,70
23	EXPEDIENTE 23	1	4	2	7	2	5	7	9	1,56
24	EXPEDIENTE 24	2	5	3	10	2	5	7	10	1,70
25	EXPEDIENTE 25	2	5	3	10	1	4	5	9	1,67
26	EXPEDIENTE 26	2	5	3	10	2	5	7	10	1,70
27	EXPEDIENTE 27	2	5	3	10	2	5	7	10	1,70
28	EXPEDIENTE 28	2	5	3	10	1	4	5	9	1,67
29	EXPEDIENTE 29	1	4	2	7	2	5	7	9	1,56
30	EXPEDIENTE 30	2	5	3	10	2	5	7	10	1,70
31	EXPEDIENTE 31	2	5	3	10	1	4	5	9	1,67
32	EXPEDIENTE 32	1	4	2	7	1	4	5	8	1,50
33	EXPEDIENTE 33	2	5	3	10	2	5	7	10	1,70
34	EXPEDIENTE 34	2	5	3	10	2	5	7	10	1,70
35	EXPEDIENTE 35	1	4	2	7	1	4	5	8	1,50
36	EXPEDIENTE 36	1	4	2	7	2	5	7	9	1,56
37	EXPEDIENTE 37	1	4	2	7	1	4	5	8	1,50
38	EXPEDIENTE 38	1	4	2	7	2	5	7	9	1,56
39	EXPEDIENTE 39	2	5	3	10	1	4	5	9	1,67
40	EXPEDIENTE 40	1	4	2	7	2	5	7	9	1,56
41	EXPEDIENTE 41	1	4	2	7	1	4	5	8	1,50

42	EXPEDIENTE 42	1	4	2	7	2	5	7	9	1,56
43	EXPEDIENTE 43	2	5	3	10	2	5	7	10	1,70
44	EXPEDIENTE 44	1	4	2	7	1	4	5	8	1,50
45	EXPEDIENTE 45	1	4	2	7	1	4	5	8	1,50
46	EXPEDIENTE 46	1	4	2	7	1	4	5	8	1,50
47	EXPEDIENTE 47	1	4	2	7	1	4	5	8	1,50
48	EXPEDIENTE 48	2	5	3	10	1	4	5	9	1,67
49	EXPEDIENTE 49	1	4	2	7	2	5	7	9	1,56
50	EXPEDIENTE 50	2	5	3	10	1	4	5	9	1,67
51	EXPEDIENTE 51	1	4	2	7	1	4	5	8	1,50
52	EXPEDIENTE 52	2	5	3	10	2	5	7	10	1,70
53	EXPEDIENTE 53	2	5	3	10	1	4	5	9	1,67
54	EXPEDIENTE 54	1	4	2	7	2	5	7	9	1,56
55	EXPEDIENTE 55	2	5	3	10	2	5	7	10	1,70
56	EXPEDIENTE 56	2	5	3	10	2	5	7	10	1,70
57	EXPEDIENTE 57	2	5	3	10	2	5	7	10	1,70
58	EXPEDIENTE 58	1	4	2	7	2	5	7	9	1,56
59	EXPEDIENTE 59	2	5	3	10	1	4	5	9	1,67
60	EXPEDIENTE 60	1	4	2	7	1	4	5	8	1,50
61	EXPEDIENTE 61	1	4	2	7	2	5	7	9	1,56
62	EXPEDIENTE 62	1	4	2	7	1	4	5	8	1,50
63	EXPEDIENTE 63	1	4	2	7	1	4	5	8	1,50
64	EXPEDIENTE 64	2	5	3	10	2	5	7	10	1,70
65	EXPEDIENTE 65	1	4	2	7	1	4	5	8	1,50
66	EXPEDIENTE 66	2	5	3	10	1	4	5	9	1,67

67	EXPEDIENTE 67	1	4	2	7	1	4	5	8	1,50
68	EXPEDIENTE 68	1	4	2	7	2	5	7	9	1,56
69	EXPEDIENTE 69	1	4	2	7	2	5	7	9	1,56
70	EXPEDIENTE 70	2	5	3	10	1	4	5	9	1,67
71	EXPEDIENTE 71	2	5	3	10	2	5	7	10	1,70
72	EXPEDIENTE 72	2	5	3	10	2	5	7	10	1,70
73	EXPEDIENTE 73	1	4	2	7	2	5	7	9	1,56
74	EXPEDIENTE 74	1	4	2	7	2	5	7	9	1,56
75	EXPEDIENTE 75	1	4	2	7	1	4	5	8	1,50
76	EXPEDIENTE 76	2	5	3	10	2	5	7	10	1,70
77	EXPEDIENTE 77	2	5	3	10	2	5	7	10	1,70
78	EXPEDIENTE 78	1	4	2	7	1	4	5	8	1,50
79	EXPEDIENTE 79	1	4	2	7	1	4	5	8	1,50
80	EXPEDIENTE 80	2	5	3	10	2	5	7	10	1,70
81	EXPEDIENTE 81	2	5	3	10	2	5	7	10	1,70
82	EXPEDIENTE 82	2	5	3	10	1	4	5	9	1,67
83	EXPEDIENTE 83	1	4	2	7	1	4	5	8	1,50
84	EXPEDIENTE 84	1	4	2	7	1	4	5	8	1,50
85	EXPEDIENTE 85	1	4	2	7	2	5	7	9	1,56
86	EXPEDIENTE 86	2	5	3	10	1	4	5	9	1,67
87	EXPEDIENTE 87	1	4	2	7	2	5	7	9	1,56
88	EXPEDIENTE 88	2	5	3	10	2	5	7	10	1,70
89	EXPEDIENTE 89	2	5	3	10	1	4	5	9	1,67
90	EXPEDIENTE 90	2	5	3	10	2	5	7	10	1,70
91	EXPEDIENTE 91	2	5	3	10	2	5	7	10	1,70

92	EXPEDIENTE 92	1	4	2	7	2	5	7	9	1,56
93	EXPEDIENTE 93	1	4	2	7	2	5	7	9	1,56
94	EXPEDIENTE 94	1	4	2	7	2	5	7	9	1,56
95	EXPEDIENTE 95	1	4	2	7	2	5	7	9	1,56
96	EXPEDIENTE 96	1	4	2	7	2	5	7	9	1,56
97	EXPEDIENTE 97	1	4	2	7	2	5	7	9	1,56
98	EXPEDIENTE 98	2	5	3	10	2	5	7	10	1,70
99	EXPEDIENTE 99	1	4	2	7	1	4	5	8	1,50
100	EXPEDIENTE 100	2	5	3	10	1	4	5	9	1,67
101	EXPEDIENTE 101	2	5	3	10	1	4	5	9	1,67
102	EXPEDIENTE 102	2	5	3	10	1	4	5	9	1,67
103	EXPEDIENTE 103	1	4	2	7	2	5	7	9	1,56
104	EXPEDIENTE 104	1	4	2	7	2	5	7	9	1,56
105	EXPEDIENTE 105	1	4	2	7	1	4	5	8	1,50
106	EXPEDIENTE 106	2	5	3	10	1	4	5	9	1,67
107	EXPEDIENTE 107	1	4	2	7	1	4	5	8	1,50
108	EXPEDIENTE 108	2	5	3	10	2	5	7	10	1,70
109	EXPEDIENTE 109	1	4	2	7	1	4	5	8	1,50
110	EXPEDIENTE 110	1	4	2	7	2	5	7	9	1,56
111	EXPEDIENTE 111	1	4	2	7	2	5	7	9	1,56
112	EXPEDIENTE 112	2	5	3	10	2	5	7	10	1,70
113	EXPEDIENTE 113	2	5	3	10	1	4	5	9	1,67
114	EXPEDIENTE 114	1	4	2	7	2	5	7	9	1,56
115	EXPEDIENTE 115	2	5	3	10	2	5	7	10	1,70
116	EXPEDIENTE 116	1	4	2	7	1	4	5	8	1,50

117	EXPEDIENTE 117	1	4	2	7	2	5	7	9	1,56
118	EXPEDIENTE 118	2	5	3	10	2	5	7	10	1,70
119	EXPEDIENTE 119	2	5	3	10	1	4	5	9	1,67
120	EXPEDIENTE 120	2	5	3	10	2	5	7	10	1,70
121	EXPEDIENTE 121	1	4	2	7	2	5	7	9	1,56
122	EXPEDIENTE 122	1	4	2	7	1	4	5	8	1,50
123	EXPEDIENTE 123	2	5	3	10	1	4	5	9	1,67
124	EXPEDIENTE 124	2	5	3	10	1	4	5	9	1,67
125	EXPEDIENTE 125	2	5	3	10	2	5	7	10	1,70
126	EXPEDIENTE 126	1	4	2	7	2	5	7	9	1,56
127	EXPEDIENTE 127	1	4	2	7	1	4	5	8	1,50
128	EXPEDIENTE 128	2	5	3	10	2	5	7	10	1,70
129	EXPEDIENTE 129	1	4	2	7	1	4	5	8	1,50
130	EXPEDIENTE 130	2	5	3	10	2	5	7	10	1,70
131	EXPEDIENTE 131	1	4	2	7	2	5	7	9	1,56
132	EXPEDIENTE 132	1	4	2	7	2	5	7	9	1,56
133	EXPEDIENTE 133	2	5	3	10	2	5	7	10	1,70
134	EXPEDIENTE 134	1	4	2	7	1	4	5	8	1,50
135	EXPEDIENTE 135	2	5	3	10	1	4	5	9	1,67
136	EXPEDIENTE 136	2	5	3	10	1	4	5	9	1,67
137	EXPEDIENTE 137	2	5	3	10	2	5	7	10	1,70
138	EXPEDIENTE 138	1	4	2	7	1	4	5	8	1,50
139	EXPEDIENTE 139	1	4	2	7	1	4	5	8	1,50
140	EXPEDIENTE 140	2	5	3	10	1	4	5	9	1,67
141	EXPEDIENTE 141	1	4	2	7	2	5	7	9	1,56

142	EXPEDIENTE 142	1	4	2	7	2	5	7	9	1,56
143	EXPEDIENTE 143	1	4	2	7	2	5	7	9	1,56
144	EXPEDIENTE 144	2	5	3	10	2	5	7	10	1,70
145	EXPEDIENTE 145	2	5	3	10	1	4	5	9	1,67
146	EXPEDIENTE 146	2	5	3	10	2	5	7	10	1,70
147	EXPEDIENTE 147	2	5	3	10	1	4	5	9	1,67
148	EXPEDIENTE 148	2	5	3	10	1	4	5	9	1,67
149	EXPEDIENTE 149	2	5	3	10	1	4	5	9	1,67
150	EXPEDIENTE 150	2	5	3	10	2	5	7	10	1,70
151	EXPEDIENTE 151	2	5	3	10	1	4	5	9	1,67
152	EXPEDIENTE 152	2	5	3	10	1	4	5	9	1,67
153	EXPEDIENTE 153	1	4	2	7	1	4	5	8	1,50
154	EXPEDIENTE 154	2	5	3	10	1	4	5	9	1,67
155	EXPEDIENTE 155	2	5	3	10	2	5	7	10	1,70
156	EXPEDIENTE 156	2	5	3	10	2	5	7	10	1,70
157	EXPEDIENTE 157	2	5	3	10	1	4	5	9	1,67
158	EXPEDIENTE 158	2	5	3	10	2	5	7	10	1,70
159	EXPEDIENTE 159	1	4	2	7	2	5	7	9	1,56
160	EXPEDIENTE 160	2	5	3	10	1	4	5	9	1,67
161	EXPEDIENTE 161	1	4	2	7	2	5	7	9	1,56
162	EXPEDIENTE 162	1	4	2	7	2	5	7	9	1,56
163	EXPEDIENTE 163	1	4	2	7	2	5	7	9	1,56
164	EXPEDIENTE 164	1	4	2	7	1	4	5	8	1,50
165	EXPEDIENTE 165	2	5	3	10	2	5	7	10	1,70
166	EXPEDIENTE 166	1	4	2	7	2	5	7	9	1,56

167	EXPEDIENTE 167	1	4	2	7	2	5	7	9	1,56
168	EXPEDIENTE 168	2	5	3	10	2	5	7	10	1,70
169	EXPEDIENTE 169	2	5	3	10	1	4	5	9	1,67
170	EXPEDIENTE 170	2	5	3	10	2	5	7	10	1,70
171	EXPEDIENTE 171	2	5	3	10	1	4	5	9	1,67
172	EXPEDIENTE 172	2	5	3	10	1	4	5	9	1,67
173	EXPEDIENTE 173	2	5	3	10	2	5	7	10	1,70
174	EXPEDIENTE 174	2	5	3	10	2	5	7	10	1,70
175	EXPEDIENTE 175	2	5	3	10	1	4	5	9	1,67
176	EXPEDIENTE 176	1	4	2	7	2	5	7	9	1,56
177	EXPEDIENTE 177	2	5	3	10	1	4	5	9	1,67
178	EXPEDIENTE 178	2	5	3	10	1	4	5	9	1,67
179	EXPEDIENTE 179	2	5	3	10	2	5	7	10	1,70
180	EXPEDIENTE 180	1	4	2	7	2	5	7	9	1,56
181	EXPEDIENTE 181	2	5	3	10	2	5	7	10	1,70
182	EXPEDIENTE 182	1	4	2	7	1	4	5	8	1,50
183	EXPEDIENTE 183	1	4	2	7	2	5	7	9	1,56
184	EXPEDIENTE 184	2	5	3	10	2	5	7	10	1,70
185	EXPEDIENTE 185	1	4	2	7	2	5	7	9	1,56
186	EXPEDIENTE 186	2	5	3	10	2	5	7	10	1,70
187	EXPEDIENTE 187	2	5	3	10	1	4	5	9	1,67
188	EXPEDIENTE 188	2	5	3	10	1	4	5	9	1,67
189	EXPEDIENTE 189	2	5	3	10	1	4	5	9	1,67
190	EXPEDIENTE 190	1	4	2	7	1	4	5	8	1,50
191	EXPEDIENTE 191	2	5	3	10	1	4	5	9	1,67

192	EXPEDIENTE 192	2	5	3	10	1	4	5	9	1,67
193	EXPEDIENTE 193	2	5	3	10	2	5	7	10	1,70
194	EXPEDIENTE 194	1	4	2	7	2	5	7	9	1,56
195	EXPEDIENTE 195	2	5	3	10	2	5	7	10	1,70
196	EXPEDIENTE 196	2	5	3	10	2	5	7	10	1,70
197	EXPEDIENTE 197	1	4	2	7	1	4	5	8	1,50
198	EXPEDIENTE 198	2	5	3	10	2	5	7	10	1,70
199	EXPEDIENTE 199	2	5	3	10	1	4	5	9	1,67
200	EXPEDIENTE 200	2	5	3	10	1	4	5	9	1,67
201	EXPEDIENTE 201	2	5	3	10	1	4	5	9	1,67
202	EXPEDIENTE 202	2	5	3	10	2	5	7	10	1,70
203	EXPEDIENTE 203	2	5	3	10	1	4	5	9	1,67
204	EXPEDIENTE 204	2	5	3	10	2	5	7	10	1,70
205	EXPEDIENTE 205	2	5	3	10	1	4	5	9	1,67
206	EXPEDIENTE 206	2	5	3	10	2	5	7	10	1,70
207	EXPEDIENTE 207	1	4	2	7	1	4	5	8	1,50
208	EXPEDIENTE 208	2	5	3	10	2	5	7	10	1,70
209	EXPEDIENTE 209	2	5	3	10	2	5	7	10	1,70
210	EXPEDIENTE 210	1	4	2	7	2	5	7	9	1,56
211	EXPEDIENTE 211	2	5	3	10	2	5	7	10	1,70
212	EXPEDIENTE 212	1	4	2	7	2	5	7	9	1,56
213	EXPEDIENTE 213	2	5	3	10	1	4	5	9	1,67
214	EXPEDIENTE 214	2	5	3	10	2	5	7	10	1,70
215	EXPEDIENTE 215	1	4	2	7	2	5	7	9	1,56
216	EXPEDIENTE 216	1	4	2	7	2	5	7	9	1,56

217	EXPEDIENTE 217	1	4	2	7	2	5	7	9	1,56
218	EXPEDIENTE 218	2	5	3	10	2	5	7	10	1,70
219	EXPEDIENTE 219	2	5	3	10	1	4	5	9	1,67
220	EXPEDIENTE 220	2	5	3	10	1	4	5	9	1,67
221	EXPEDIENTE 221	1	4	2	7	1	4	5	8	1,50
222	EXPEDIENTE 222	1	4	2	7	2	5	7	9	1,56
223	EXPEDIENTE 223	2	5	3	10	2	5	7	10	1,70
224	EXPEDIENTE 224	1	4	2	7	2	5	7	9	1,56
225	EXPEDIENTE 225	1	4	2	7	2	5	7	9	1,56
226	EXPEDIENTE 226	1	4	2	7	2	5	7	9	1,56
227	EXPEDIENTE 227	2	5	3	10	1	4	5	9	1,67
228	EXPEDIENTE 228	1	4	2	7	2	5	7	9	1,56
229	EXPEDIENTE 229	2	5	3	10	2	5	7	10	1,70
230	EXPEDIENTE 230	2	5	3	10	1	4	5	9	1,67
231	EXPEDIENTE 231	1	4	2	7	2	5	7	9	1,56
232	EXPEDIENTE 232	2	5	3	10	1	4	5	9	1,67
233	EXPEDIENTE 233	2	5	3	10	2	5	7	10	1,70
234	EXPEDIENTE 234	2	5	3	10	2	5	7	10	1,70
235	EXPEDIENTE 235	2	5	3	10	1	4	5	9	1,67
236	EXPEDIENTE 236	1	4	2	7	1	4	5	8	1,50
237	EXPEDIENTE 237	1	4	2	7	1	4	5	8	1,50
238	EXPEDIENTE 238	1	4	2	7	2	5	7	9	1,56
239	EXPEDIENTE 239	2	5	3	10	2	5	7	10	1,70
240	EXPEDIENTE 240	2	5	3	10	1	4	5	9	1,67
241	EXPEDIENTE 241	1	4	2	7	2	5	7	9	1,56

242	EXPEDIENTE 242	2	5	3	10	1	4	5	9	1,67
243	EXPEDIENTE 243	1	4	2	7	1	4	5	8	1,50
244	EXPEDIENTE 244	2	5	3	10	1	4	5	9	1,67
245	EXPEDIENTE 245	1	4	2	7	2	5	7	9	1,56
246	EXPEDIENTE 246	2	5	3	10	1	4	5	9	1,67
247	EXPEDIENTE 247	1	4	2	7	2	5	7	9	1,56
248	EXPEDIENTE 248	2	5	3	10	1	4	5	9	1,67
249	EXPEDIENTE 249	1	4	2	7	2	5	7	9	1,56
250	EXPEDIENTE 250	2	5	3	10	1	4	5	9	1,67
251	EXPEDIENTE 251	2	5	3	10	1	4	5	9	1,67
252	EXPEDIENTE 252	1	4	2	7	1	4	5	8	1,50
253	EXPEDIENTE 253	2	5	3	10	1	4	5	9	1,67
254	EXPEDIENTE 254	1	4	2	7	2	5	7	9	1,56
255	EXPEDIENTE 255	2	5	3	10	1	4	5	9	1,67
256	EXPEDIENTE 256	2	5	3	10	1	4	5	9	1,67
257	EXPEDIENTE 257	1	4	2	7	1	4	5	8	1,50
258	EXPEDIENTE 258	1	4	2	7	1	4	5	8	1,50
259	EXPEDIENTE 259	2	5	3	10	1	4	5	9	1,67
260	EXPEDIENTE 260	1	4	2	7	1	4	5	8	1,50
261	EXPEDIENTE 261	1	4	2	7	2	5	7	9	1,56
262	EXPEDIENTE 262	2	5	3	10	1	4	5	9	1,67
263	EXPEDIENTE 263	2	5	3	10	1	4	5	9	1,67
264	EXPEDIENTE 264	1	4	2	7	2	5	7	9	1,56
265	EXPEDIENTE 265	2	5	3	10	1	4	5	9	1,67
266	EXPEDIENTE 266	1	4	2	7	2	5	7	9	1,56

267	EXPEDIENTE 267	1	4	2	7	1	4	5	8	1,50
268	EXPEDIENTE 268	1	4	2	7	2	5	7	9	1,56
269	EXPEDIENTE 269	1	4	2	7	2	5	7	9	1,56
270	EXPEDIENTE 270	1	4	2	7	2	5	7	9	1,56
271	EXPEDIENTE 271	2	5	3	10	1	4	5	9	1,67
272	EXPEDIENTE 272	1	4	2	7	2	5	7	9	1,56
273	EXPEDIENTE 273	1	4	2	7	2	5	7	9	1,56
274	EXPEDIENTE 274	2	5	3	10	2	5	7	10	1,70
275	EXPEDIENTE 275	2	5	3	10	2	5	7	10	1,70
276	EXPEDIENTE 276	2	5	3	10	1	4	5	9	1,67
277	EXPEDIENTE 277	1	4	2	7	2	5	7	9	1,56
278	EXPEDIENTE 278	1	4	2	7	1	4	5	8	1,50
279	EXPEDIENTE 279	2	5	3	10	1	4	5	9	1,67
280	EXPEDIENTE 280	2	5	3	10	1	4	5	9	1,67
281	EXPEDIENTE 281	2	5	3	10	1	4	5	9	1,67
282	EXPEDIENTE 282	1	4	2	7	2	5	7	9	1,56
283	EXPEDIENTE 283	1	4	2	7	2	5	7	9	1,56
284	EXPEDIENTE 284	1	4	2	7	1	4	5	8	1,50
285	EXPEDIENTE 285	2	5	3	10	1	4	5	9	1,67
286	EXPEDIENTE 286	2	5	3	10	2	5	7	10	1,70
287	EXPEDIENTE 287	1	4	2	7	1	4	5	8	1,50
288	EXPEDIENTE 288	2	5	3	10	1	4	5	9	1,67
289	EXPEDIENTE 289	1	4	2	7	2	5	7	9	1,56
290	EXPEDIENTE 290	1	4	2	7	1	4	5	8	1,50
291	EXPEDIENTE 291	1	4	2	7	2	5	7	9	1,56

292	EXPEDIENTE 292	1	4	2	7	2	5	7	9	1,56
293	EXPEDIENTE 293	1	4	2	7	2	5	7	9	1,56
294	EXPEDIENTE 294	1	4	2	7	2	5	7	9	1,56
295	EXPEDIENTE 295	1	4	2	7	1	4	5	8	1,50
296	EXPEDIENTE 296	2	5	3	10	1	4	5	9	1,67
297	EXPEDIENTE 297	1	4	2	7	2	5	7	9	1,56
298	EXPEDIENTE 298	2	5	3	10	1	4	5	9	1,67
299	EXPEDIENTE 299	2	5	3	10	1	4	5	9	1,67
300	EXPEDIENTE 300	1	4	2	7	2	5	7	9	1,56
301	EXPEDIENTE 301	1	4	2	7	1	4	5	8	1,50
302	EXPEDIENTE 302	1	4	2	7	2	5	7	9	1,56
303	EXPEDIENTE 303	1	4	2	7	1	4	5	8	1,50
304	EXPEDIENTE 304	2	5	3	10	2	5	7	10	1,70
305	EXPEDIENTE 305	1	4	2	7	1	4	5	8	1,50
306	EXPEDIENTE 306	1	4	2	7	1	4	5	8	1,50
307	EXPEDIENTE 307	2	5	3	10	2	5	7	10	1,70
308	EXPEDIENTE 308	2	5	3	10	1	4	5	9	1,67
309	EXPEDIENTE 309	1	4	2	7	2	5	7	9	1,56
310	EXPEDIENTE 310	2	5	3	10	1	4	5	9	1,67
311	EXPEDIENTE 311	1	4	2	7	2	5	7	9	1,56
312	EXPEDIENTE 312	2	5	3	10	2	5	7	10	1,70
313	EXPEDIENTE 313	1	4	2	7	2	5	7	9	1,56
314	EXPEDIENTE 314	2	5	3	10	1	4	5	9	1,67
315	EXPEDIENTE 315	2	5	3	10	2	5	7	10	1,70
316	EXPEDIENTE 316	2	5	3	10	1	4	5	9	1,67

317	EXPEDIENTE 317	1	4	2	7	2	5	7	9	1,56
318	EXPEDIENTE 318	2	5	3	10	2	5	7	10	1,70
319	EXPEDIENTE 319	2	5	3	10	2	5	7	10	1,70
320	EXPEDIENTE 320	2	5	3	10	2	5	7	10	1,70
321	EXPEDIENTE 321	2	5	3	10	1	4	5	9	1,67
322	EXPEDIENTE 322	1	4	2	7	1	4	5	8	1,50
323	EXPEDIENTE 323	1	4	2	7	2	5	7	9	1,56
324	EXPEDIENTE 324	1	4	2	7	1	4	5	8	1,50
325	EXPEDIENTE 325	1	4	2	7	1	4	5	8	1,50
326	EXPEDIENTE 326	2	5	3	10	1	4	5	9	1,67
327	EXPEDIENTE 327	2	5	3	10	2	5	7	10	1,70
328	EXPEDIENTE 328	2	5	3	10	2	5	7	10	1,70
329	EXPEDIENTE 329	2	5	3	10	2	5	7	10	1,70
330	EXPEDIENTE 330	2	5	3	10	1	4	5	9	1,67
331	EXPEDIENTE 331	1	4	2	7	2	5	7	9	1,56
332	EXPEDIENTE 332	1	4	2	7	1	4	5	8	1,50
333	EXPEDIENTE 333	2	5	3	10	2	5	7	10	1,70
334	EXPEDIENTE 334	1	4	2	7	2	5	7	9	1,56
335	EXPEDIENTE 335	2	5	3	10	2	5	7	10	1,70
336	EXPEDIENTE 336	2	5	3	10	1	4	5	9	1,67
337	EXPEDIENTE 337	2	5	3	10	2	5	7	10	1,70
338	EXPEDIENTE 338	2	5	3	10	1	4	5	9	1,67
339	EXPEDIENTE 339	2	5	3	10	1	4	5	9	1,67
340	EXPEDIENTE 340	2	5	3	10	1	4	5	9	1,67
341	EXPEDIENTE 341	2	5	3	10	2	5	7	10	1,70

342	EXPEDIENTE 342	1	4	2	7	2	5	7	9	1,56
343	EXPEDIENTE 343	2	5	3	10	1	4	5	9	1,67
344	EXPEDIENTE 344	2	5	3	10	2	5	7	10	1,70
345	EXPEDIENTE 345	1	4	2	7	1	4	5	8	1,50
346	EXPEDIENTE 346	2	5	3	10	2	5	7	10	1,70
347	EXPEDIENTE 347	2	5	3	10	2	5	7	10	1,70
348	EXPEDIENTE 348	2	5	3	10	2	5	7	10	1,70
349	EXPEDIENTE 349	2	5	3	10	1	4	5	9	1,67
350	EXPEDIENTE 350	1	4	2	7	2	5	7	9	1,56
351	EXPEDIENTE 351	2	5	3	10	1	4	5	9	1,67
352	EXPEDIENTE 352	2	5	3	10	1	4	5	9	1,67
353	EXPEDIENTE 353	2	5	3	10	2	5	7	10	1,70
354	EXPEDIENTE 354	2	5	3	10	2	5	7	10	1,70
355	EXPEDIENTE 355	1	4	2	7	1	4	5	8	1,50
356	EXPEDIENTE 356	2	5	3	10	2	5	7	10	1,70
357	EXPEDIENTE 357	1	4	2	7	2	5	7	9	1,56
358	EXPEDIENTE 358	1	4	2	7	2	5	7	9	1,56
359	EXPEDIENTE 359	1	4	2	7	2	5	7	9	1,56
360	EXPEDIENTE 360	2	5	3	10	1	4	5	9	1,67
361	EXPEDIENTE 361	1	4	2	7	2	5	7	9	1,56
362	EXPEDIENTE 362	2	5	3	10	2	5	7	10	1,70
363	EXPEDIENTE 363	1	4	2	7	2	5	7	9	1,56
364	EXPEDIENTE 364	1	4	2	7	2	5	7	9	1,56
365	EXPEDIENTE 365	2	5	3	10	2	5	7	10	1,70
366	EXPEDIENTE 366	1	4	2	7	2	5	7	9	1,56

367	EXPEDIENTE 367	1	4	2	7	2	5	7	9	1,56
368	EXPEDIENTE 368	1	4	2	7	1	4	5	8	1,50
369	EXPEDIENTE 369	1	4	2	7	2	5	7	9	1,56
370	EXPEDIENTE 370	1	4	2	7	1	4	5	8	1,50
371	EXPEDIENTE 371	1	4	2	7	2	5	7	9	1,56
372	EXPEDIENTE 372	1	4	2	7	2	5	7	9	1,56
373	EXPEDIENTE 373	2	5	3	10	1	4	5	9	1,67
374	EXPEDIENTE 374	2	5	3	10	2	5	7	10	1,70
375	EXPEDIENTE 375	1	4	2	7	1	4	5	8	1,50
376	EXPEDIENTE 376	1	4	2	7	1	4	5	8	1,50
377	EXPEDIENTE 377	2	5	3	10	2	5	7	10	1,70
378	EXPEDIENTE 378	2	5	3	10	2	5	7	10	1,70