



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

“Implantación de un modelo para Clasificación Automática de Textos en
Grupo Coril”

**TESIS PARA OBTENER EL TITULO PROFESIONAL DE
INGENIERO DE SISTEMAS**

AUTOR:

Vargas Salas Jose Fernando

ASESOR:

Mg. Manuel Hilario Falcón

LÍNEA DE INVESTIGACIÓN:

SISTEMAS DE INFORMACIÓN TRANSACCIONALES

LIMA- PERU

(2018)

PÁGINA DEL JURADO

 UCV UNIVERSIDAD CÉSAR VALLEJO	ACTA DE APROBACIÓN DE LA TESIS	Código : F07-PP-PR-02.02 Versión : 09 Fecha : 23-03-2018 Página : 1 de 1
--	---------------------------------------	---

El Jurado encargado de evaluar la tesis presentada por don (a) **VARGAS SALAS JOSE FERNANDO** cuyo título es:

Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril

Reunido en la fecha, escuchó la sustentación y la resolución de preguntas por el estudiante, otorgándole el calificativo de: 12 (números) doce (letras)

Lima, San Juan de Lurigancho 06 de diciembre del 2018


.....
MG. RENEE RIVERA CRISÓSTOMO
PRESIDENTE


.....
DR. HILARIO FALCON MANUEL
SECRETARIO


.....
MG. MARÍA ACUÑA MELÉNDEZ
VOCAL

					
Elaboro	Dirección de Investigación	Revisó	Responsable del SGC	Aprobó	Vicerrectorado de Investigación

DEDICATORIA

Está dedicado a mi familia, en especial a mis padres, por su paciencia y apoyo emocional. A mis maestros, por estar presente en momentos difíciles donde hay oscuridad sembraron luz de conocimiento.

Y a todas aquellas personas, que estuvieron a mi lado al cumplimiento de este objetivo.

Vargas Salas Jose Fernando

AGRADECIMIENTO

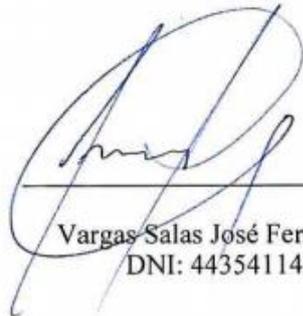
La gratitud es parte de la vida por lo que es una bendición. A todas las personas que me aportaron sus dones en mi camino durante mi vida, me alentaron también a los que corrigieron mis errores y llenaron de buenas experiencias a todas ellas muchas gracias.

DECLARACION DE AUTENTICIDAD

Yo, Vargas Salas José Fernando, con DNI N° 44354114 a efecto de cumplir con las disposiciones vigentes consideradas en el Reglamento de Grados y Títulos de la Universidad César Vallejo, Facultad de Ingeniería, Escuela de Ingeniería de Sistemas, declaro bajo juramento que toda la documentación que acompaño es veraz y auténtica.

Asimismo, declaro también bajo juramento que todos los datos e información que se presenta en la presente tesis son auténticos y veraces. 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 de la Universidad César Vallejo.

Lima, Diciembre de 2018



Vargas Salas José Fernando
DNI: 44354114

PRESENTACIÓN

Señores miembros del Jurado, presento ante ustedes la tesis titulada “Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril”, la misma que someto a su consideración, esperando que cumpla con los requisitos de aprobación para obtener el título Profesional de Ingeniero de Sistemas. El presente trabajo de investigación consta de seis capítulos.

El primer capítulo, consiste en la introducción del proyecto, donde se expone la realidad problemática, los trabajos previos y teorías relacionadas al tema, que son el sustento base de esta tesis, además se manifiestan los problemas, las justificaciones, los objetivos y las hipótesis que persigue la investigación.

En el capítulo dos, se detalla la metodología aplicada describiendo el diseño de la investigación, las variables y su operacionalización, además se determinan la población y muestra sobre la cual se realizaron las pruebas de pre-test y post-test, se explican las técnicas e instrumentos de recolección de datos y se describen los métodos de análisis de datos.

En el capítulo tres, se muestran los resultados obtenidos por cada indicador al realizar las pruebas respectivas tanto antes como después de la implementación del sistema de información, los cuales fueron descritos en el capítulo anterior, con sus tablas y gráficos correspondientes.

El capítulo cuatro muestra las comparaciones de los resultados del trabajo con los resultados obtenidos en otras investigaciones con la intención de respaldar estos trabajos o discrepar de ellos en el caso de no coincidir con la solución planteada.

En el capítulo cinco, se dan a conocer las conclusiones finales del proyecto de investigación por cada indicador según los resultados obtenidos en el capítulo anterior.

El capítulo seis VI se ven las recomendaciones correspondientes, así mismo se muestran las referencias de la investigación y para finalizar se muestran los anexos de la investigación.

ÍNDICE

FACULTAD DE INGENIERÍA	i
PÁGINA DEL JURADO	ii
DEDICATORIA.....	iii
AGRADECIMIENTO	iv
DECLARACION DE AUTENTICIDAD	v
PRESENTACIÓN	vi
ÍNDICE.....	vii
RESUMEN	xiii
ABSTRACT	xiv
I. INTRODUCCIÓN	15
1.1. Realidad Problemática.....	16
1.2. Trabajos previos.....	17
1.2.1. Trabajos previos internacionales	17
1.2.2. Trabajos previos nacionales.....	22
1.3. Teorías relacionadas al tema.....	30
1.3.1. Redes Neuronales	30
1.3.2. Procesamiento de Lenguaje Natural	31
1.3.3. Clasificación automática de documentos	32
1.3.4. Indicador de precisión en la clasificación de documentos	33
1.3.5. Indicador de tiempo promedio de registro de documentos	34
1.3.6. Deep Learning	34
1.3.7. NLTK	35
1.3.8. OCR.....	35
1.3.9. Perceptrón multicapa	36
1.3.10. TensorFlow	39
1.4. Formulación de Problema	40
1.4.1. Problema General:	40
1.4.2. Problemas Específicos:.....	40
1.5. Justificación del Estudio.....	40
1.5.1. Justificación teórica	40
1.5.2. Justificación práctica	40
1.5.3. Justificación Metodológica.....	41
1.5.4. Justificación social.....	41

1.6.	Hipótesis.....	41
1.6.1.	Hipótesis General:	41
1.6.2.	Hipótesis Específicos:	41
1.7.	Objetivos	42
1.7.1.	Objetivo General.....	42
1.7.2.	Objetivos Específicos	42
II.	METODO.....	43
2.1.	Diseño de Investigación.....	44
2.1.1.	Tipo de Investigación	44
2.2.	Variables, operacionalización	44
2.3.	Población y muestra.....	47
2.3.1.	Población	47
2.3.2.	Muestra	47
2.3.3.	Muestreo	48
2.4.	Técnicas e instrumento de recolección de datos.....	49
2.5.	Métodos de análisis de datos	49
2.5.1.	Procesamiento de datos	49
2.5.2.	Definición de Variables	50
2.5.3.	Hipótesis Estadística.....	50
2.5.4.	Nivel de Significancia	51
2.5.5.	Pruebas de Normalidad.....	51
2.5.6.	Pruebas de hipótesis Estadísticas.....	51
2.5.7.	Interpretación de datos.....	51
2.6.	Aspectos éticos	51
III.	RESULTADOS.....	52
3.1.	Análisis Descriptivo.....	53
3.1.1.	Análisis Inferencial.....	55
3.2.	Prueba de Hipótesis 1.....	58
3.3.	Prueba de Hipótesis 2.....	59
IV.	DISCUSIÓN.....	61
4.	DISCUSIÓN.....	62
V.	CONCLUSIONES	63

VI. RECOMENDACIONES	65
Bibliografías	67
VII. DESARROLLO	75
7.1. Datos utilizados.....	75
7.2. Corpus NLTK.....	75
7.3. Etapa de Entrenamiento.....	76
7.3.1. Pre – procesamiento de textos	76
7.3.2. Extracción de palabras claves.....	77
7.3.3. Ranking de palabras Claves	77
7.4. Etapa de evaluación semántica	77
VIII. PROCESOS DEL DESARROLLO DE LA APLICACIÓN.....	80
8.1. Entorno de trabajo.....	80
8.2. Instalación de Programas necesarios.	80
IX. IMPLANTACION DEL MODELO DE CLASIFICACION.....	84

INDICE DE TABLAS

Tabla 1 Matriz Operacional de la Variable	45
Tabla 2 Indicadores	46
Tabla 3 Número de documentos gestionados en un mes.....	47
Tabla 4 Frecuencia de Clasificación de algoritmos.....	50
Tabla 5 Cuadro de validación cruzada para la selección del modelo.....	53
Tabla 6 Indicador Precisión de clasificación.....	53
Tabla 7 Tiempo Promedio de Registro de documentos.....	54
Tabla 8 Matriz de Consistencia Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril	74
Tabla 9 Clasificación de documentos.....	76
Tabla 10 Comparación de Precisión de Clases.....	78
Tabla 11 Palabras Claves de Clasificación.....	79

INDICE DE FIGURAS

Figura 1 : La pirámide Deep learning.....	35
Figura 2: Perceptron multicapa	36
Figura 3 Tabla de Distribuciones.....	48
Figura 4: Instrumentos de recolección de datos	49
Figura 5 Porcentaje de Precisión de Clasificación	54
Figura 6 : Indicador de tiempo en Segundos	54
Figura 7: Kolmogorov- Precisión.....	55
Figura 8: Pre-Test Porcentaje de Precisión.....	56
Figura 9: Post-Test Porcentaje de Precisión	56
Figura 10: KOLMOGOROV-SMIRNOV	57
Figura 11: Tiempo Pre-test Registro de Documento	57
Figura 12: Tiempo Promedio en Registrar un Documento Post-Test	58
Figura 13 : Prueba de rangos Wilcoxon	60
Figura 14: Conjunto de Datos en Excel.....	75
Figura 15 Ananconda Python 3.0	80
Figura 16: Entorno Jupyter	81
Figura 17 NLTK - CORPUS	81
Figura 18: Inicio de Sesión login.....	88
Figura 19: Registro de tipos de documento	88
Figura 20: Registro de Usuarios	89
Figura 21: Registro de proveedores	89
Figura 22 Gestión Documental.....	90
Figura 23 Agregar Documentos	90
Figura 24 Visualización de documentos.....	91
Figura 25 Confirmación del Sistema	92
Figura 26 Resultado de la Clasificación	92

ANEXOS

ANEXO 1 : Matriz de Consistencia	74
ANEXO 2: Desarrollo de Teorías	75
ANEXO 3: Desarrollo de Software	80
ANEXO 4: Implantación del Modelo.....	84
ANEXO 5: Código Fuente	85
ANEXO 6: Interfaz de Usuario	88
ANEXO 7: Ficha de registro indicadores tiempo pre - test.....	93
ANEXO 8: Ficha de registro indicadores tiempo post-test	99
ANEXO 9: Ficha de porcentaje de precisión	106
ANEXO 10 : Autorización de publicación de tesis en el repositorio institucional UCV..	111
ANEXO 11: Acta de aprobación de Originalidad de tesis	112
ANEXO 12: Autorización de la versión final del trabajo de investigación	113
ANEXO 13: Porcentaje del Turnintin	114

RESUMEN

Las tecnologías de la información y la comunicación en los últimos tiempos han permitido al ser humano tener una cantidad excesiva de documentos en lenguaje natural. Dentro de toda la información que se genera hoy en día, podemos encontrar una nueva forma de comunicación, este trabajo de tesis surge con la necesidad de brindar una solución a los problemas que son el procesamiento del lenguaje natural y la ingeniería de software.

Generalmente para el diseño de un sistema de clasificación de información se recurren a una gran variedad de recursos lingüísticos el resultado son sistemas de gran portabilidad en diferentes idiomas. Podemos encontrar textos sobre productos, política, noticias, personas famosas, etcétera. Sin embargo, la capacidad del ser humano para analizar la información de estos textos tiene un tope, el cual en cierta forma, mediante el lenguaje de procesamiento natural se puede maximizar.

En la presente tesis se propone una metodología y se desarrolla una aplicación que permite el análisis de textos, clasificándolos en un repositorio de base de datos. Enfocándonos en textos cortos del idioma español y considerando una de las herramientas más importantes el NLTK.

PALABRAS CLAVES

Red Bayesiana, ERP, LPN, Perceptron, NLTK, OCR, MARKOV, Sistema Web, Gestión Documentaria.

ABSTRACT

The technologies of information and communication in recent times have allowed the human being to have an excessive amount of documents in natural language. Within all the information that is generated today, we can find a new form of communication, this thesis work arises with the need to provide a solution to the problems that are natural language processing and software engineering.

Generally, for the design of an information classification system, a great variety of linguistic resources are used, the result being highly portable systems in different languages. We can find texts about products, politics, news, famous people, etcetera. However, the capacity of the human being to analyze the information of these texts has a cap, which in a certain way, through the language of natural processing can be maximized.

In the present thesis a methodology is proposed and an application is developed that allows the analysis of texts, classifying them in a database repository. Focusing on short texts of the Spanish language and considering one of the most important tools the NLTK.

KEYWORDS

Bayesian Network, ERP, LPN, Perceptron, NLTK, OCR, MARKOV, Web System, Documentary Management.

I. INTRODUCCIÓN

1.1. Realidad Problemática

Los sistemas tecnológicos de acceso a la información están evolucionando, en las empresas existe una gran demanda de aplicaciones informáticas de documentación, lo que representa actualmente un gran mercado. (Maseda García, 1997, Pag.133).

La base de la documentación es el contenido de la información, el éxito de la empresa en un mercado competitivo depende muchos factores entre ellos el conocimiento y el aprendizaje, estos factores determinan si la empresa puede adaptarse a los cambios frente a la incertidumbre, dinamismo y la complejidad de los nuevos entornos. (Molina Pinto, 1998, Pag.171-183).

En el Perú la mayor parte de las empresas cuenta con una infraestructura básica en conectividad y comunicaciones. (INEI, 2015). Existe cierta cautela, confusión y en muchos casos desorientación al momento de implementar una solución documental, actualmente existen muchas aplicaciones informáticas que implementan esta solución, lo cierto es que todas comparten características en común pero responden a diferentes enfoques. Maseda García (1997, Pag.137).

En general la innovación tecnológica está a disposición para cualquier empresa y es económicamente accesible. Maseda García (1997, Pag.139) menciona que el uso del escáner de una impresora ha mejorado el reconocimiento óptico de caracteres, gráficos así como los tamaños gracias algoritmos que pueden minimizar la densidad de los archivos.

El procesamiento de lenguaje natural se lleva practicando desde hace tiempo y se comienza a recoger los frutos de años de inspección para dirigirlos a maximizar los recursos disponibles y optimizar los conocimientos gracias a los algoritmos de procesamiento que integran el uso de redes neuronales. (Cortes Vásquez, 2009, pag.53).

Grupo Coril necesita herramientas informáticas que generen conocimiento, el estándar de los programas informáticos, no son aquellos que simplifiquen o automaticen los procesos, (Matich 2011, pág. 4) menciona que son aquellos que puedan imitar el comportamiento humano, estamos viviendo tiempos donde la tecnología está a disposición, donde las aplicaciones cada día están desplazando el esfuerzo tratando de aliviar las cargas del trabajo inventando, diseñando y mejorando los procesos.

El personal de Grupo Coril recurre a un ERP de trámite documentario en el que de forma recurrente se desperdicia el tiempo en el registro manual, estas funciones están apartadas a la labor principal del empleado.

Por lo anterior, contribuir mediante una propuesta que facilite la clasificación de documentos por categorías y además que realice la tarea con buena precisión tendría un

gran aporte en el campo de la ingeniería, aun siendo experimental elevara la calidad de la gestión documentaria.

1.2. **Trabajos previos**

En materia de estudio y teniendo en cuenta las dos variables que son aplicación del Lenguaje de Procesamiento Natural y clasificación Documentaria se encontró los siguientes trabajos previos:

1.2.1. **Trabajos previos internacionales**

En la tesis de Daniel Isaac Paz Villaseca con título “Herramientas Deep Learning para Predictive Analytics. Una aplicación.”, para obtener la título de profesional de magíster en ingeniería industrial en la Universidad de Concepción – Chile 2017.

Se buscó implementar una aplicación Bussines Analytics. Utilizando las redes neuronales artificiales profundas con respecto a la elaboración de pronósticos, se hace la comparación entre los métodos estadísticos tradiciones versus el pronóstico de una red neuronal recurrente, adicional se realiza la comparación de modelos de pronósticos para minimizar el error.

Se llegó a la conclusión que mediante los algoritmos presentados se obtuvieron un mayor desempeño para el pronóstico de ventas y resultaron más eficientes que los reportes tradicionales (históricos).

En la tesis Betancourt Salazar Silvana del Rocío, Ponce Proaño Miguel Alejandro, con título “Sistema de reconocimiento óptico de caracteres, como solución al procesamiento digital”, con motivo de obtener el título profesional de Ingeniero Informático de la Universidad Central del Ecuador”, 2012.

Implemento en un sistema de procesamiento digital utilizando el reconocimiento óptico de caracteres, con librerías open source y programas básicos para su funcionamiento, las cuales permitieron el procesamiento de texto, almacenamiento del contenido del documento, procesamiento en batch de varios documentos, durante el proceso la información digitalizada se almacenaba en una base de datos, permitiendo la generación de reportes, numero de caracteres, número de líneas, rendimiento de software.

En conclusión se logró obtener buenos resultados de las pruebas realizadas en reconocimiento óptico de caracteres y procesamiento de texto llegando a un control de un 70%, el cual es aceptable y cumple con los requerimientos mínimos para su correcta usabilidad.

En la tesis de Andres Fabian Rodriguez Peña con el título “Implementar un herramienta tecnológica de información para la producción y gestión de documentos en la organización Serdán”, para obtener el título Ingeniero Industrial en la Universidad Libre de Colombia, en Colombia 2016.

En la investigación propone obtener la eficiencia y la confianza de la gestión documental para el uso de la información oportuna y toma de decisiones, utilizando la metodología PETI (Planeación estratégica de Tecnologías de la información).

En conclusión se logró buenos resultados con gestión documentaria por el uso del proceso de planeación dinámico, en la que las estrategias sufren continua adaptación, innovación y cambios.

En la tesis de Elvia Leticia Williams Map, con el título “Propuesta para la digitalización del fondo documental del Dr. Belisario Porras” con motivo de optar el título de “Master en gestión documental y administración de archivos “, en la Universidad de Andaluza en el país de España, 2013.

Implemento el sistema de gestión documentaria donde se gestiona los documentos históricos del presidente Dr. Belisario Porras, aplicando la digitalización de documentos.

En conclusión el sistema permite el procesamiento sobre texto e imágenes con el objetivo de preservar la seguridad de la información.

En la tesis de Nadia Patricia Araujo Arredondo con el título “Método Semisupervisado para la clasificación Automática de Textos de Opinión” para obtener el grado de Maestría en Ciencias en el área de Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica en Tonantzintla, Puebla, 2009.

Implemento la clasificación de la polaridad de opiniones, construyo un corpus en español y desarrollo el método de clasificación semi-supervisado utilizando para su clasificación el método de máquina de vectores de soporte el cual demostró un gran desempeño en una gran variedad de clasificación de textos.

En conclusión al realizar los experimentos de clasificación de polaridad de opiniones se concluyó que utilizando la clasificación semi-supervisada se permite alcanzar mejores resultados con un porcentaje del 10% de exactitud a comparación de los métodos tradicionales.

En la tesis de Roque Enrique López Condori con el título “Método de Clasificación Automática de Textos basado en Palabras Claves utilizando Información Semántica: Aplicación a Historias Clínicas”, para obtener la título de profesional de Ing. de Sistemas en la universidad Nacional de San Agustín en el 2014.

Implemento un método de clasificación de historias clínicas basándose en palabras claves utilizando la información textual, selecciono las palabras más representativas. Estableció un ranking de palabras claves y evaluó el método propuesto. Es importante mencionar que se implementó para el idioma inglés debido a los corpus disponibles utilizando la medición de Resnik que mide la relación semántica entre conceptos.

En conclusión se mostró una propuesta para la clasificación de historias clínicas, gracias a las mediciones de los algoritmos propuestos se pudo concluir cual es clasificador con mayor desempeño ayudando a la clasificación de documentos.

En la tesis de Alberto Téllez Valero con el título de “Extracción de Información con Algoritmos de Clasificación”, para obtener la título de profesional de Maestro de ciencias en Instituto Nacional de Astrofísica, Óptica y Electrónica. En Tonantzintla, Pue, el 2015.

Implemento un clasificador utilizando métodos empíricos que mejoran el entendimiento del lenguaje comparando los resultados obtenidos por los algoritmos de clasificación.

En conclusión la arquitectura propuesta presenta una carencia del entendimiento del lenguaje reemplazando los métodos de clasificación supervisada mostrando de manera detalla las ventajas y desventajas, como las palabras que no aportan con un significado a la clasificación.

En la tesis de Raúl Ignacio Sanhueza Arancibia con el título “Formas de representación de textos cortos para su posterior Clasificación Automática” para obtener título profesional de Ingeniero Civil en Informática en la Pontificia Universidad Católica de Valparaíso, 2013.

Implemento un método de clasificación de textos utilizando en el enfoque de máquinas de aprendizaje con la representación vectorial. La investigación mejoro gracias al árbol de decisiones porque mediante los grafos se pudo entregar buenos resultados.

En conclusión se clasificaron 5239 twets, por su grado de intención sin embargo este proceso es muy lento por lo que queda descartado, en general para un tiempo optimo se considera un volumen de 800 datos, que es menor a 5239.

En la tesis de J. David Lojo Vicente con el título de “Clasificación Automática de Documentación Clínica” para obtener el grado de Doctor en la Universidad de Santiago de Compostela en Da Coruña, 2012.

Implemento la clasificación automática de en un entorno de la documentación clínica. Utilizando el aprendizaje supervisado donde los documentos son informes de alta hospitalaria.

En conclusión se clasifico la documentación de las altas, utilizando los algoritmos de clasificación KNN y SVM (Maquinas de Soporte Vectorial). Una de las conclusiones más importantes es que la clasificación manual es muy caro e ineficiente frente a las soluciones automatizadas.

En la tesis de Moisés Pastor Jorge Carrillo de Albornoz Cuadrado con el título “Un Modelo Lingüístico-Semántico Basado en emociones para la Clasificación de Textos según su Polaridad e Intensidad” para obtener la título de profesional de Doctor en Informática de la Universidad Complutense de Madrid, 2011.

Implemento de un clasificador de análisis sentimental como la negación, los cuantificadores o los modales. La tecnología utilizada es un modelo basados en emociones desde la perspectiva emocional utilizando las palabras más frecuentes o patrones de expresión y sustantivos subjetivos desarrollando un análisis profundo utilizando el software Word Net Affect.

En conclusión los resultados nos indican que a pesar de que las palabras más frecuentes pueden ser un indicador alternativo para la clasificación de textos, cabe mencionar que gran parte de los errores de clasificación se debe al mal uso del lenguaje de expresión de parte de usuario.

En la tesis de Lic. Hugo Librado Jacobo con el título “Análisis automático de opiniones de productos en redes sociales”, para obtener el grado de Maestría en Ciencias de la Computación en el Instituto de Politécnico Nacional en México, 2016.

Implemento una aplicación de software que permita analizar textos cortos en idioma español de redes sociales, twitter y determinar su polaridad. En el desarrollo de la aplicación propuso una metodología diferente a las existentes haciendo uso de la comparación de los resultados obtenidos en otros sistemas.

En conclusión el trabajo desarrollado transmite una complejidad diferenciada de la unipolaridad porque trata de en diferentes sentidos en un árbol de decisión, considerando los términos que resultan multipolares.

En la tesis de Arkaitz Zubiaga Mendiakua con el título de “Aproximaciones a SVM semisupervisado multiclase para clasificación de páginas web” para obtener la título de profesional de Master de Tecnologías Web en la Universidad de Educación a Distancia en España, el año 2008.

Implemento el clasificador SVM para la clasificación de páginas web. El estudio está centrado al análisis del algoritmo SVM. Utilizando el método semisupervisado que utiliza una colección de datos no etiquetada refinando el clasificador de forma inductiva.

En conclusión se realizó una comparación del método de clasificación SVM semisupervisado ante la problemática de clasificación de páginas web destacando una mayor efectividad en el tratamiento de los datos no etiquetados.

En la tesis de Lic. María Verónica Rosas con el título de “Un Framework de Ingeniería del lenguaje para el pre-procesado Semántico de Textos”, para obtener la título de profesional el grado de Maestría en Ingeniería de Software en Argentina, 2012.

Implemento los métodos supervisados para representación semántica de documentos utilizando WSD para el entrenamiento. El WSD demuestra uno de los mejores comportamientos en base a resultados porque constituyen una herramienta al realizar trabajos de representación semántica.

De los resultados obtenidos en los experimentos el procesamiento de textos se concluye que la información semántica representación de los textos aporta beneficios en el enfoque del conocimiento y en la tarea de la desambiguación.

En la tesis de Dr. Juan M. Ale con el nombre “Procesamiento de Lenguaje Natural en Sistemas de Análisis de Sentimientos” para obtener el grado de Ingeniero de Informática en la Universidad de Buenos Aires Argentina, 2013.

Implemento un análisis de texto relacionado a los sentimientos dentro de un mismo documento. Seleccionado aquellas características utilizo el “parser” que permite las frases nominales del documento. Con el uso de la evaluación del porcentaje polaridad de los métodos de clasificación será suficiente para ejecutar la clasificación y obtener los parámetros necesarios para la evaluación de la misma.

En conclusión se investigaron las técnicas para el análisis de sentimientos adaptadas al idioma español alcanzando un 95% de dominio de la clasificación.

En la tesis de Pablo Toloza con el título “Construcción automática de grafos RDF a partir de frases en lenguaje natural para la integración de bases de datos heterogéneas”, para obtener la título de profesional el grado de Master en Madrid España, 2014.

Implemento dentro del Sistema p-medice un servicio que tiene el propósito de la búsqueda en lenguaje natural, el servicio permitirá desarrollar al usuario consultas a la base de datos a partir de frases naturales. Implemento NLTK con el objetivo de proporcionar un marco intuitivo, uniforme el que construye automáticamente las palabras gracias al corpus de entrenamiento que posee.

En conclusión se desarrolló una aplicación en REST para interpretar el lenguaje natural disminuyendo el tiempo utilizado por otros etiquetadores el resultado es sencillo de aprender e intuitivo y su desempeño en cuando al procesamiento es aceptable.

En la tesis de Toselli Alejandro Hector con el título “Reconocimiento de textos manuscrito Continuo”, para obtener la título de profesional el grado de Doctor en de informática en la Universidad Politécnica de Valencia en España, 2014.

Implemento un Sistema automático de textos implementando la metodología de cadena oculta de Markov. También utilizo el modelamiento de N-Gramas para la representación gráfica de grafos en la fase de entrenamiento se utilizó el HTK un software de reconocimiento de habla.

En conclusión se logró tener una tasa de error de 9.4 % obtenida en un resolución de 28. El test final se utilizando la topología HMMS con unas 5000 pruebas logrando un tasa de error que corresponde a un 9.3% para la estimación de una medida de confianza.

En la tesis de Gómez Durand Mauricio con el título “Reconocimiento de caracteres ópticos OCR por medio de la correlación y redes neuronales”, para obtener la título de profesional el grado de Ingeniero Electrónico en la Universidad Pontificia Bolivariana en Bolivia, 2014.

Implemento el reconocimiento de manuscritos por medio del método de correlación de redes neuronales. Usando el lápiz óptico gracias al método de correlación y redes neuronales en los cuales se aplicaron los conceptos teóricos, matemáticos y aplicaciones.

En conclusión se logró un porcentaje en uso de la red de base radial, obteniendo un porcentaje de 91% gracias al método Matlab y tuvo una efectividad de un 87%. Para el entrenamiento de la red de back – propagation.

1.2.2. Trabajos previos nacionales

En la tesis de Raúl Elías Rosado Exellmes con título “Aplicación de una interfaz de lenguaje natural textual para la gestión administrativa del sistema de registro académico del instituto de educación superior tecnológico público Juli”, para obtener la título de profesional de Ingeniero de Sistemas”, en Puno Perú 2016.

Presenta una aplicación que aprovecha el procesamiento de lenguaje natural para la generación de información sin utilizar el SQL realizando consultas simples en ILNBD (Interfaz de lenguaje natural de base de datos) que se encarga de la red de transmisiones

dándole facilidad al usuario de personalizar reportes siempre que se entienda su forma de uso y limitaciones.

En conclusión la investigación comprobó que aplicar ILNBD mejora la gestión administrativa, se reduce el tiempo de administración, adquiere una aceptación en 66.66% en los usuarios, además el lenguaje de procesamiento natural es muy útil para que el software entienda el lenguaje humano.

En la tesis de Eduart Rudolf Garrido Rojas, con título “Diseño e implementación de reconocimiento de números manuscritos”, con motivo de optar por el título ingeniero electrónico en la Universidad Pontificia Católica del Perú, 2010.

Estudio la efectividad que tiene el reconocimiento de caracteres para modelos de Matlab, en su investigación implemento el reconocimiento de manuscritos gracias a la inteligencia artificial, con el método Matlab de redes recurrentes y BackPropagation se propuso el análisis de las pruebas que resultaron en un 95.9% de efectividad con un porcentaje de error de 0.8%., usando como metodología los métodos Markovianos que son generados por mecanismos interrelacionados con un número finito de estados el cual nos da como resultado las probabilidades de transición de letras que pueden ser de un carácter finito dándonos una palabra.

En conclusión la metodología desarrollada obtuvo un reconocimiento óptico de caracteres del 90.11% y el error de 3.67% con la base de datos “National Institute of Standards and Technology” con el que se puede optimizar y obtener mejores resultados.

En la tesis de Malaver Elera, Manuel Alexander con el título “Aplicación de redes neuronales para determinar el pronóstico de las ventas en la empresa catering & buffets Mys ubicada en la ciudad de Piura” para obtener la título de profesional de Ingeniero de Sistemas en la Universidad Cesar Vallejo en el Perú, 2015.

Implemento el pronóstico de ventas con un bajo porcentaje de error aplicando el modelo de red supervisada. Kaastra y Boyd fue el modelo que implemento para el desarrollo la red neuronal multicapas. Los resultados de su trabajo fueron que las redes neuronales multicapas son las más adecuadas aplicadas para este tipo de organizaciones.

Como conclusión el porcentaje de error del modelo de red neuronal planteada es del 1%. Donde la función sigmoideal fue el más adecuado para predecir una mejor ventaja competitiva en las ventas.

Rogelio Nazar, Patricio Arriagada. 2017. “Sociedad española para el lenguaje de procesamiento natural”, *Calidad de revistas Científicas españolas* .58, 13-20.Paginas, España.

Implemento una metodología que ayudo en la clasificación y detección de nombres propios (NP), la hipótesis trata sobre el contexto de la aparición de nombres propios que pueden dar seguimiento al nombre de la entidad. Por tal motivo elaboro el método de clasificación que se entrenó con el corpus “Freeling” supervisado, un anotador en la selección de los idiomas y el corpus CONNLL-2002 en su edición en castellano.

En conclusión se presentó un trabajo de metodología para la clasificación y detección de nombres de una entidad, la propuesta es de carácter investigativo y de bajo coste computacional al no requerir ningún procesamiento de corpus. La propuesta CONNLL-2002 se procesa en menos de 1 segundo en una PC de escritorio el cual presenta una ventaja importante sobre otros modelos más complejos.

En la tesis de Saavedra Rosales, Yefffer José con el título “Sistema web para la gestión documental en la empresa Development It E.I.R.L” para obtener la título de profesional de Ingeniería de Sistemas en la Universidad Cesar Vallejo en el Perú, 2015.

Implemento un sistema web para la gestión documental.

En conclusión se puede destacar el tiempo de procesamiento que disminuyo en un 84% para gestión documental.

En la tesis de Jhony Richard Quispe Obregón, Jewer Smith Vilchez Huachaca con el título “Relación entre la implementación de un sistema de tramite documentario y la gestión documentaria en la municipalidad del Rímac” para obtener la título de profesional de Ingeniero Empresarial y de Sistemas en la Universidad San Ignacio de Loyola en el Perú, 2017.

En la investigación se implementó el sistema de tramite documentario en la municipalidad de Rímac en una entrevista con los gerentes muestran su preocupación por algunos inconvenientes en el uso del aplicativo.

En conclusión el sistema documentario recopila en tiempo real el impacto de la gestión documental.

En la tesis de Rodrigo David Barahona Llore con el título “Perceptron Wavelert y algoritmos genéticos para la predicción de captura de la anchoveta” para obtener la título de profesional de Ingeniero Civil en informática en la Pontificia Universidad Católica de Valparaíso, 2014.

Implemento una red neuronal artificial que posee la capacidad de “aprender con la experiencia”, aplicando redes neuronales recurrentes evolutivas y evaluando el rendimiento de Jordan y Elman en el proceso de asertividad utilizando métricas residuales.

En conclusión se obtuvo el mejor rendimiento utilizando el modelo Elvan a comparación de Jordán, debido a su metodología. Se establecieron los modelo predictivos para las anchovetas en el volumen mensual y estableciendo paramétricas que ayuden a seleccionar el modelo con un buen resultado.

En la tesis de Luis Alejandro Pinto Valdiviezo con el nombre de “Modelo algorítmico para la clasificación de documentos de carácter judicial en lenguaje portugués según su contenido” para obtener la título de profesional de Ingeniero Informático en la Pontificia Universidad Católica del Perú en el Perú, 2015.

Implemento un modelo de algoritmo de clasificación de documentos en el lenguaje Portugués, para la investigación propuso 4 algoritmos de clasificación, realizo la tarea de recopilación de información para la selección del modelo utilizando un cuadro comparativo en base a resultados, finalmente para el análisis de la información utilizo el software WEKA para el tratamiento del texto.

En conclusión se mostró el uso del Software WEKA para el tratamiento de los datos, la aplicación ayudo a probar los diversos modelos de clasificación, seleccionando de clasificación más confiable.

En la tesis de Br. Barrios Méndez Arnaldo Abel con el título “Sistema Web y móvil de Trámite Documentario para mejorar la Gestión de Movimientos de equipos Informáticos del área de soporte técnico en una Universidad” en la universidad cesar vallejo para obtener el título de ingeniero de sistemas, 2016.

Implemento la metodología ICONIX que permite realizar el análisis y diseño del Sistema haciendo uso del UML para el modelamiento. Para el desarrollo del software se utilizó Java para la arquitectura web y Android para dispositivos móviles.

En conclusión se obtuvo un reducción de tiempo en 2.76 minutos a un tiempo de 1.83 que es un porcentaje del 34%, como consecuencia el nivel de satisfacción de 2.13% a 4.63%.

En la tesis de Mgtr. Patricia Malena Cepeda Gamio con el título “Programa de automatización de los procedimientos de trámite documentario en la calidad del servicio a los usuarios del Hospital Nacional Arzobispo Loayza – Lima, 2016” en la Universidad Cesar Vallejo para obtener el grado académico de doctora en la gestión Publica en el Perú, 2016.

Implemento un programa de automatización de procesos de trámite documentario que permite el CRUD de un expediente, consultas en tiempo real, facilidad de búsquedas y permite emitir informes estadísticos utilizando la metodología Hipotética deductiva.

En conclusión se concluye que la aplicación mejora la calidad del servicio a los usuarios aumentando la productividad de la gestión documental.

En la tesis de Egusquiza Escriba, Xiomí Geraldine con el título “Sistema Web para el proceso de gestión documental para la empresa de prevención Global” en la Universidad Cesar Vallejo para obtener el título profesional de Ingeniero de Sistemas en el Perú, 2015.

Implemento de un sistema de archivos para la gestión documental bajo la metodología RUP para el desarrollo del software.

En conclusión se mejoró el proceso de la gestión documental en un 40% mejorando los sistemas de trámite documentario tradicionales.

En la tesis de Miguel Ángel Flores Márquez con el título “Sistema informático para el proceso de Tramite Documentario en la municipalidad de Chaclacayo” en la Universidad Cesar Vallejo para obtener el título profesional de Ingeniero de Sistemas en el Perú, 2017.

Implemento un sistema de trámite documentario bajo la metodología RUP y un ambiente de programación WEB. Para el desarrollo se implementó Spring MVC para el desarrollo del Back-End y para el Fron-End Angular JS.

En conclusión luego de la implementación del aplicativo WEB mejoro el acceso a la información, reduciendo el tiempo a 37 segundos el porcentaje que equivale a un 47 % del tiempo.

En la tesis de Maldonado Barrios Alexander con el título “Sistema Web para el proceso de Gestión Documental de la División de Seguridad de Contraminas – PNP-Rímac” en la Universidad Cesar Vallejo para obtener el título profesional de Ingeniero de Sistemas en el Perú, 2018.

Implemento un software para determinar la influencia del Sistema web en el proceso de trámite documentario de la división de Seguridad de la información. Para su desarrollo utilizo la metodología RUP, estableciendo parámetros y buenas prácticas. Utilizando Java como plataforma y mysql para la base de datos.

En conclusión el sistema web mejora el proceso de trámite documentario, porque localizo los documentos en un 19% a comparación de otros sistemas por lo que se afirma que el sistema Web Incrementa el servicio de tramite documentario.

En la tesis de Moran Trujillo, Luis Ángel con el título “Sistema de Informático para la gestión documental para la empresa Hector Gonzales Sandi Agencia afianzada de Aduana

S.A.” en la Universidad Cesar Vallejo para obtener el título profesional de Ingeniero de Sistemas en el Perú, 2017.

Implemento un Sistema de Información para determinar la influencia de la gestión documental por tal motivo se desarrolló una metodología ágil con el propósito de almacenar y recuperar documentos y todo el flujo documentario. La metodología utilizada es RUP y SCRUM. Que mejora el control y el orden de toda la información documental.

En conclusión se puede afirmar que la gestión documental tenía un valor de 41.8 % antes de implementar el sistema y aumento su valor a 87.7%, de la misma forma el porcentaje de documentos localizados de 41% a 86.6%.

En la tesis de Alva Machengo Fernando Emilio con el título “Sistema de Información de detección de plagio en documentos digitales usando el método Document FingerPrinting”, para obtener la título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2010.

Implemento un sistema que permitió almacenar una serie de documentos que sea capaz de detectar plagio por copia directa la metodología de desarrollo es SCRUM. Para la clasificación de palabras se utilizó los métodos de redes bayesianas.

En conclusión, la implementación del método Document FingerPrinting se justifica porque obtiene un mayor puntaje, además de que se explica de las técnicas de reconocimiento de patrones de texto que son utilizados en diferentes algoritmos de clasificación para demostrar su eficiencia.

En la tesis de Carrera Jiménez Dorila Sarita con el título “Análisis y diseño de un sistema de tramite documentos de pago de proveedores vía Intranet”, para obtener la título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2009.

Implemento un sistema de registro de documentos digitalizados, la metodología es RUP y para el Back End JEE2 usando la programación orientada a objetos, toda la arquitectura es WEB en tres capas MVC.

En conclusión se mejoró el proceso de pagos de documentos se controló el seguimiento al flujo de pago todo se realizó gracias a la implementación del sistema.

En la tesis de Escobar Sedano Mayra Carolina con el título “Análisis, Diseño e implementación de un sistema de apoyo al seguimiento de proceso judiciales para un estudio de abogados”, para obtener la título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2010.

Implemento un sistema de toma de decisiones con la metodología RUP porque proporciona muchas herramientas para la programación orientada a objetos, se utilizó la

metodología de desarrollo XP para aumentar la velocidad del desarrollo y minimizar la documentación.

En conclusión al comparar las metodologías RUP y XP se demostró que ambos llegan a los mismos resultados sin embargo la metodología tradicional tiene un gran impacto en el costo del proyecto, además de la implementación completa de los requerimientos en el diagrama de RUP.

En la tesis de Gómez Ruiz Bach. Erickson Piere con el título “Implementación de un Sistema de Documentación bajo la plataforma Web para la gestión y control documental de la empresa Corporación JUJEDU E.I.R.L. – talara 2017”, para obtener el título de profesional de Ingeniero Informático en la Universidad de Católica de los Ángeles de Chimbote. Perú, 2017.

Implemento un sistema de información web usando la metodología de desarrollo RUP, entre los procesos principales se encuentra la automatización la programación orientada a objeto, conceptos de desarrollo y diagrama de procesos.

En conclusión se tiene un nivel de satisfacción, gracias a la implementación del sistema actual. Permitiendo que un 75% de los trabajadores expresen su interés en la gestión documental.

En la tesis de Liberato Navarro Albert Felix, Nadya Elida Marcial Reynalde con el título “Desarrollo de un sistema de gestión documental, fichas de resumen y lista de publicación para el proyecto PROCAL-PROSER”, para obtener el título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2014.

Implemento un sistema de gestión documental con la guía de PMBOK el mismo que propone la dirección de la toma de proyectos. La metodología es RUP, la programación es orientada a objetos en PHP y MYSQL para el gestor de base de datos.

En conclusión mejoro la gestión documentaria porque se implementó la elaboración de publicaciones, asignación de etiquetas, autores y grupos. Además se identificó las falencias.

En la tesis de Herrera Garcia Bach. Maikel Edu con el título “Implementación de un sistema Web de Gestión de trámite documentario para optimizar los procesos en el área de mesa de partes en la municipalidad distrital de Bellavista –Sullana”, para obtener el título de profesional de Ingeniero Informático en la Universidad de Católica de los Ángeles de Chimbote. Perú, 2018.

Implemento un Sistema de trámite documentario en un ERP usando la metodología UML para desarrollar el comportamiento de las aplicaciones. El lenguaje utilizado es

PHP el servidor de aplicaciones es XAMP con MYSQL, realizo el desarrollo bajo la metodología de desarrollo ágil XP.

En conclusión se logró mejorar los procesos de la organización gracias a la estructuración del modelado de procesos, también se logró mejorar el tiempo de respuesta del trámite documentario, con la ayuda del Sistema Web.

En la tesis de Romero Galindo Raúl Miguel con el título “Análisis, Diseño e implementación de un Sistema de información aplicado a la gestión Educativa en centros de educación especial”, para obtener la título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2014.

Implemento un Sistema que permita generar el documento de capacitación para los niños con habilidades especiales, se utilizó la metodología RUP para la correcta gestión de requerimientos el entorno de desarrollo se utilizó WEB con ASP en C#, ADO para conexión a SQL.

En conclusión el trabajo fue estrictamente académico sin fin lucrativo logrando incorporar un sistema para automatización de programas educativos. Se logró la integración de módulos bajo la plataforma .net, con un diseño a escalabilidad.

En la tesis de Villar García Gustavo con el título “Implementación de un Sistema de gestión de Investigación para la especialidad de Ingeniería informática de la PUCP”, para obtener la título de profesional de Ingeniero Informático en la Universidad de Católica del Perú, 2014.

Implemento un Sistema de Gestión de la Investigación en la plataforma WEB tomando como metodología la programación extrema (XP) de acuerdo a la problemática de la empresa, también se añadió los generadores de reportes del estado de la investigación. El desarrollo está en RUBY y MYSQL.

En conclusión el sistema de gestión permitió la recopilación de la información que es tema de investigación, logrando centralizar la información por medio de la navegación del software.

En la tesis de Monge Osorio Manuel Alejandro con el título “Diseño y arquitectura de una Red Neuronal Artificial Perceptron Multicapa sobre un FPGA aplicada al reconocimiento de Caracteres”, para obtener la título de profesional de Ingeniero Electrónico en la Universidad de Católica del Perú, 2008.

Implemento un sistema de reconocimiento de caracteres utilizando las redes neuronales artificiales usando el algoritmo Perceptron – back propagation. Diseño una red de una neurona artificial y verifico su funcionamiento.

En conclusión los resultados obtenidos por la neurona requerían una mayor cantidad de datos para mejorar la probabilidad. Se concluye que las redes neuronales pueden implementarse en hardware y en software. La velocidad del procesamiento de la red neuronal es veloz por lo que se puede implementar en tiempo real.

En la tesis de Ascarza Mendoza Franco Javier con el título “Segmentación automática de textos, mediante las redes neuronales convolucionales en imágenes de documentos históricos”, para obtener la título de profesional de Magister en Informática con Mención en Ciencias de la Computación Universidad de Católica del Perú, 2018.

Implemento una red neuronal convolucional profunda para la segmentación de imágenes con el objetivo de construir una red de aprendizaje, también se evaluó el desempeño de los algoritmos.

En conclusión se procesó las imágenes bajo una secuencia de procesos el pre-procesamiento, la red neuronal y el entrenamiento. Los resultados obtenidos por la clasificación de imágenes demuestra la calidad de la segmentación de imágenes.

1.3. Teorías relacionadas al tema

1.3.1. Redes Neuronales

Son sistemas de procesamiento de la información, está inspirado en las redes neuronales biológicas, estas consisten en nodos o neuronas que están conectadas entre sí y tienen un peso que es modificable. De forma simple es sumar los valores de entrada que son comparadas al valor del umbral para enviar una activación o salida. El procesamiento de la información se da mediante parámetros que definen el “conocimiento” que una red neuronal posee. En los sistemas convencionales el modelo conexionista se echa a perder por un resultado dañado, por otro lado el modelo de red neuronal permite reconocer objetos que solo presenten como entrada, una parte del mismo. Por último el modelo de red neuronal es “entrenado” con el aprendizaje que consigue con el ajuste de los pesos entre conexiones. (Rodríguez, R. P. 2017.)

1.3.2. **Procesamiento de Lenguaje Natural**

(Siddiqui & Tiwari, 2008) mencionó, que la Inteligencia Artificial (AI) tiene como ramas principales el procesamiento del lenguaje natural. El lenguaje de procesamiento natural se enfoca en la actualidad en problemas prácticos, en los que están mejor dirigidos. El lenguaje de procesamiento natural se diferencia de otros sistemas por el uso del conocimiento de la lengua. En ese sentido, los sistemas deben enfocar todos los grados de la lengua (morfológico, sintácticos, léxico), además de las intenciones del uso (pragmático). (Jurafsky & Martin, 2000).

(Ferreira, 2010) menciona, el procesamiento natural es la tecnología que permite desarrollar algoritmos, que diferentes lenguaje de programación han implementado en sus frameworks, estos, a su vez han permitido incorporar librerías haciendo uso de la lingüística, estas reglas pueden ser ecuaciones predictivas de casos frecuentes de palabras. Se puede mencionar, que el proceso inicia con el análisis de las estructuras en los párrafos ingresadas por una persona que genere inputs de datos, con esta información lingüística se pueden realizar operación de detección de errores, buscar palabras en un párrafo, implementar una conversación, etcétera.

1.3.2.1. ***Técnicas básicas para el tratamiento de lenguaje natural***

Las técnicas del procesamiento de lenguaje natural se desarrollan a través de etapas, estas pueden ser de manera secuencial o paralela la etapa inicial es el procesamiento de las palabras que incurre en un proceso textual, bajo esta premisa, el párrafo puede ser mencionado como una simple cadena de silabas, los procesos que se deben realizar en esta fases son: 1) la segmentación del párrafo, 2) el filtrado de palabras no relevantes en el texto y, 3) la ubicación de caracteres procesados. Tiene una apariencia simple por la secuencia de pasos, sin embargo, estos procesos resultan complicaciones en algunos sistemas. (Ferreira, 2010).

1.3.2.2. ***Arquitectura de un Sistema de procesamiento de lenguaje natural***

La arquitectura del lenguaje de procesamiento natural se define por niveles el fonológico en su definición son los sonidos relacionados a las palabras, morfológico son unidades pequeñas llamadas morfemas que en conjunto forman la palabra, sintáctico son formadas por unión de palabras buscando la estructura de una oración , semántico es el significado de las palabras en las oraciones estructuradas y pragmático es el significado

que le precede a las oraciones de acuerdo a las situaciones y el uso afectado. (Vásquez, 2009).

1.3.2.3. *Clasificación del Lenguaje de Procesamiento Natural*

En el campo de la clasificación del procesamiento del lenguaje natural se busca respuesta en la especialidad de la detección y clasificación los nombres propios nombrados de manera genérica como “entidades”. Estas se catalogan de manera principal como entidades incluyen fechas, números y signos, para ser tratados y clasificados como valores o cantidades en la tarea de reconocimiento de texto. El procesamiento de lenguaje natural requiere el análisis de texto, sintáctico y en algunos casos hasta textuales, sin embargo la investigación continúa abierta hasta la actualidad. (Procesamiento del Lenguaje Natural, Revista n° 58, marzo de 2017).

1.3.3. **Clasificación automática de documentos**

(Figuerola, 2013, Pág. 2), menciona que la clasificación de documentos tiene como finalidad fundamental el poder localizar con facilidad los documentos con unas características determinadas. La clasificación automática es aplicable a documentos en formato digital estas permiten resolver tareas de asignación automática, como sucede en las páginas web según menciona (Sebastiani, 2012) mencionado por Figuerola.

La clasificación se da mediante técnicas simples como la detección de palabras claves, sin embargo existe muchos obstáculos como la sinonimia y la polisemia, bien conocidos desde hace tiempo en las ciencias de la documentación que da como resultado el fracaso en la mayoría de los casos.

La aplicación de técnicas a la clasificación automática puede producir buenos resultados si son procesadas de forma específica. Esta especialidad la tienen los algoritmos que funcionan como manejadores de texto que podrían clasificar por palabras que conforman dichos texto, el cual requiere una dimensionalidad alta, puesto que el vocabulario del lenguaje natural es muy amplio.

1.3.3.1. *Que es un Corpus*

(Editorial tecnologías del lenguaje, pag43, 2017), menciona que es una colección de datos seleccionados y ordenados, para ser usada como muestra. Un corpus automatizado es aquel que tiene uniformidad y está generalizado para la tarea de recuperación de la

información. Por tal motivo hay que señalar que la recuperación empírica no habría podido generarse sin haberse desarrollado antes esta tecnología.

Cada tipo de corpus tienen un propósito que es justificado por la estructura y por el proceso en el cual se han planteado, por ende estas se utilizan dentro del marco de desarrollo.

El uso que tienen estos corpus es la información contenida, dentro de sus características se pueden mencionar el etiquetado o la anotación. Por medio del etiquetado podemos determinar la categoría gramatical de cada elemento léxico y entender las apariciones que tendrán dentro del contexto en la documentación.

A diferencia de un archivo plano de texto o de una biblioteca electrónica, los corpus suelen incorporar diferentes tipos de clasificación donde cada unidad de texto contiene una naturaleza que permite ser explotada y procesada de manera específica algunos ejemplos pueden ser:

- 1) Aspectos bibliográficos del texto: autor, título, año de publicación, tema, género, etc.
- 2) Aspectos de estructura del texto: marcado tipográfico, divisiones textuales, párrafos, citas, títulos, fragmentos no analizables, etc.
- 3) Caracterización de las unidades léxicas: léxico general, extranjerismos, nombres propios, abreviaturas, siglas, etc.
- 4) Caracterización morfosintáctica de las unidades léxicas: categoría gramatical, forma, gráfica del lema, categorías morfológicas, etc.
- 5) Constituyentes sintácticos: representación de los árboles sintácticos, funciones sintácticas de cada constituyente, etc.
- 6) Caracterización semántica: desambiguación de unidades homógrafas, caracterización de usos polisémicos, etc.

1.3.4. **Indicador de precisión en la clasificación de documentos**

Precisión de documentos clasificados: (Cruz, Troyano, Enriquez, & Ortega, 2008) Es el proceso de clasificar documentos, basados en contenido, es la tarea que trata de identificar la naturaleza del documento, aplicando un corpus de entrenamiento de documentos clasificados, conociendo el significado de las palabras a nivel sintáctico, semántico y pragmático.

- La fórmula para el porcentaje de precisión de documentos.

$$PC = \left(\frac{DC}{DU} \right) * 100$$

PC: Porcentaje de precisión de documentos.

DC: Cantidad de documentación clasificados correctamente.

DU: Cantidad de documentos analizados.

1.3.5. Indicador de tiempo promedio de registro de documentos

Tiempo promedio de registro de documentos: (Amendola, 2003), menciona que es un indicador capaz medir la confiabilidad que tiene los usuarios en el uso del sistema, que también es llamado “Tiempo promedio de registro de documentos”. Es por esa razón que es el tiempo que transcurre entre el ingreso al módulo de documentos y la clasificación de la misma.

- La fórmula para el tiempo promedio de registro de documentos.

$$TPR = \sum_{i=1}^n TR$$

TPR = Tiempo promedio para clasificar un documento.

TR : Tiempo utilizado en la clasificación de documentos

n : Cantidad de documentos analizados

1.3.6. Deep Learning

Es aprendizaje profundo comprende un área dentro del machine learning que compone de procesos como redes neuronales, inteligencia artificial, optimización, son un desarrollo revolucionario para predictores potentes. En la base están dos estilos básicos de aprendizaje a partir de la información supervisada y no supervisada (Villaseca, 2017) (como se cita a Lewis, 2016).

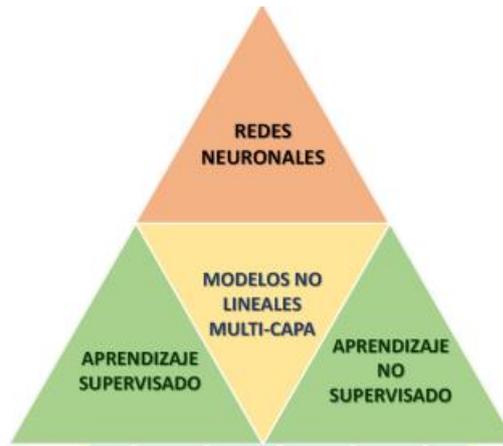


Figura 1 : La pirámide Deep learning

1.3.7. **NLTK**

Es la plataforma líder para crear aplicaciones en Python con el procesamiento de lenguaje natural. Nos proporciona interfaces con más de 50 corpus etiquetados y recursos léxicos como WordNet. Algunas de las herramientas que proporciona para la clasificación son tokenización, derivación, etiquetado análisis y razonamiento semántico. Es el más adecuado para los estudiantes, ingenieros, empresas, investigadores y usuarios.

NLTK está creado desde el 2001 como parte de un curso de informática de una institución educativa Universidad de Pennsylvania. Desde entonces se está ampliando gracias a su código abierto y a la comunidad. (Loper E, pp. 63-70, 2002).

1.3.8. **OCR**

Sanchez (2009) mencionó, el OCR que es la tecnología para el reconocimiento de caracteres, contiene una serie de procesos basados en la predicción. El formato de cómo se representan las letras es mediante símbolos ASCII. Las comparaciones y las unidades tratables son procesos que ayudaran a intercambiar pesos entre sí para determinar en forma práctica las diferencias que existen entre el alfabeto, los números y los códigos. Existe una probabilidad alta de que, la precisión que se obtiene al tratar de identificar los símbolos de un conjunto de datos enviados no se puedan reconocer al 100%. En tal sentido, se puede afirmar que cuantos más tipos de símbolos existan en un conjunto de datos es mayor el porcentaje de error de la clasificación.

1.3.9. Perceptrón multicapa

Barahona (2014) dio a entender que se desprende la arquitectura Feed-forward, tiene un patrón de entrada con una gran cantidad de neuronas que se activan posee una capa oculta. Una deficiencia del algoritmo es que es de aprendizaje lento, porque una alteración a los pesos genera un cambio en toda la red perdiendo el efecto de los ciclos previos.

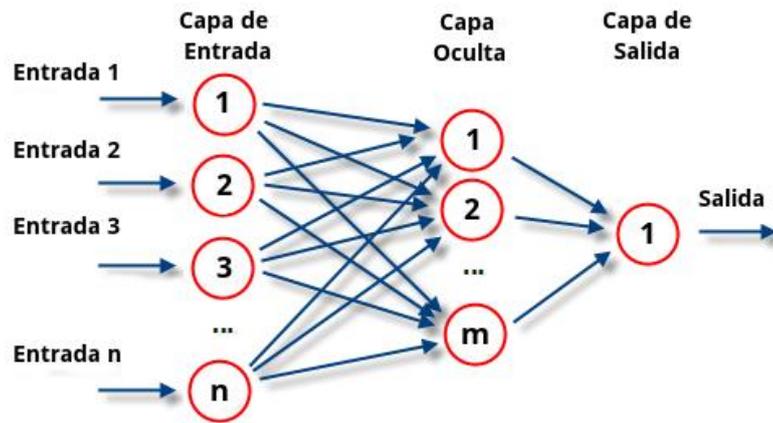


Figura 2: Perceptron multicapa

1.3.9.1. Regla de aprendizaje de Perceptron

Se realiza de acuerdo a la cantidad de neuronas (Collins, M. (2002, julio) la deducción matemática de una capa de entrada está representando por la siguientes ecuación.

q : Equivalente al número de vectores.

m : Equivales a la cantidad de neuronas sobre la capa oculta.

i : Equivalente a la cantidad de neuronas en la capa de salida.

Cuando un patrón de entrenamiento, esta propagada se produce una entrada n en cada una de las neuronas de la capa siguiente, la entrada de la neurona j .

$$n_j^0 = \sum_{i=1}^q W_{ji}^0 P_i + b_j^0$$

W_{ji}^0 : Peso de la neurona del componente i de la entrada de la neurona j

P_i : Vector p , componente i del que contiene el patrón de entrenamiento de q .

b_j^0 : Capa oculta de ganancia de la neurona.

Calculo del error debido al patrón p .

$$ep^2 = \frac{1}{2} \sum_{k=1}^s (S_k)^2$$

ep^2 : Error medio cuadrático para cada patrón de entrada p .

S_k : Error de la neurona k de la capa de salida con l neuronas.

1.3.9.2. *Red Bayesiana*

La red bayesiana es un generador estadístico de probabilidades, determinado en un conjunto de arcos y nodos ambas partes son los componentes de un grafo. El enfoque bayesiano es uno de los métodos más útiles en el procesamiento de lenguaje natural estadístico. (Mannig Christopher D. (1999, pág. 202).

1.3.9.3. *Naives Bayes*

El modelo presentado por (Sanchis, 2004) es un clasificador ampliamente utilizado para el aprendizaje debido a su gran eficiencia y su capacidad de combinar pruebas de una gran cantidad de características. En nuestro caso se estimaría la evaluación de las palabras del documento con el teorema de Bayes. Se tiene un conjunto de documentos $D = \{d_1, d_2, d_3, \dots, d_n\}$ que están relacionadas a las clases definidas $C = \{c_1, c_2, c_3, \dots, c_n\}$, cada documento es representado por un vector $\vec{d}_j = \{w_1, w_2, w_3, \dots, w_{|T|j}\}$ donde T es el conjunto de términos que pertenece a c_i dado el documento \vec{d}_j .

$$P(c_i|\vec{d}_j) = \frac{P(c_i) \cdot P(\vec{d}_j|c_i)}{P(\vec{d}_j)}$$

Para el cálculo de $P(c_i|\vec{d}_j)$ se asume como características independientes a los datos analizados, un supuesto que no cumple con la realidad. De lo anterior mencionado, mejora la simplicidad de los cálculos complejos, permitiendo combinar características por cada palabra, en nuestro caso T es la dimensión del vector de características.

$$P(\vec{d}_j|c_i) = \prod_{t=1}^T P(\vec{W}_{tj}|c_i)$$

La clase c_i puede ser utilizada para el cálculo de la estimación de la división de los documentos de entrenamiento de $P(c_i)$ para $P(c_i)$.

$$P(c_i) = \frac{N_j}{N}$$

Donde N es el número total de documentos, N_j el número de documentos para la clase c_i .

La predicción $P(\vec{d}_j|c_i)$ para $P(\vec{d}|c_i)$ está dada por:

$$P(\vec{d}_j|c_i) = \frac{1 + N_{ij}}{M + \sum_{k=1}^M N_{ik}}$$

N_{ij} es el contador de palabras del conjunto i , en los documentos de clase c_i . Se agrega + 1 para que la probabilidad no resulte Cero. M es el número de términos en el vocabulario.

1.3.9.4. ***K-Vecinos más cercanos***

(Fac.Ingenieria, 2008), menciona que es un método para clasificar textos mediante el análisis, este algoritmo no tiene fase de entrenamiento los K-NN vecinos necesitan una definición a priori del número de K-vecinos K , en resumen consiste en estimar el valor de un dato desconocido a partir de las características del dato más próximo.

La fórmula de la distancia euclidiana entre dos vectores $d = d_1, d_n$ y $e = e_1, e_n$, en se obtiene con el coseno entre dos vectores.

$$d(d, e) = \sqrt{(d_1 - e_1)^2 + (d_2 - e_2)^2 + \dots + (d_n - e_n)^2} = \sqrt{\sum_{i=1}^n (d_i - e_i)^2}$$

Para encontrar el coseno entre dos vectores se utiliza

$$d(d, e) = \frac{\sum_{i=1}^n (d_i \cdot e_i)}{\sqrt{\sum_{i=1}^n d_i^2 \cdot \sum_{i=1}^n e_i^2}}$$

1.3.9.5. *K-Modelo oculto de Markov*

(Sonnhammer, 1998) menciona, es una herramienta para calcular probabilidades, los valores pueden ser discretas, valores reales, o cualquier dato. Por tal motivo se asume que las observaciones son muestreadas en intervalos que resulta un valor entero.

Poseen una capa oculta donde concurren dos procesos estocásticos, el primero de ellos ocurre de en la capa oculta y el segundo proceso está asociado a un estado automático el cual emite un resulta de un conjunto de premisas por tanto se asume que cada estado satisface a las propiedades de markov.

Formulación de Markov $Q = \{q_1, q_2, \dots, q_n\}$ es el conjunto de estados $a : Q . Q \rightarrow [0,1]$.

$$\sum_{q_j \in Q} a(q_i, q_j) = 1 \quad \forall q_i \in Q$$

1.3.10. **TensorFlow**

Fritz (1997) mencionó, TensorFlow es un framework de código abierto para el estudio de procesos automáticos, desarrollado por la comunidad de Google para realizar sus procesos de sistemas, estas herramientas son capaces de emular redes neuronales para clasificar, predecir correlaciones y patrones, que son producto del estudio del pensamiento usados por los humanos. En la actualidad es utilizado para la implementación de estudios en los problemas actuales, Google continuamente está cambiando el rol de su predecesor de código privado, DistBelief. TensorFlow que fue originalmente implementado por el team de Google Brain para el uso de Google antes de ser de licencia de código abierto.

1.3.10.1. *Python*

Clavijo, Bernal y Silva (2006) mencionaron, Python es un lenguaje multiparadigma. Lo que significa que es fácil de implementar porque permite varios estilos de programación, la herramienta está orientada a objetos, tiene desarrollo imperativo y funcional. Algunos paradigmas son soportados gracias al uso de las extensiones. Es tipado, dinámico por lo que se administra de manera eficiente en la memoria.

1.4. Formulación de Problema

1.4.1. Problema General:

¿En qué medida permite mejorar la gestión documental la implantación del modelo de clasificación automática de documentos en Grupo Coril?

1.4.2. Problemas Específicos:

¿En qué medida permite la implantación del modelo mejorar el porcentaje de precisión de documentos clasificados?

¿En qué medida la aplicación permite disminuir el tiempo para gestión de clasificación de documentos?

1.5. Justificación del Estudio

1.5.1. Justificación teórica

Esta investigación también ayudara a divulgar los conceptos relacionados al aprendizaje, (Gonzales, 2004, Pag.573) menciona que lenguaje de procesamiento natural tienen como modelo asociado los métodos interpretativos, deductivos, variables de entrada dentro del ámbito investigativo, por lo que este trabajo, bien puede servir de base para el desarrollo de otras aplicaciones similares.

(Gelpunk, 2009) menciona, existe una gran cantidad de información digital no procesada, el material es de naturaleza experimental e investigativa algunos temas son la interacción de los proteínas y los genes, lo que resulta en un trabajo tedioso sin el procesamiento del lenguaje natural resultando imposible para una persona poder valerse de este volumen de información, por lo que resulta beneficioso el lenguaje de procesamiento natural, además de ser apropiado para los grandes volúmenes de información.

1.5.2. Justificación práctica

Se considera una justificación práctica porque automatiza los procesos con nuevas herramientas de procesamiento de lenguaje natural, (Gonzales, 2004, Pag.573), menciona que el análisis del contenido ahorra tiempo, dinero, recursos humanos y recursos computacionales, a su vez este tiempo recuperado, se pueden invertir en otros usos, por

ejemplo tareas dedicadas al negocio de la empresa, finalmente el procesamiento de la información rompe la forma tradicional de tomar requerimientos en la actualidad.

1.5.3. **Justificación Metodológica**

La investigación metodológica del estudio propone un nuevo método para la clasificación documentaria, (Nazar, 2017, Pag.14), menciona que la metodología para la detección y clasificación se basa en el diseño de un algoritmo que entrena con los resultados. En nuestra investigación de forma concreta el clasificador toma las oraciones contenidas y abstrae durante el aprendizaje aquellas características para la clasificación automática.

1.5.4. **Justificación social**

Este proyecto de investigación servirá a la sociedad para poder brindarles una herramienta que facilite la clasificación de diferentes textos utilizando lenguaje de procesamiento natural el cual permitirá el registro de los datos al sistema documentario en un formato estructurado.

(Isidoro Gil, 2017) menciona, el futuro próximo se tiene que estar atento a la evolución del procesamiento del lenguaje natural porque este tipo de investigación se lleva practicando hace décadas.

1.6. **Hipótesis**

1.6.1. **Hipótesis General:**

HG: A través de la implantación del modelo de clasificación de documentos se mejorará la gestión documental en Grupo Coril.

1.6.2. **Hipótesis Específicos:**

H1: A través de la comparación de los resultados obtenidos por la clasificación se mejorará el porcentaje de precisión documentos clasificados.

H2: El uso de la aplicación reducirá el tiempo empleado para gestión documentaria en Grupo Coril.

1.7. Objetivos

1.7.1. Objetivo General

Implantación de un modelo de Clasificación de textos que permitan analizar documentos en idioma español y determine su clasificación para mejorar la gestión documental en Grupo Coril.

1.7.2. Objetivos Específicos

Comparar los resultados obtenidos por clasificación automática para determinar el porcentaje de precisión documentos clasificados.

Reducir el tiempo de clasificación de documentos para gestión documentaria en Grupo Coril.

II. METODO

2.1. **Diseño de Investigación**

La investigación realizada es de tipo experimental que corresponde al diseño experimental, se caracteriza por la manipulación intencionada de la variable independiente y se analizan las consecuencias que la manipulación tendrá sobre una o más variables dependientes en términos generales.

2.1.1. **Tipo de Investigación**

El tipo de investigación del proyecto es aplicada, porque tiene el objetivo de resolver el problema de la clasificación automática de textos. A su vez es descriptiva porque explica y realiza un estudio profundo del tema.

(Hedrick et al., 1993). Menciona, los diseños experimentales tienen el mismo propósito de corroborar la relación existente entre dos o más variables. Cuando es imposible la selección de una muestra aleatoria, entonces los experimentos nos permiten calcular el impacto del programa o la metodología, como requerimiento se necesita que entre las variables se llegue a establecer la comparación necesaria. (p. 58).

2.2. **Variables, operacionalización**

- Variable independiente: Implantación de un modelo.
- Variable dependiente: Clasificación automática de textos

Tabla 1

Matriz Operacional de la Variable

Variable	Definición conceptual	Dimensión	Indicadores	Escala de Medición
VARIABLE DEPENDIENTE: Clasificación automática de textos.	Su finalidad es poder localizar con facilidad los documentos con características determinadas, se da mediante la detección de palabras claves de forma específica que lo tienen los dimensionadores del vocabulario lingüístico. (Figuerola 2013, Pág. 2)	Precisión	Porcentaje de la precisión de documentos clasificados.	Nominal
		Tiempo	Tiempo promedio de registrar un documento	Intervalo
VARIABLE INDEPENDIENTE: Implantación de un modelo.	Cuando una tecnología revoluciona la forma tradicional de cómo se desarrollan los procesos, entonces estamos en el último eslabón de la implementación de software. (Clavijo, Bernal y Silva (2006).	La implantación del método de clasificación automática de textos permitirá mostrar lo útil que es en tareas de extraer información de los documentos escaneados. Esta aplicación tiene un porcentaje de servicios web para mostrar los resultados obtenidos.		

Tabla 2
Indicadores

Indicadores	Descripción	Objetivo	Instrumento	Unidad de medida	Formula
Porcentaje de la precisión de documentos clasificados.	Determina el porcentaje de la precisión que tiene de la clasificación de texto al clasificarse por el usuario y luego por el algoritmo.	Aumentar el porcentaje de precisión en la clasificación de documentos.	Ficha de observación	Porcentaje	$PC = \left(\frac{DC}{DU}\right) * 100$ <p>PC: Porcentaje de precisión de documentos. DC: Cantidad de documentación clasificados correctamente. DU: Cantidad de documentos analizados.</p>
Tiempo promedio de registrar un documento.	Permite gestionar el tiempo de la gestión documentaría.	Disminuir o eliminar el área de gestión documentaria manual.	Cronometro. Ficha de Observación	Tiempo	$TPR = \sum_{i=1}^n TR$ <p>TPR = Tiempo promedio para clasificar un documento. TR: Tiempo utilizado en la clasificación de documentos n: Cantidad de documentos analizados</p>

2.3. Población y muestra

2.3.1. Población

Según Judith Scharager (2001, pág. 1), “son llamadas también muestras dirigidas o intencionales”, son seleccionadas con mecanismos informales lo que implica que es imposible calcular el porcentaje de la precisión del error estándar en nuestra población, las muestras no probabilísticas tienen carencia teórica, son consideradas poco minuciosas, utilizadas frecuentemente cuando nuestro estudio está dirigido a un grupo muy específico.

Para el desarrollo de esta investigación, la población estará conformada por el número de documentos gestionados en un mes en la empresa Grupo Coril donde se utilizará tres algoritmos de tipo finito Navies Bayes, K vecinos y el modelo oculto de Markov, todos ellos serán escogidos para ser hacer pruebas de probabilidad y especificidad que brindara el indicador de porcentaje de precisión.

Tabla 3

Número de documentos gestionados en un mes

Día de la semana	Día1	Día 2	Día 3	Día 4	Día 5	Total
1	0	0	89	37	47	173
2	31	52	0	0	26	109
3	60	37	20	31	0	148
4	0	97	37	43	53	230
5	68	33	24	50	58	233
Total	159	219	170	161	184	893

2.3.2. Muestra

Blanca Arteaga. (2017), menciona que los resultados de la muestra tienen su fundamento en la estadística inferencial según lo mencionado para la investigación se conoce el tamaño de la población y se procede a obtener el tamaño de la muestra.

$$n = (Z^2 * N) / (Z^2 + 4 * N * EE^2)$$

n = Es tamaño de la muestra.

N = tamaño de la población

$$Z = \text{Nivel de Confianza al 95\% (1.96) elegido } Z = \frac{EE}{2} = \frac{0.05}{2} = 0.025$$

$$Z_{0.025} = 1.96$$

normal	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-4	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.0001	0.0001	0.0001	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.0002	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.0003	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.0004	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.0006	0.00058	0.00056	0.00054	0.00052	0.0005
-3.1	0.00097	0.00094	0.0009	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.001
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.0024	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.0028	0.00272	0.00264
-2.6	0.00466	0.00453	0.0044	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.0057	0.00554	0.00539	0.00523	0.00508	0.00494	0.0048
-2.4	0.0082	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.0099	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.0139	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.0116	0.0113	0.01101
-2.1	0.01786	0.01743	0.017	0.01659	0.01618	0.01578	0.01539	0.015	0.01463	0.01426
-2	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.0197	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.0268	0.02619	0.02559	0.025	0.02442	0.02385	0.0233

Figura 3 Tabla de Distribuciones

EE = error estimado al 5%

Reemplazando los datos de la formula

$$n = (1.96^2 * 893) / (1.96^2 + 4 * 893 * 0.05^2) = 269$$

Por lo tanto el tamaño de la muestra para el presente estudio es 269 documentos.

2.3.3. Muestreo

Alaminos, A., & Castejón, J. L. (2006), mencionan que una muestra probabilística es ciega al seleccionar el número de elementos muestrales, mientras que la no probabilística intencional se requiere conocer previamente las características de los individuos que forman la muestra, por lo que requiere un conocimiento elevado de la población que se quiere estudiar.

El muestreo de nuestra investigación será no probabilístico, debido a que desarrollara métodos de clasificación de textos mencionados en la población, que pueda reducir el grado de complejidad de nuestra muestra de documentos seleccionados.

2.4. Técnicas e instrumento de recolección de datos

Cuando el planteamiento de la investigación es coherente con el marco conceptual se inicia el proceso de recolección de datos el cual es siempre posible crear un nuevo instrumento de recolección de datos específicamente.

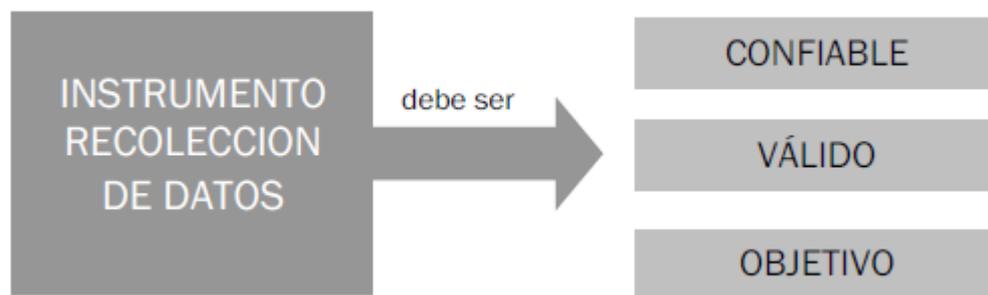


Figura 4: Instrumentos de recolección de datos

¿Qué tipo de datos nos permiten responder a nuestros objetivos? , ¿Cuál nuestro input donde obtenemos la información?, ¿A través de que método se puede recolectar datos?

- Experimentos: Constituyen el modo de recolección más clásico en la investigación científica, los experimentos son una manera directa, precisa, confiable y muy valiosa de recolectar datos, por lo que se utilizar los algoritmos empaquetados en las librerías de Python, también se utilizara un conjunto de entrenamiento de textos clasificados en español para respaldar la información sobre nuestras variables de mejora de precisión.
- Cronometro: Es un medidor que permite la medición de una pequeña fracción de tiempo. En la presente investigación será la herramienta utilizada para verificar el tiempo promedio de clasificación de un documento.

2.5. Métodos de análisis de datos

2.5.1. Procesamiento de datos

Hair (1999, pág. 201). Menciona que la clasificación de la información se exhibe mediante distribuciones de frecuencias de las variables cuantitativas. La información de partida ahora es una matriz de dimensión $I \times J$, que representan las frecuencias absolutas. En resumen el método de

análisis de datos es un conjunto de número positivos dispuesto en una matriz. I una manera de llegar a una tabla de contingencia es $I \times J$.

Tabla 4

Frecuencia de Clasificación de algoritmos

Descripción de Algoritmos	K-Vecinos	Bayesianos	Markov
C1	0.92	0.82	0.5
C2	0.95	0.7	0.65
C3	0.927	0.95	0.91
Promedio	0.93233333	0.82333333	0.68666667

2.5.2. **Definición de Variables**

I_a = Indicador de la aplicación actual

I_p = Indicador de la aplicación propuesta

2.5.3. **Hipótesis Estadística**

Hipótesis **H0**: La implantación del modelo de clasificación de documentos no mejora la gestión documental en Grupo Coril.

Hipótesis **Ha**: La implantación del modelo de clasificación de documentos mejora la gestión documental en Grupo Coril.

Hipótesis Específicas:

Hipótesis **H10**: La comparación de los resultados obtenidos por la clasificación no mejora el porcentaje de precisión documentos clasificados en Grupo Coril.

$$H1_0: I1_a \leq I1_p$$

Hipótesis **H2a**: La comparación de los resultados obtenidos por la clasificación mejora el porcentaje de precisión documentos clasificados en Grupo Coril.

$$H1_a: I1_a > I1_p$$

Hipótesis H3₀: La aplicación no reduce el tiempo de clasificación de documentos en la gestión documental para la empresa Grupo Coril.

$$H2_0: I2_a \leq I2_p$$

Hipótesis H3_a: La aplicación reduce el tiempo de clasificación de documentos en la empresa Grupo Coril.

$$H2_a: I2_a > I2_p$$

2.5.4. Nivel de Significancia

Nivel de significancia (α): 0.05 Nivel de confianza ($\gamma = 1-\alpha$): 0.95

2.5.5. Pruebas de Normalidad

“Cuando la muestra es menor, $n < 50$ se aplica la prueba de shapiro-Wilk para la prueba de normalidad. Esta prueba ayuda a calcular la estadística de prueba W, que si llega a ser mayor al nivel de significancia o se asume que la distribución es normal, si no la distribución es no normal”.

2.5.6. Pruebas de hipótesis Estadísticas

Se basan en la respuesta de la muestra aleatoria, en una población de estudio la cual nos guía por un proceso para aceptar o rechazar la hipótesis presentada.

2.5.7. Interpretación de datos

Se tiene la distribución de frecuencias absolutas, cuadros estadísticos, gráficos de superficie de barra para la comparación entre métodos de clasificación y grafica de líneas para medición de los datos.

2.6. Aspectos éticos

En la presente investigación se podrá obtener información de artículos científicos y revistas indexadas con respecto al estudio de algoritmos y métodos de clasificación de textos.

III. RESULTADOS

3. RESULTADOS

En este capítulo se describen los resultados de la recolección de documentos empleados para evaluar el método propuesto. Haciendo uso de los indicadores “Porcentaje de precisión de la clasificación obtenida” y “Tiempo promedio de registrar un documento”, además se explicara el desarrollo de aplicación en la empresa Grupo Coril y la comparación de la tasa de aciertos con los métodos de clasificación.

3.1. Análisis Descriptivo

Para el análisis descriptivo se utilizó el ERP Sinergia para evaluar el tiempo promedio de clasificación de documento y para el porcentaje de precisión de documento se utilizó el método clasificación del algoritmo bayesiano, para este resultado se aplicó el pre-test el cual permitió conocer las variables iniciales; en un segundo momento se implementó el sistema de clasificación automática de textos un post-test realizando una nueva evaluación al tiempo de registro de documentos y el porcentaje de precisión de clasificación documentaria.

Primer Indicador: Porcentaje de la precisión de documentos clasificados correctamente.

Técnica de estimación de rendimiento el cual se utilizara la validación cruzada, la precisión del modelo será estimado por el total de documentos clasificados de manera correcta sobre la totalidad documentos. El proyecto se ha definido con un total de 269 documentos clasificados por un especialista. A continuación se muestra en la tabla 5 los resultados de la precisión para la clasificación de documentos.

Tabla 5

Cuadro de validación cruzada para la selección del modelo

	Red Bayesiana Instancias clasificadas de manera correcta	KNN(k vecinos)	Cadenas de Markov
Test	82.07 %	0%	0%

Tabla 6

Indicador Precisión de clasificación

	N-Documentos	Erro de Clasificación	Desviación estándar	Varianza
Pre-test	269	65	0.420	0.176
Post-Test	269	4	0.121	0.15

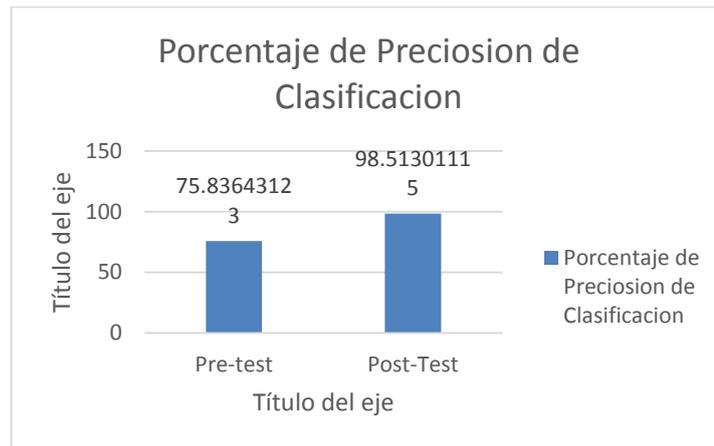


Figura 5 Porcentaje de Precisión de Clasificación

En el caso de la precisión de documentos clasificados correctamente, se tiene en el pre-test un valor a 65 errores. En caso de post-test se obtuvo un valor de 4 errores de clasificación, esta diferencia muestra la diferencia antes y después de la implantación del método de clasificación de textos.

Como se puede observar, la clasificación de datos cruzada el rendimiento del método de clasificación de una red bayesiana es la más apropiada para la clasificación de documentos.

Segundo Indicador: Tiempo Promedio de registrar documentos.

Los resultados obtenidos del tiempo promedio de registrar documentos

Tabla 7

Tiempo Promedio de Registro de documentos

	N-Documentos	Tiempo en segundos	Desviación estándar	Varianza
Pre-test	269	715.16	34.7561	1207.99
Post-Test	269	358.01	33.6198	1130.29

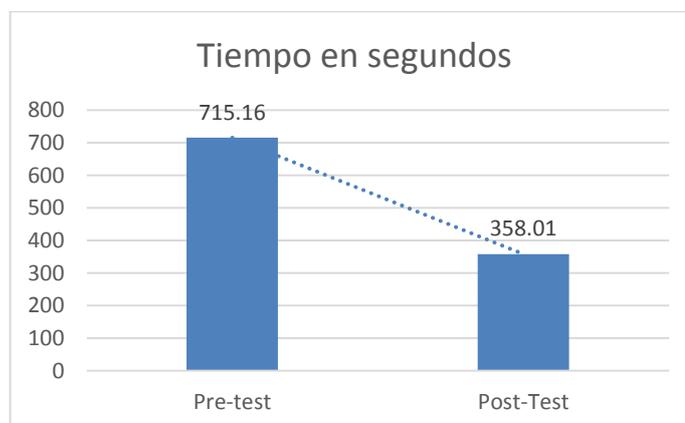


Figura 6 : Indicador de tiempo en Segundos

En el caso del tiempo de registro de documentos de las pruebas pre-test de la muestra se tiene un valor de 715.16 segundos por mantenimiento. En el caso de post-test se obtuvo un promedio de 358.01 segundos, esta información muestra la diferencia antes y después de la implantación del método de clasificación de textos.

3.1.1. Análisis Inferencial

Pruebas de normalidad

Los datos de las muestras para los indicadores precisión de clasificación y tiempo, fueron examinados mediante pruebas de normalidad, determinando la hipótesis a realizar.

En el indicador de Porcentaje de precisión del resultado obtenido por la clasificación la muestra es de 269 documentos, que es mayor a 50, se ejecutaron las pruebas de normalidad de Kolmogorov-Smirnov dado a que es un indicador de porcentaje, dicha prueba se realizó utilizando los algoritmos de clasificación en el software PYTHON y para el indicador de “Tiempo promedio de registrar un documento”, la muestra es de 269 documentos el cual supera las 50 muestras, se ejecutaron las pruebas de normalidad de Kolmogorov-Smirnov.

Para efectos de cálculos se utilizó el software IBM SPSS Statics Version 23 con un nivel de confiabilidad del 95% bajo las siguientes premisas:

Si $\text{sig} \geq 0.05$ se trata de una distribución normal.

Si $\text{sig} < 0.05$ se trata de una distribución no normal.

De la variable $\text{sig} =$ nivel crítico.

Primer indicador: El porcentaje de precisión de documentos clasificados, tiene el objetivo de seleccionar nuestra prueba de hipótesis; para someter los datos a un indicador de clasificación documental el cual tiene la distribución normal.

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Pre-test Calsifico	,479	269	,000	,517	269	,000
Post-test Clasifico	,534	269	,000	,097	269	,000

Figura 7: Kolmogorov- Precisión

Para la muestra mayor a 50 datos se toma la prueba de normalidad de kolmogorov Smirnov para el P-Valor (Pre-Test) precisión de documentos clasificados correctamente antes del sistema 0.00 el cual es menor al valor de 0.05 que es el nivel de insignificancia. Para el P-Valor (Post-Test)

precisión de documentos clasificados correctamente con el nuevo sistema 0.00 es menor al valor 0.05. En ambos casos la variable porcentaje de precisión de registro de documentos es menor al nivel de significancia por lo que se concluye que no es una distribución normal. Por lo se confirma la distribución no normal los cuales se pueden apreciar en el gráfico de figura 8 y la figura 9.

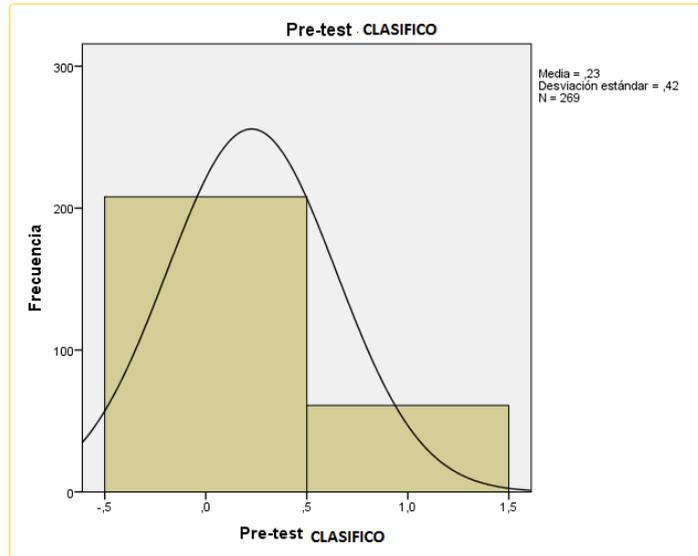


Figura 8: Pre-Test Porcentaje de Precisión

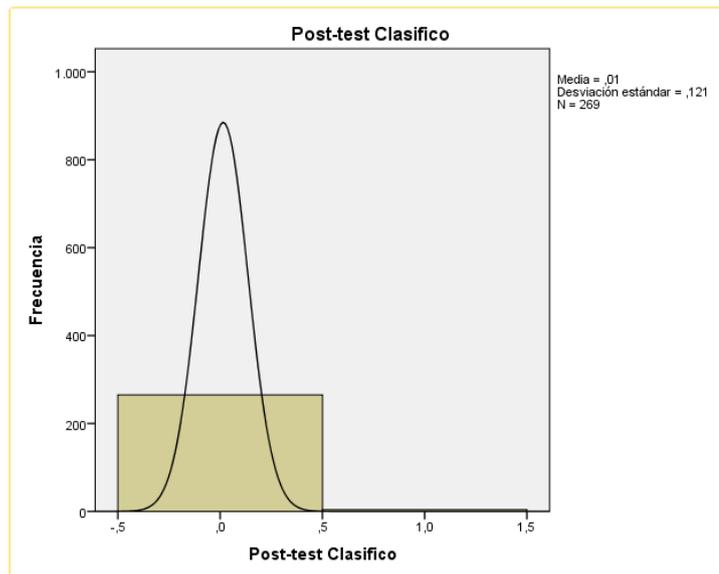


Figura 9: Post-Test Porcentaje de Precisión

Segundo indicador: **Tiempo Promedio para registrar un documento** tiene el objetivo de seleccionar la prueba de hipótesis; los datos fueron sometidos a un indicador de tiempo promedio el cual tiene la distribución normal.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Pre -test Tiempo Segundos	,081	269	,000	,950	269	,000
Post -test Tiempo Segundos	,086	269	,000	,957	269	,000

Figura 10: KOLMOGOROV-SMIRNOV

En la figura 5 el valor de Sig. Para la muestra mayor a 50 datos se toma la prueba de normalidad de kolmogorov Smirnov para el P-Valor (Pre-Test) tiempo de registro de documentos antes del sistema 0.00 el cual es menor al valor de 0.05 que es el nivel de insignificancia . Para el P-Valor (Post-Test) tiempo de registro de documentos con el nuevo sistema 0.00 es menor al valor 0.05. En ambos casos la variable tiempo de registro de documentos es menor al nivel de significancia por lo que se concluye que no es una distribución normal. Por lo se confirma la distribución no normal los cuales se pueden apreciar en el gráfico de figura 11 y la figura 12.

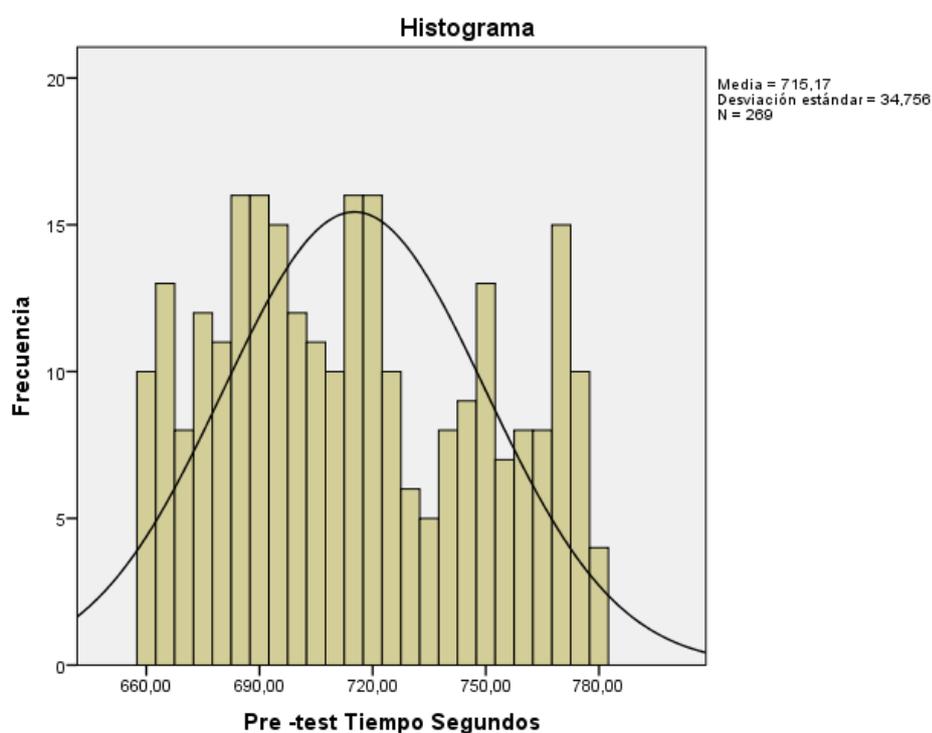


Figura 11: Tiempo Pre-test Registro de Documento

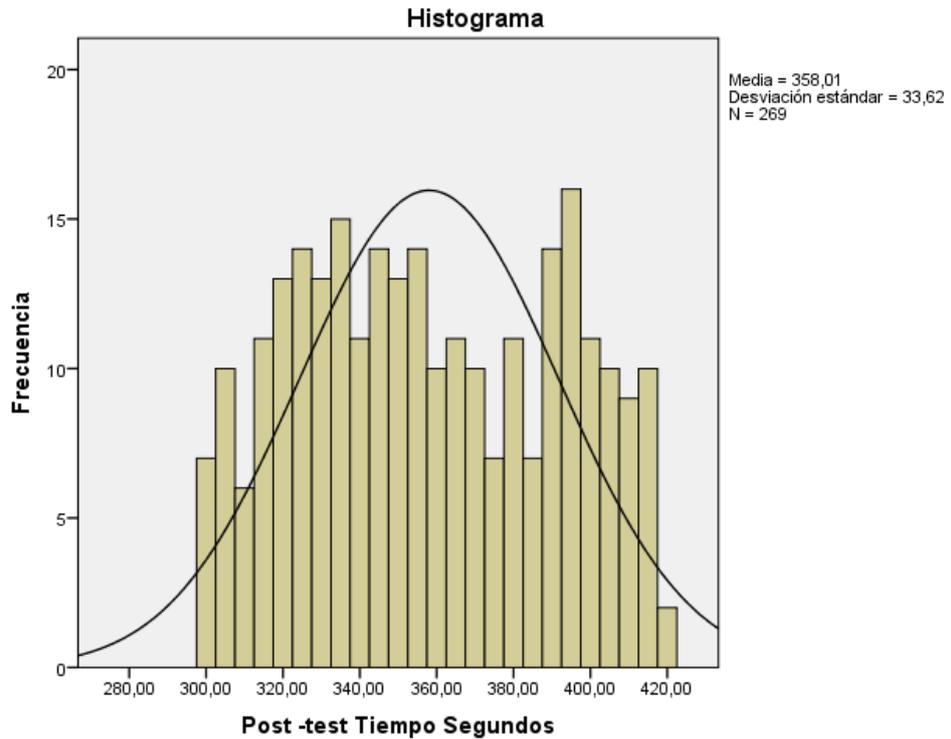


Figura 12: Tiempo Promedio en Registrar un Documento Post-Test

3.2. Prueba de Hipótesis 1

La implantación de modelo de clasificación de documentos aumenta el porcentaje de precisión de documentos clasificándolos de manera efectiva en la empresa GRUPO CORIL.

Indicador: Porcentaje de la precisión de documentos clasificados correctamente.

Hipótesis Estadística

Definición de variables:

PDCCa: Porcentaje de la precisión de documentos clasificados correctamente antes de la implantación del modelo de clasificación de documentos.

PDCCd: Porcentaje de la precisión de documentos clasificados correctamente después de la implantación del modelo de clasificación de documentos.

Hipótesis H₂₀: El modelo de clasificación disminuye el porcentaje de precisión de documentos clasificados correctamente en empresa Grupo Coril.

$$H_{20}: PDCC_a \leq PDCC_p$$

Sea: El indicador de la aplicación actual mejoro el indicador de la aplicación propuesta.

Hipótesis H_{2a}: El modelo de clasificación aumenta el porcentaje de precisión de documentos clasificados correctamente en empresa Grupo Coril.

$$H_{3a}: PDCC_a > PDCC_p$$

Sea: El indicador de la aplicación propuesta mejoro el indicador de aplicación actual.

Prueba de rangos con WilCoxon

Las pruebas de normalidad que se encuentran en la figura 6 y figura 7 de los resultados que se obtuvieron se notó que son de distribución no normal en el Pre-test como Post-Test por lo que se aplicara la prueba de Wilcoxon, a continuación los resultados se muestran:

Prueba de rangos con signo de Wilcoxon

Rangos				Estadísticos de prueba ^a	
		N	Rango promedio	Suma de rangos	
Post-test Clasifico - Pre-test Calsifico	Rangos negativos	61 ^a	33,00	2013,00	Post-test Clasifico - Pre-test Calsifico
	Rangos positivos	4 ^b	33,00	132,00	Z
	Empates	204 ^c			Sig. asintótica (bilateral)
	Total	269			

- a. Post-test Clasifico < Pre-test Calsifico
 b. Post-test Clasifico > Pre-test Calsifico
 c. Post-test Clasifico = Pre-test Calsifico

- a. Prueba de rangos con signo de Wilcoxon
 b. Se basa en rangos positivos.

Si sig < 0.05 Se acepta la prueba de hipótesis alternativa.

Si sig >= 0.05 Se rechaza la prueba de hipótesis alternativa.

Validación de la hipótesis

Al aplicar la prueba de rangos de Wilcoxon, se aprecia que el valor de sig. Es de 0.000 el cual es menor que 0.05. La hipótesis alternativa es aceptada. La implantación del modelo de clasificación de documentos mejora el porcentaje de precisión de documentos con probabilidad del 95%.

3.3. Prueba de Hipótesis 2

La implantación de modelo de clasificación de documentos automática reduce el tiempo promedio de la gestión documental para la empresa GRUPO CORIL.

Indicador: Tiempo promedio para registrar un documento.

Hipótesis Estadística

Definición de variables:

TPRDa: Tiempo promedio de registro de documentos antes de la implantación del modelo de clasificación de documentos.

TPRDd: Tiempo promedio de registro de documentos después de la implantación del modelo de clasificación de documentos.

IV. DISCUSIÓN

4. DISCUSIÓN

En los resultados de la investigación se compararon el porcentaje de precisión de documentos clasificados y el tiempo de la gestión documental en Grupo Coril.

1.- Se aplicó las pruebas de precisión de clasificación, en el sistema anterior los usuarios que participaron en la ficha de observación obtuvieron una precisión al clasificar de en 75.83%, cuando se implementó el método de clasificación automática aumento el porcentaje de precisión en un 98%.

Según la investigación Luis Alejandro Pinto Valdiviezo al utilizar el clasificador KNN obtuvo un porcentaje de instancias incorrectas clasificadas en un 79% de precisión. Con esta información se concluye que el método de clasificación influye en el porcentaje de precisión en base a la clasificación.

2.-El tiempo de la gestión documental que se evaluó en Grupo Coril durante el Pre-Test tiene un equivalente a 715 segundos, y con la implantación del método de clasificación automática se redujo a 358 segundos. Por lo tanto se puede determinar que la implantación del modelo de clasificación automática produjo una disminución 357 segundos.

Según la investigación realizada por Saavedra Rosales, Jeffer Jose Los resultados obtenidos de la gestión documentaria en el pre-test equivalen a 728 segundo que se redujeron a aproximadamente a 97 segundos. Con esta conclusión se demuestra que los sistemas automatizados influyen en la gestión del tiempo para el registro documentario.

V. CONCLUSIONES

5. CONCLUSIONES

Con la implantación del método de clasificación de textos automática se mejoró la gestión documental.

1.- Se concluye con las pruebas de precisión para clasificación documentaria, que los usuarios que participaron en la ficha de observación, utilizando el software anterior, obtuvieron una clasificación de 75.83% de precisión, cuando se implementó el método de clasificación automática aumento el porcentaje de precisión en un 98%.

2.- Se concluye que aplicando las pruebas de medición de tiempo en la gestión documental, los usuarios que participaron obtuvieron el promedio 715 segundos por registro, pero con la implantación del método de clasificación automática se redujo a un promedio de 358 segundos por registro lo que equivale al 50%.

3.- Por ultimo luego de presentar los siguientes resultados queda demostrado que disminuye de manera significativa el proceso del indicador de tiempo de registro de documentos y aumenta en un porcentaje significativo la precisión documentaria, por lo que se concluye que la implantación del método de clasificación de documentos en Grupo Coril - 2018 genera efectos positivos.

VI. RECOMENDACIONES

6. RECOMENDACIONES

Se recomienda migrar el alcance del aplicativo a entornos de impresoras, existen marcas como Xerox que de manera gratuita implementan herramientas de desarrollo como solución a los problemas de las empresas, para inscribirse no es necesario pertenecer a un grupo corporativo por tal motivo Xerox permite a todo tipo de desarrolladores puedan usar el api de servicios de la impresora.

Se recomienda ampliar las palabras claves para una mejor clasificación documentaria, investigar sobre otros algoritmos de precisión y clasificación.

Se recomienda mejorar el corpus lingüístico implementado en Grupo Coril.

Se recomienda entre los principales trabajos para el futuro incrementar la robustez de la entrada de los datos analizando y explotando al máximo los algoritmos de clasificación.

La clasificación automática de textos está ganando mucha importancia por la masiva cantidad de documentos en formato digital. Como se mostró en capítulos anteriores los métodos de clasificación existentes pueden competir con el desempeño de los empleados, estas tareas pueden ser automatizadas ahorrando tiempo y dinero.

En resumen se recomienda en el campo del procesamiento de lenguaje natural, investigar respecto a temas como la minería de datos, polaridad de opiniones, Deep learnig, etcétera. La arquitectura expuesta posee una fuerte carencia del entendimiento del lenguaje de procesamiento de máquinas así que se tiene mucho que explorar.

Bibliografías

- Alaminos, A., & Castejón, J. (2006). Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión. *Ingeniero de Sistema*. Universidad de Alicante, España.
- Ale, J. (2013). *Procesamiento de Lenguaje Natural en Sistemas de Análisis de Sentimientos*. Buenos Aires Argentina, Argentina.
- Amendola, L. (2003). INDICADORES DE CONFIABILIDAD. *Universidad Politécnica Valencia España*. Obtenido de http://www.mantenimientoplanificado.com/Articulos%20gesti%C3%B3n%20mantenimiento_archivos/indicadores%20confiabilidad%20amendola.pdf
- Araujo, N. (2009.). *Método Semisupervisado para la clasificación Automática de Textos de Opinión*. Instituto Nacional de Astrofísica, Óptica y Electrónica en Tonantzintla. Puebla.
- Arkaitz, M. (2018). *Aproximaciones a SVM semisupervisado multiclase para clasificación de páginas web*. Universidad de Educación a Distancia, España.
- Arteaga, B. (2017). Sobre la población y muestra en investigaciones empíricas. *France, Europe: Aula Magna*. doi:edsbas.65519B90
- Ascarza, F. (2018). *Segmentación automática de textos, mediante las redes neuronales convolucionales en imágenes de documentos históricos*. Universidad de Católica del Perú, Perú.
- Atuan Toledo. (2018). Categorización de comentarios, según la emoción expresada, utilizando procesamiento de lenguaje natural y lingüística computacional. Obtenido de <https://repositorio.usm.cl/bitstream/handle/11673/40150/3560900257789utfsm.pdf?sequence=1&isallowed=y>
- Barahona, D. (2016). *Perceptron Wavelert y algoritmos genéticos para la predicción de captura de la anchoveta*. Universidad Católica de Valparaíso., Brasil.
- Barahona, R. (s.f.). Perceptron wavelet recursivo y algoritmos genéticos para la predicción de captura de la anchoveta. 2. Obtenido de http://opac.pucv.cl/pucv_txt/txt-3000/UCE3375_01.pdf
- Barrios, A. (2016). *Sistema Web y móvil de Trámite Documentario para mejorar la Gestión de Movimientos de equipos Informáticos del área de soporte técnico en una Universidad*. Cesar Vallejo, Peru.
- Bermonti, M. (s.f.). Desarrollo y adaptación a formato computarizado de una intervención cognitiva para desarrollar destrezas de lectura. (A. f. Central., Ed.) doi:10150148
- Betancourt, S. (2012). Sistema de reconocimiento óptico de caracteres, como solución al procesamiento digital. *Ecuador*. Universidad Central del Ecuador.
- Castillo, J., & Cárdenas, M. (2017). In XIX Workshop de Investigadores en Ciencias de la Computación. Desarrollo de sistemas de análisis de texto.

- Cepeda, P. (2016). Programa de automatización de los procedimientos de trámite documentario en la calidad del servicio a los usuarios del Hospital Nacional Arzobispo Loayza. *Ingeniero de Sistemas*. Cesar Vallejo, Lima, Peru.
- Collins, M. (2002,, Julio)). Association for Computational Linguistics. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing. *10*, 1-8.
- Condori, R. (2014). *Método de Clasificación Automática de Textos basado en Palabras Claves utilizando Información Semántica: Aplicación a Historias Clínicas*. Universidad Nacional de San Agustín.
- Cruz, F. L., Troyano, J. A., Enriquez, F., & Ortega, J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del lenguaje natural*, *41*, 73-80.
- David, L. (2012). *Clasificación Automática de Documentación*. Santiago., Da Coruña.
- Egusquiza, X. (2015). Sistema Web para el proceso de gestión documental para la empresa de prevención Global. *Ingeniero de Sistemas*. Cesar Vallejo, Peru.
- Escobar, M. (2010). *Diseño e implementación de un sistema de apoyo al seguimiento de proceso judiciales para un estudio de abogados*. Universidad de Católica del Perú, Peru.
- Ferrerira, A. (2010). ELE-Tutor Inteligente: Un analizador computacional para el tratamiento de errores gramaticales en español como Lengua Extranjera. *ELE-Tutor Inteligente*, *43*.
- Figuerola, C. (2013). Clasificación automática de documentos. doi:FFI2011-27763
- Flores, E. (2017). Detección de reutilización de código fuente monolingüe y translingüe. ISSN: 1135-5948. Procesamiento del Lenguaje Natural. Obtenido de <http://hdl.handle.net/10045/64021>
- Flores, M. (2017). Sistema informático para el proceso de Tramite Documentario en la municipalidad de Chaclacayo. *Ingeniero de Sistemas*. Cesar Vallejo, Peru.
- Flórez, J., & Torres, H. (s.f.). Estrategia de regresión basada en el método de los k vecinos más cercanos para la estimación de la distancia de falla en sistemas radiales. En *Revista Facultad de Ingeniería* (págs. 100-108). España.
- Garrido, E. (2010). *Diseño e implementación de reconocimiento de números manuscritos*. Universidad Pontificia Católica, Peru.
- Gómez, B. (2017). *Implementación de un Sistema de Documentación bajo la plataforma Web para la gestión y control documental de la empresa Corporación JUJEDU E.I.R.L. – talaria 2017*. Universidad de Católica de los Ángeles de Chimbote, Chimbote, Peru.
- Gómez, M. (2014). *Reconocimiento de caracteres ópticos OCR por medio de la correlación y redes neuronales*. Universidad Pontificia Bolivariana, Bolivia.

- Gómez, P. (2016). La lucha contra el terrorismo y la delincuencia organizada: una visión desde la lingüística y la ingeniería del conocimiento. *Miscelánea*, 107-123,138-139. Obtenido de <https://search.proquest.com/docview/1860847317?accountid=37408>
- González, A. (1993). El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural. *Rev. Esp. Doc.* 27, 573.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). Análisis multivariante. *Prentice Hall*, 491.
- Herrera, B. (2018). *Implementación de un sistema Web de Gestión de tramite documentario para optimizar los procesos en el área de mesa de partes en la municipalidad distrital de Bellavista –Sullana*. Universidad de Católica de los Ángeles de Chimbote, Peru.
- INEI. (2015)). *Tecnología de información y comunicación en las empresas. Encuesta económica anual 201*. Obtenido de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digiales/Est/Lib1406/libro.pdf
- Jiménez, S. (2009). *Análisis y diseño de un sistema de tramite documentos de pago de proveedores vía Intranet*. Universidad de Católica del Perú, Perú.
- Liberato, A. N. (2014). *Desarrollo de un sistema de gestión documental, fichas de resumen y lista de publicación para el proyecto PROCAL-PROSER*. Universidad de Católica del Perú, Peru.
- Librado, H. (2016). *Análisis automático de opiniones de productos en redes sociales*. Politécnico Nacional en México, México.
- Loper, E., & S., B. (2002, Julio). NLTK: The natural language toolkit. In Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics. *Association for Computational Linguistics*, 1, 63-70.
- Losada, B., & Zapata, M. (2010). Transformación de Lenguaje Natural a Lenguaje Controlado Para La Educación de Requisitos a Partir de Documentación Técnica. *Cuaderno Activa*, 2, 9–14.
- Malaver, M. (2015). Aplicación de redes neuronales para determinar el pronóstico de las ventas en la empresa catering & buffets Mys ubicada en la ciudad de Piura. *Ingeniero de Sistemas*. Cesar Vallejo, Peru.
- Maldonado, A. (2018). Sistema Web para el proceso de Gestión Documental de la División de Seguridad de Contraminas – PNP-Rímac. *Ingeniero de Sistemas*. Cesar Vallejo, Peru.
- Manning, C. (1999). Foundations of Statistical Natural Language Processing. doi:E0123734
- Manuel, M. (2008). *Diseño y arquitectura de una Red Neuronal Artificial Perceptron Multicapa sobre un FPGA aplicada al reconocimiento de Caracteres*. Universidad de Católica del Perú, Peru.
- María, R. (2012). Un Framework de Ingeniería del lenguaje para el pre-procesado Semántico de Textos. *Maestría en Ingeniería de Software*.

- Martín Antonin, M. A., Gómez Guinovart, X., Gonzalo Arroyo, J., Llisterra Boix, J., Rafel i Fontanals, J., Rodríguez Hontoria, H., . . . F. (2003). Tecnologías del lenguaje. Libro colección de manuales. Obtenido de <http://www.digitaliapublishing.com/a/964/tecnologias-del-lenguaje>
- Maseda, F. (1997). Acceso a la información basado en su contenido la clave de la gestión documental de la empresa. *Revista general de información y documentación. Revista general de información y documentación. Universidad complutense de Madrid.* doi:ISSN 1132-1873
- Matich, D. (s.f.). Redes Neuronales: Conceptos básicos y aplicaciones. Cátedra de Informática Aplicada a la Ingeniería de Procesos–Orientación. Obtenido de <tp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf>.
- Monteserin, A., Armentano, M. G., & P. D. (2017, Octubre). Minería de argumentos en ambientes de aprendizaje colaborativo soportado por computadoras. In Brazilian Symposium on Computers in Education. *Simpósio Brasileiro de Informática na Educação-SBIE, 28(1457)*. Obtenido de <http://br-ie.org/pub/index.php/sbie/article/view/7676>
- Moran, L. (2017). *Sistema de Informático para la gestión documental para la empresa Hector Gonzales Sandi Agencia afianzada de Aduana S.A.* Cesar Vallejo, Peru.
- Obregón, M. (2016). Sistema web para la gestión documental de la empresa municipal inmobiliaria de Lima S.A. *Ingeniero de Sistemas.* Universidad César.
- Pastor, M. (2011). Un Modelo Lingüístico-Semántico Basado en emociones para la Clasificación de Textos según su Polaridad e Intensidad. *Ingeniero de Sistema.* Complutense de Madrid, Madrid, España.
- Paz, D. (2017). Herramientas Deep Learning para Predictive Analytics. Una aplicación. *Ingeniero de Sistemas.* Universidad de Concepción, Chile.
- Pérez, J. (2011). Red neuronal convolucional rápida sin fotogramas para reconocimiento de dígitos. (M. Reina, Ed.) doi:CP41012, 2011.
- Pinto, L. (2015). *Modelo algorítmico para la clasificación de documentos de carácter judicial en lenguaje portugués según su contenido.* Pontifica Universidad Católica del Perú, Peru.
- Pinto, M. (1998). Gestión de Calidad en Documentación. *Revista de Biblioteconomía y Documentación, 1,* 171-183. doi:1575 - 243
- Quispe, J. (2017). Relación entre la implementación de un sistema de trámite documentario y la gestión documentaria en la municipalidad del Rímac. *Ingeniero de Sistemas.* Universidad San Ignacio de Loyola, Peru.
- Rodriguez, A. (2016). Implementar un herramienta tecnológica de información para la producción y gestión de documentos en la organización Serdán. *Ingeniero Industrial.* Universidad Libre de Colombia, Colombia.
- Rodríguez, R. (2017). Redes Neuronales Artificiales. Conceptos básicos y algunas aplicaciones en Energía. Obtenido de <http://dspace.uib.es/xmlui/handle/11201/2511>

- Rogelio, N. (2017). Sociedad española para el lenguaje de procesamiento natural. *Calidad de revistas Científicas españolas*, 13-20.
- Romero, R. (2014). Análisis, Diseño e implementación de un Sistema de información aplicado a la gestión Educativa en centros de educación especial. *Ingeniero de Sistemas*. Universidad de Católica del Perú, Peru.
- Rosado, R. (2016). *Aplicación de una interfaz de lenguaje natural textual para la gestión administrativa del sistema de registro académico del instituto de educación superior tecnológico público Juli*. Univesidad UPAL, Peru.
- Saavedra, Y. (2015). Sistema web para la gestión documental en la empresa Development It E.I.R.L. *Ingeniero de Sistemas*. Universidad Cesar Vallejo, Peru.
- Sánchez, C. (s.f.). Reconocimiento Óptico de Caracteres. Obtenido de <http://www.it.uc3m.es/jvillena/irc/practicass/08-09/09.pdf>
- Sanchis, A. (2004). *Estimación y aplicación de medidas de confianza en reconocimiento automático del habla*. Universidad Politécnica de Valencia, Valencia.
- Sanhueza, R. (2013). *Formas de representación de textos cortos para su posterior Clasificación Automática*. Universidad Católica de Valparaíso, Chile.
- Scharager, J., & Reyes, P. (2001). *Muestreo no probabilístico*. Pontificia Universidad Católica de Chile, Chile.
- Siddiqui, T., & Tiwary, U. (2008). Natural language processing and information retrieval. Nueva Dehli: Oxford University Press.
- Sonnhammer, E., Von Heijne, G., & Krogh, A. (1998, Julio). A hidden Markov model for predicting transmembrane helices in protein sequences. *6*, 175-182.
- Tolosa, P. (2014). Construcción automática de grafos RDF a partir de frases en lenguaje natural para la integración de bases de datos heterogéneas. *Master en Madrid*. España.
- Toselli, H. (2014). *Reconocimiento de textos manuscrito Continuo*. Universidad Politécnica de Valencia, España.
- Valero, T. (2015). Extracción de Información con Algoritmos de Clasificación. *Maestro de ciencias*. Instituto Nacional de Astrofísica, Óptica y Electrónica, En Tonantzintla.
- Vásquez, A., Quispe, J. P., & Huayna, A. M. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*. 45-54. Obtenido de <http://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923/5121>
- Villar, G. (2014). *Implementación de un Sistema de gestión de Investigación para la especialidad de Ingeniería informática de la PUCP*. Universidad de Católica del Perú, Perú.
- Villaseca, P. (2017). Herramientas deep learning para predictive analytics. Una aplicación. Tesis Doctoral. Universidad de Concepción. Obtenido de <http://repositorio.udec.cl/handle/11594/2618>

Wikipedia. (2017, 06 20). *Perceptrón multicapa*. Obtenido de Perceptrón multicapa:

https://es.wikipedia.org/wiki/Perceptr%C3%B3n_multicapa#/media/File:RedNeuronalArtificial.png

Williams, E. (2013). *Propuesta para la digitalización del fondo documental del Dr. Belisario Porras*.

Universidad de Andaluía, España.

ANEXOS

ANEXO 1 : Matriz de Consistencia

Tabla 8

Matriz de Consistencia Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril

Problemas	Objetivos	Hipótesis	Variable	Dimensiones	Indicadores
¿En qué medida permite mejorar la gestión documental la implantación del modelo de clasificación automática de documentos en Grupo Coril?	Implantación de un modelo de Clasificación de textos que permitan analizar documentos en idioma español y determine su clasificación para mejorar la	HG: A través de la implantación del modelo de clasificación de documentos se mejorará la gestión documental en Grupo Coril.	VARIABLE	Procesamiento de documentos	Porcentaje de la precisión de documentos clasificados.
¿En qué medida permite la implantación del modelo mejorar el porcentaje de precisión de documentos clasificados?	Comparar los resultados obtenidos por clasificación automática para determinar el porcentaje de precisión documentos clasificados.	H1: A través de la comparación de los resultados obtenidos por la clasificación se mejorará el porcentaje de precisión documentos clasificados.	DEPENDIENTE: Clasificación automática de textos		
¿En qué medida la aplicación permite disminuir el tiempo para gestión de clasificación de documentos?	Reducir el tiempo de clasificación de documentos para gestión documentaria en Grupo Coril.	H2: El uso de la aplicación reducirá el tiempo empleado para gestión documentaria en Grupo Coril.		Registro	Tiempo promedio de registrar un documento

ANEXO 2: Desarrollo de Teorías

VII. DESARROLLO

7.1. Datos utilizados

Para el desarrollo de esta aplicación se tomara un conjunto de documentos que fueron brindados por la empresa Grupo Coril.

Actualmente Grupo Coril tiene una gran cantidad de documentos clasificados las cuales han sido recopiladas, sintetizadas y estructuradas en el siguiente documento de Excel donde la columna A es la clase, y la columna B es el contenido.

A	B
CLASE	DOCUMENTO
Expediente	MARTINEZ VILLOGAS ROBERT AV. EXCO rao. 478 COLAS Recbi de: DATA GLOBAL SERVES SIC. Identificado con RUC R-U.c. 10460381571 RECIBO POR HC
Expediente	ARTICULO 33 DE LA LEY DEL LA RECEPCION 15 JUNZO" 1200.00 (0.00) Fecha emisión del Total por honorarios: Retención (8 Total Neto Recibido:
Expediente	CORILKAPITAL CLIENTE DIRECCION Ruc FECHA RECIBI CONFORME R.U.C.: 20545003431 FACTURA 0002470 002- 13 ADQUIRENTE O USUARIO uuA. S
Expediente	CORILKAPITALMARKETSSA. R.U.C.: 20545003431 FACTURA 0002470 002- 13 ADQUIRENTE O USUARIO DEL 2C Surco. L ION CANCELADO Q6 CLIENTE DIRE
Expediente	Carvajal Tecnología y Servicios S.A.C. Carvajal Pisos -SAN' IAGO Ok R.U.C_ 20505779291 FACTURA ELECTRONICA NO F001-00024416 TECNOLOGIA Y SER
Expediente	R.u.c. N' Cliente; 20141718780 CODICO 76201104 DESCRIPCION CARCO MENSUAL CENT RETAIL CADENA CARGO MENSUAL CENT RETAIL t CADENA cont
Expediente	11200 Página d. I PRECIO u NITAR'O 37'60C 10.V ONEDA MES JULIO 2018 (TRANSACCIONES MAYO) Menta 'renta y Gobraoor ZOILA ORTIZ BCP '93.1469'
Expediente	MARTINEZ VILLOGAS ROBERT AV. EXCO rao. 478 COLAS Recbi de: DATA GLOBAL SERVES SIC. Identificado con RUC R-U.c. 10460381571 RECIBO POR HO
Expediente	Inciso A DEL ARTICULO 33 DE LA LEY DEL LA RECEPCION 15 JUNZO" 1200.00 (0.00) Fecha emisión del Total por honorarios: Retención (8 Total Neto Rec
Expediente	vida 14. 2-018 RECEPCION deu - S. a realizar d. de oba C" Cúb 10 — Ir. 316 — Lima
Expediente	QUISPE SIRBELION NOELIA MARIA E LOTE. 07 A. H. ACA PACHACIJTEC PROV. DEL CALIE PROV. CONST. CALLAO VENTANILLA Recibi do: CR HOLDING SA
Expediente	ENTA Ru.c. 10442779711 RECIBO POR HONORARIOS ELECTRONICO Nro.. 26 RECEPCION 500.00 (0.00) 15 Fuha de emisión 14 de del 2018 Total por honc
Expediente	ASESORA CORIL CONSULTING S.A.C. C. W3 uro. Chocar" d' Est.n.,v.. S.M. R.U.C: 20536879898 FACTURA 001- N? 002476 RECEPCION 1 5 JUN 2018 CLENTE
Expediente	R.U.C: 20536879898 FACTURA ASESORA CORIL CONSULTING SA.C. CSle Monterosa lm. W3 d. - Lima - L*na cent. 611 DIRECCION . FECHA 001- sus S/ REC
Expediente	ANAMARIA PARDO SILVESTRE sunz Recibi de: SAN PEDRO SA Identificado con Ruc DomiciWado en AV. La suma DOS Y Y por concepto de AUXLAR ADI
Expediente	RECEPCION *XES 2,13500 (0'0) 2.135W Fecha de ernisión del 201e Total por honorarios: Retención 8 R: Total Neto Recibido:
Expediente	CORILKAPITALMARKETSSA. CONFOHVE R.U.C:: 20545003431 FACTURA del a - Lime . Lima CANCELADO 002- DEU20 1 S 0002468 1 3 JUN 2018 ADQUIREN

Figura 14: Conjunto de Datos en Excel

7.2. Corpus NLTK

El corpus NLTK es un conjunto de datos más utilizado por la comunidad para la clasificación de textos. Adicional se implementa gramáticas de práctica, modelos entrenados.

Para instalar el corpus, primero se debe instalar el NLTK a continuación se realizaran los siguiente pasos.

Ejecutar en el intérprete Anaconda Python

```
>>> import NLTK
```

```
>>> nltk.download()
```

Se abrirá una nueva ventana, seleccione los paquetes que desea descargar.

La clasificación de documentos de NLTK se caracteriza por estar bien formateadas y escritas de manera concisa y eficiente para nuestro caso el muestreo de esta investigación será de tipo intencionado, ya que se escogerán los documentos que cumplan con las especificaciones para ser las pruebas clasificación de textos automáticas.

El objetivo de esta etapa es la obtención de un conjunto de datos aplicables a los modelos de clasificación disponibles, dado que dichos modelos son importantes para detallar la estimación de pesos por palabras, y otros atributos correspondientes, la cual determinara la frecuencia relativa de la palabra en la colección total de documentos.

Tabla 9

Clasificación de documentos

Cantidad	Tipo de documento
91	Carta
4	Carta De Renuncia
3	Carta Notarial
1	Certificado
1	Circular
743	Comprobante De Pago
2	Comprobante De Percepcion Vent
1	Esquela
15	Factura Electronica
1	Formulario
3	Oficio
1	Papeleta De Infraccion
10	Requerimiento
2	Resolucion
3	Resolucion Coactiva
8	Sesion De La Comision Administ
15	Transferencia

7.3. Etapa de Entrenamiento

Está definido como la categorización de textos lo que requiere un clasificador automático de un conjunto de datos clasificados manualmente al que se conocerá como conjunto de entrenamiento, se asume entonces que en esta etapa se incluirá las características que definan la clase de documento a la que pertenece de un conjunto de clases ya definidas.

7.3.1. Pre – procesamiento de textos

Está definida para eliminar las partes del documento que no aportan con nada para la clasificación automática, esto consta de tres fases:

- Tokenizacion: Consiste en dividir las palabras en tokens(unidad de seguridad) estas pueden ser palabras, símbolos, números, etcétera.

- Eliminación de palabras auxiliares: Consiste en eliminar las palabras que aparecen frecuentemente dentro del texto, pronombres, preposiciones, conjunciones, artículos, etcétera.
- Etiquetación morfosintáctica: Consiste en etiquetar cada palabra de un texto con su categoría (Sustantivo, adjetivo, verbo, etc.). Para esta tarea utilizaremos un etiquetador NLTK (Kit de herramientas de lenguaje natural).

7.3.2. Extracción de palabras claves

(Bermonti, 2016), menciona que las palabras claves representan el contenido esencial del documento. Se podría mencionar algunas palabras como por ejemplo Recibo por honorarios, factura, nota de crédito, certificado de retenciones, etc. Que aparecen frecuentemente en los documentos del módulo de documentación de grupo coril, por tal motivo la frecuencia relativa de la palabra del documento tendrá mayor valor para una clase, el cual será determinado con un peso por palabra.

Fórmula para calcular el peso por palabra.

$$w_{i \text{ clase}} = t f_i \cdot \log\left(\frac{N_{clases}}{n_{i \text{ clases}}}\right)$$

Donde $t f_i$ es el número de documentos de la clase en los que la palabra i -ésima aparece, este valor es normalizado entre el total de documentos en la clase; N_{clases} es el total de clases; y $n_{i \text{ clases}}$ es el número de clases que tienen los documentos con la i ésima palabra.

7.3.3. Ranking de palabras Claves

(Alberto Esteban, 2017), menciona que para calcular la relevancia de una frase y establecer un ranking de palabras se verifica la relevancia según el tipo de fuente de información a la que pertenece, en esta fase es donde radica el mayor grado de complejidad para nuestro conjunto de entrenamiento obtenida previamente. Si dos conceptos o términos tienden a estar frecuentemente de manera secuencial esto es un indicativo de que la relación semántica entre ellos es más fuerte. Por ejemplo Gerente General y Grupo Coril, tienen más relación que las palabras Gerente General y certificado.

7.4. Etapa de evaluación semántica

Para la evaluación de este método se ha propuesto tres algoritmos Naives Bayes, Cadenas de Markov y k-vecinos.

Se realizaron los experimentos utilizando 8 clases de tipos de comprobante que posee el Corpus NLTK. La evaluación del rendimiento del clasificador propuesto fue realizada en base a exactitud y precisión que son los estándares de medición de clasificación de textos.

En la figura 6 se muestra una comparación de la exactitud (tasa de aciertos) de los métodos propuestos en nuestros indicadores que ayudan a mejorar la clasificación automática de textos.

Como se puede observar en la figura 6 utilizando el método de clasificación de documentos para 8 clases de tipo de documento se obtiene con exactitud 82.07% usando el método de clasificación del algoritmo de naives bayes a comparación de los algoritmos k-vecinos y cadenas de markov.

En la tabla 6, los resultados obtenidos no son muy buenos esto se debe a que existe una gran cantidad clases de documentos. En esta evaluación solo se evaluó el clasificador Navies Bayes, como se explicó anteriormente en las bases teóricas el funcionamiento de este clasificador.

Tabla 10
Comparación de Precisión de Clases

Clase	Precisión
Acta	0.81
Aviso de cobranza	0.76
Boleta de Venta	0.75
Carta	0.87
Carta de Renuncia	0.75
Cedula de notificación	0.63
Certificado	0.66
Comprobante de Pago	0.68

Es importante mencionar que los tres métodos implementados están comparados gracias a la tokenización utilizadas en nuestro conjunto de entrenamiento para cada método. También se explicaron los algoritmos para la aplicación adicional se explicó cómo es que el método utiliza palabra claves para su clasificación y cuál de todos los métodos es el más representativo en la clasificación documentaria. En la tabla 7 se muestra las palabras claves con mayor peso seleccionadas por nuestro clasificador para cada una de las clasificaciones presentes en el corpus

NLTK. Como se pueden observar estas palabras son claves para la clasificación. El cual es reflejo de que la semántica ayuda a la clasificación de textos.

Tabla 11

Palabras Claves de Clasificación

Nombre	Palabras claves
Acta	acta
Aviso de cobranza	Aviso, Cobranza,
Boleta de Venta	Boleta, Venta,
Carta	Presente, estimados, atentamente, por medio, agradeciendo
Carta de Renuncia	Carta de Renuncia, Renuncia, oportunidad.
Cedula de notificación	Cédula, Expediente, Tribunal
Certificado	Instrumentos, Certificado, Rentas
Comprobante de Pago	Factura, Invoice, Electronica

VIII. PROCESOS DEL DESARROLLO DE LA APLICACIÓN

La programación del software se culminó. A continuación los pasos a realizar:

Habilitar un espacio de trabajo para realizar pruebas con las tecnologías mencionadas, desarrollar el test para verificar los resultados obtenidos, programar el servicio y procesar la data de la empresa.

8.1. Entorno de trabajo

Para el entorno de trabajo se utilizó un sistema operativo Ubuntu linux, con las siguientes características:

- Ubuntu 16 de 64-bit
- Kernel Linux 3.3.0-59
- AMD FX 4100, 8gb ram

8.2. Instalación de Programas necesarios.

A continuación los comandos a utilizar para el desarrollo de Inteligencia Artificial:

1.-Instalación de Anaconda Python

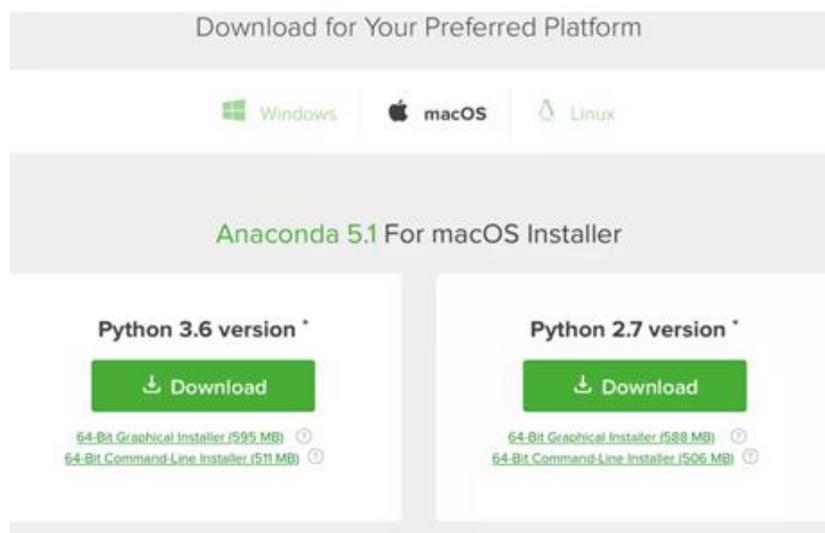


Figura 15 Anaconda Python 3.0

2.-Entorno de trabajo Jupyter Notebook



Figura 16: Entorno Jupyter

3.-Instalación de Python y escribir los comandos

- Instalación del CORPUS NLTK
- NLTK viene con gramáticas, un corpus de práctica, los métodos de entrenamiento, etc.
- Para instalar el corpus, se tiene que instalar iniciando el NLTK, a continuación, se deben ejecutar los siguientes comandos en el intérprete python.
- `import nltk`
- `nltk.download()`

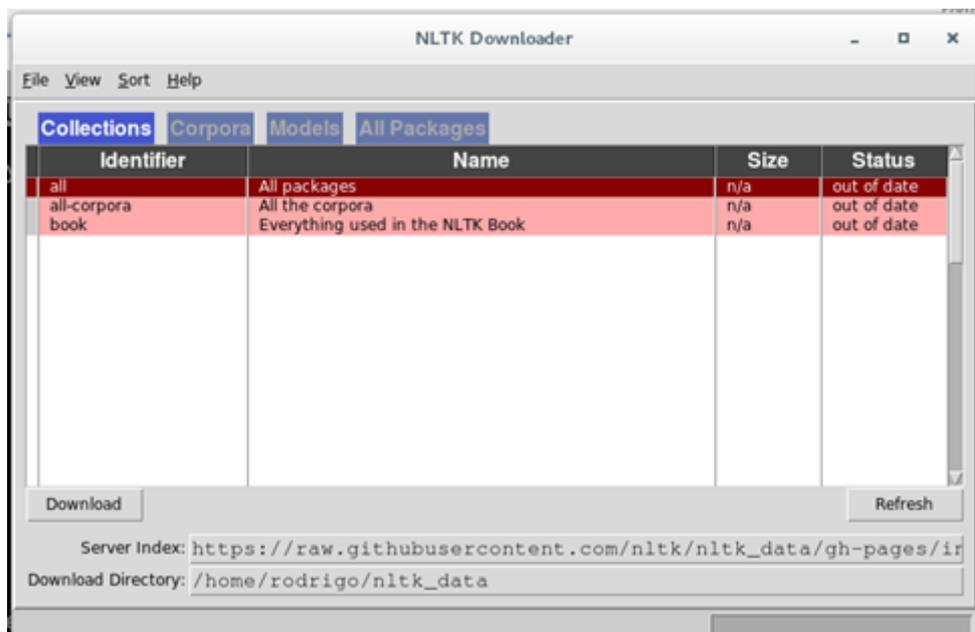


Figura 17 NLTK - CORPUS

4.- Instalar Pip:

```
jvargas@PCLIBRE:~$ sudo easy_install pip
```

5.-Instalar Numpy

```
jvargas@PCLIBRE:~$ sudo pip install -U numpy
```

6.-Creación del Servicio Web

El desarrollo del servicio REST está en el archivo “appClasificador.py”, es donde se configuran las rutas para que se puedan utilizar las librerías necesarias en este caso NLTK:

La servicio tiene la ruta (“/”): que es la url por defecto, para las pruebas se mostrará la interfaz del módulo de servicio. Esta interfaz está ubicada en la carpeta “api-erp”. Solo responderá a peticiones del tipo “POST, GET”.

Implementación con Python y NLTK+

7.-Uso de la Herramienta StringTagger

```
In [8]: 1 """
2 Uso sencillo de "StringTagger" para el etiquetado (clasificación) de texto.
3 """
4
5 from StringTagger.StringClf import Classifier
6 from StringTagger.getPage import getTextPage
```

8.-Importación del Corpus en español.

```
1 training_data = { # Datos para entrenar al clasificador
2 "Aviso de Cobranza":[
3 'http://www.wordreference.com/sinonimos/cobranza'
4 ],
5 "Boleta de Venta":[
6 'http://www.wordreference.com/sinonimos/boleta'
7 ],
8 "Carta":[
9 'http://www.wordreference.com/sinonimos/carta'
10 ],
11 "Carta de Renuncia":[
12 'http://www.wordreference.com/sinonimos/renuncia'
13 ],
14 "Cedula de notificacion":[
15 'http://www.wordreference.com/sinonimos/notificaci%C3%B3n'
16 ],
17 "Certificado":[
18 'http://www.wordreference.com/sinonimos/certificado'
19 ],
20 "Comprobante de Pago":[
21 'http://www.wordreference.com/sinonimos/factura'
22 ],
23 }
24
25 clf = Classifier() # Instancia del clasificador
26
27 for category,urls in training_data.items(): # Entrenamos al clasificador con el contenido de cada pagina
28     for url in urls:
29         clf.train(getTextPage(url),category) # El metodo "getTextPage", recibe como argumento una url para exi
30
31
```

9.-Enviamos la data que necesita para clasificarla.

```
In [10]: 1 # Iniciamos el proceso de clasificación con el metodo "String"
2 # Solo le pasamos como argumento el texto que deseamos etiquetar (clasificar)
3 string = "MARTINEZ VILLOGAS ROBERT AV. EXCO rao. 478 COLAS Recbi de: DATA GLOBAL SERVES SIC. Identificado con RUC R-
4 clas = clf.String(string)
5 print('\n')
6 print("Texto: %s " % string)
7 print("Etiqueta del Texto: %s " % clas)
```

Texto: MARTINEZ VILLOGAS ROBERT AV. EXCO rao. 478 COLAS Recbi de: DATA GLOBAL SERVES SIC. Identificado con RUC R- U.c. 10460381571 RECIBO POR HONORARIOS ELECTROMCO Nro: EOOI. 7 DornicWado en CAL ROSA OE SURCO La swna UN DOSCIENT OS y SOLES por DISEÑO GRAFCO Inciso A DEL ARTICULO 33 DE LA LEY DEL LA RECEPCION 15 JUNZO 1200.00 (0.00) Fecha emi sión del Total por honorarios: Retención (8 Total Neto Recibido:
Etiqueta del Texto: Certificado

10.-Algoritmo Bayesiano de Clasificación

```
import math, re
from collections import defaultdict
from fractions import Fraction
from .ExtractWords import getWords, wordFreq

class Classifier(object):
    def __init__(self):
        super(Classifier, self).__init__()
        self.classes = defaultdict(lambda : defaultdict(int))
        self.SamplesTotal = 0
        self.vocab = {}

    def train(self, obj, className): # Metodo para entrenar al clasificador
        frequents = None
        if isinstance(obj, str) or isinstance(obj, bytes):
            frequents = wordFreq(obj)
        elif isinstance(obj, dict):
            frequents = obj
        else:
            raise Exception("El objeto debe ser una cadena de texto o una palabra frecuente.")
        for word, freq in frequents.items():
            self.classes[className][word] += freq

    def allWords(self):
        vocab = set()
        for freqMaps in self.classes.values():
            vocab = vocab.union(freqMaps.keys())
        return sorted(vocab)

    def computeProb(self, obj, className): # Parte logica, la cual etiquetara al texto
        words = None
        if type(obj) in (str, bytes):
            words = getWords(obj)
        else:
            words = obj
        numWords = sum(self.classes[className].values())
        logSum = 0
        for word in words:
            freq = self.classes[className][word] + 1
            prob = Fraction(freq, numWords)
            logSum += math.log(prob)
        return logSum

    def String(self, obj, verbose=False): # El metodo "String" recibira al texto y lo etiquetara
        probs = {}
        for className in self.classes:
            probs[className] = self.computeProb(obj, className)
        highestClass = sorted(probs, key=probs.get)[-1]
        highest = probs[highestClass]
        if verbose:
            print(probs)
        return highestClass # "String" nos devolvera la etiqueta del texto (Clasificación)
```

ANEXO 4: Implantación del Modelo

IX. IMPLANTACION DEL MODELO DE CLASIFICACION

1.- Pasos para instalar OCR y Convertir de PDF a IMG

1.1- Instalación Tesseract en ubuntu 16.04

```
jvargas@PCLIBRE:/var/www/html/api-erp$ sudo apt-get install tesseract-ocr
```

1.2 Tesseract OCR for PHP

```
jvargas@PCLIBRE:/var/www/html/api-erp$ composer require thiagoalessio/tesseract_ocr
^[[^[[^[[Using version ^2.7 for thiagoalessio/tesseract_ocr
./composer.json has been updated
Loading composer repositories with package information
Updating dependencies (including require-dev)
```

1.3 Instalación del idioma en español de Tesseract

```
jvargas@PCLIBRE:/var/www/html/api-erp$ sudo apt-get install tesseract-ocr-spa
[sudo] password for jvargas:
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
Leyendo la información de estado... Hecho
Los paquetes indicados a continuación se instalaron de forma automática y ya no son necesarios.
linux-headers-4.15.0-33 linux-headers-4.15.0-33-generic linux-image-4.15.0-33-generic linux-modules-4.15.0-33-generic
linux-modules-extra-4.15.0-33-generic
Utilice «sudo apt autoremove» para eliminarlos.
```

1.4 Instanciamos el framework

```
8 use thiagoalessio\TesseractOCR\TesseractOCR;
9
```

2. Declaramos el controller para ser accedido

```
# Listar
documentos/documento/ocr:
  path:          /script
  defaults      : { _controller: ControllerBundle:Documentos\Documento:ocr }
  methods       : [GET,OPTIONS]
```

2.1 Instalación de ImageMagick

```
jvargas@PCLIBRE:/var/www/html/api-erp$ sudo apt-get install imagemagick
[sudo] password for jvargas:
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias
```

2.2 Instanciamos los paquetes

```
use Imagick;
use Spatie\PdfToImage\Pdf;
use PHPUnit\Framework\TestCase;
use Spatie\PdfToImage\Exceptions\InvalidFormat
```

3.0 Instalación de Naives Bayes

```
jvargas@PCLIBRE:/var/www/html/api-erp$ composer require nikknow/bayes
```

ANEXO 5: Código Fuente

ACCION REGISTRAR DOCUMENTO AUTOMATICAMENTE

En el sistema se tiene el botón guardar y clasificar, el evento mostrado en la imagen siguiente es el que utiliza el algoritmo bayesiano.

```
// ADD - POST
public function addAutoAction(Request $request){

    //try{
        // Recibo los datos
        $decoded = $this->get('Jwt')->decodeToken($request->headers->get('authToken'));
        if(!$decoded['status']) return new JsonResponse(['status'=>"logout"]);
        $user = $decoded['user'];

        $cadena = 'series,aFileName';
        $sentencia = $this->get('Read')->getPost($cadena);
        eval($sentencia);
        if(!$existent){
            return new JsonResponse(['error'=>'faltan variables.']);
        }

        $permiso=TRUE;
        if(!$permiso){
            return new JsonResponse(['error'=>'sin permiso']);
        }

        $aFileName = json_decode($aFileName,TRUE);
        $data=[
            'idE'=>NULL,
            'idP'=>NULL,
            'aFileName'=>$aFileName,
            'series'=>$series,
            'users'=>NULL
        ];

        $resp = $this->get('DocumentosServiciosDocumento')->validarAuto($data);
        if( !$resp['success'] ){
            return New JsonResponse(['error'=>$resp['msg']]);
        }

        $data['series']=$resp['data']['series'];
        $data['multiple']=$resp['data']['multiple'];
        $data['usuarios']=$resp['data']['usuarios'];
        $data['mDoc']= NULL;
        $data['d']= NULL;
        $data['u']=$user->getId();

        $begin;
        $this->getDoctrine()->getConnection()->beginTransaction();

        $resp = $this->get('DocumentosServiciosDocumento')->crearDocumentoAuto($data);
        if( !$resp['success'] ){
            $this->getDoctrine()->getConnection()->rollBack();
            return New JsonResponse(['error'=>$resp['msg']]);
        }

        $this->getDoctrine()->getConnection()->commit();
        return new JsonResponse(['status'=>'ok','msg'=>$resp['clasificacion']]);

    //}catch (\Exception $e){
        // if( isset($begin) ){
        //     $this->getDoctrine()->getConnection()->rollBack();
        // }
        // return New JsonResponse(['error'=>$e->getMessage()]);
    //}
}
```

ACEDIENDO AL SERVICIO DE REGISTRAR DOCUMENTO

En el proceso se registra el documento en la base de datos, luego sigue un proceso de digitalización de documento para el OCR en formato png.

```

$auxP = NULL;
if( $Sentidades[0]->getIdP() ){
    $auxP = '_' . $Sentidades[0]->getIdP()->getId();
}
$fecha = $Sentidades[0]->getF()->format('Y-m-d');
$hora = $Sentidades[0]->getF()->format('H:i:s');
foreach($Sentidades as $i => $Sentidad){
    $spath[] = [];
    if( count($aFileName)!=0 ){
        foreach($Sentidad->getPath() as $file){
            if ( file_exists( (Constante::$fileDocumentos) $oldidE.'/'.$file ) ){
                unlink( (Constante::$fileDocumentos) $oldidE.'/'.$file );
            }
        }
        // --- Guardamos el Scan de l(los) Documento(s)
        if( !$multiple ){
            foreach($aFileName as $j=>$file){
                $spath[] = 'Doc' . $Sentidad->getId() . '_' . $Sentidad->getIdE()->getId() . $auxP . '_' . ($j+1) . '_' . $fecha . '_' . $hora . '.pdf';

                $spath[$j] = str_replace(":",",",$spath[$j]);
                $resp = $this->copyDocumento( (Constante::$filetempo) $file['path'], $ruta $spath[$j],FALSE);
            }
        }else{
            $spath[] = 'Doc' . $Sentidad->getId() . '_' . $Sentidad->getIdE()->getId() . $auxP . '_1_' . $fecha . '_' . $hora . '.pdf';
            $spath[0] = str_replace(":",",",$spath[0]);
            $resp = $this->copyDocumento( (Constante::$filetempo) $aFileName[$i]['path'], $ruta $spath[0],FALSE);
        }
    }else{
        if( !$multiple ){
            if( !isset($oldidE) && ($idE!=$oldidE) && count($Sentidad->getPath())!=0 ){
                // --- Guardamos el Scan de l(los) Documento(s)
                foreach($Sentidad->getPath() as $j=>$file){
                    $spath[] = 'Doc' . $Sentidad->getId() . '_' . $Sentidad->getIdE()->getId() . $auxP . '_' . ($j+1) . '_' . $fecha . '_' . $hora . '.pdf';
                    $spath[$j] = str_replace(":",",",$spath[$j]);
                    $resp = $this->copyDocumento( (Constante::$fileDocumentos) $oldidE.'/'.$file, $ruta $spath[$j],FALSE);
                }
            }else{
                $spath = $Sentidad->getPath();
            }
        }else{
            $spath[] = 'Doc' . $Sentidad->getId() . '_' . $Sentidad->getIdE()->getId() . $auxP . '_1_' . $fecha . '_' . $hora . '.pdf';
            $spath[0] = str_replace(":",",",$spath[0]);
            $resp = $this->copyDocumento( (Constante::$fileDocumentos) $oldidE.'/'.$file, $ruta $spath[0],FALSE);
        }
    }
}

//////////////////////////////////////
//////////////////////////////////////DIGITALIZAMOS EL DOCUMENTO IMGAGIK//////////////////////////////////////
//////////////////////////////////////

$Sentidad->setPath($spath);
$Sentidad->setCount(count($spath));

$Sid = $Sentidad->getId();
// $spath = $Sentidad->getPath();
$Sinurl = "Sistema/Documentos/" . $Sentidad->getIdE()->getId() . "/" . $spath[0];

//if(!chmod($Sinurl, 0777)){
//    return ['success'=>false,'msg'=>'No se dio permiso al archivo '. $Sinurl];
//}

// $Sinurl = str_replace(":",",",$Sinurl);
$fichero = $Sinurl;
$newnamepng = str_replace(".pdf",".png",$spath[0]);
// $newnamepng = 'Doc18217_8450_1_2018-11-30_09:51:41.png';

```

ACEDIENDO AL SERVICIO DEL CLASIFICADOR BAYESIANO

En el proceso se clasifica el documento en la base de datos.

```

// $inurl = str_replace(":", "_", $inurl);
$fichero = $inurl;
$newnamepng = str_replace(".pdf", ".png", $path[0]);
// $newnamepng = 'Doc18217_8458_1_2018-11-30_09:51:41.png';

if (!file_exists($fichero)) {
    return ['success'=>false, 'msg'=>'No existe el archivo: ' . $fichero];
}

$imagic = new Imagick();
// Sets the image resolution
$imagic->setResolution(150, 150);
// Reads image from PDF
$imagic->readImage($fichero . '[0]');

// Writes an image
$imagenocr = "Sistema/Documentos/" . trim($entidad->getId()->getId()) . "/" . $newnamepng;
$imagic->writeImages($imagenocr, true);

$imagenocr = "Sistema/Documentos/" . trim($entidad->getId()->getId()) . "/" . $newnamepng;

////////////////////////////////////
////////////////////////////////////      OCR AL DOCUMENTO GRUPOCORIL      //////////////////////////////////////
////////////////////////////////////
// $palabra = (new TesseractOCR($imagenocr))->run();
$palabra = (new TesseractOCR($imagenocr))->lang('spa')->run();

// $ent[] = ['fichero'=>$fichero, 'imagenocr'=>$imagenocr];

// $palabra = '';
$entidad->setNamepathimg($imagenocr);
$entidad->setOcrcontenido($palabra);
$entidad->setEstado('5');

////////////////////////////////////
////////////////////////////////////      CLASIFICACION BAYESIANA      //////////////////////////////////////
////////////////////////////////////
$mTipoDoc = $this->finby_naibes_nayes($palabra);

if (!$mTipoDoc['success']){
    return ['success'=>false, 'msg'=>'Un error al clasificar: ' . $palabra];
}
$mDoc = $mTipoDoc['entidad'];
$entidad->setMDoc($mDoc);

$ent[] = ['texto'=>$palabra, 'clasificado'=>$mDoc->getD()];

$this->em->persist($entidad);
$this->em->flush();

// --- Envio de correos del documento
foreach($entidades as $i => $entidad){
    $resp = $this->generarCorreo($entidad, $usuarios);
    if( !$resp['success'] ){
        return $resp;
    }
}
}

```

ANEXO 6: Interfaz de Usuario

1.- Ingreso de usuario y contraseña.

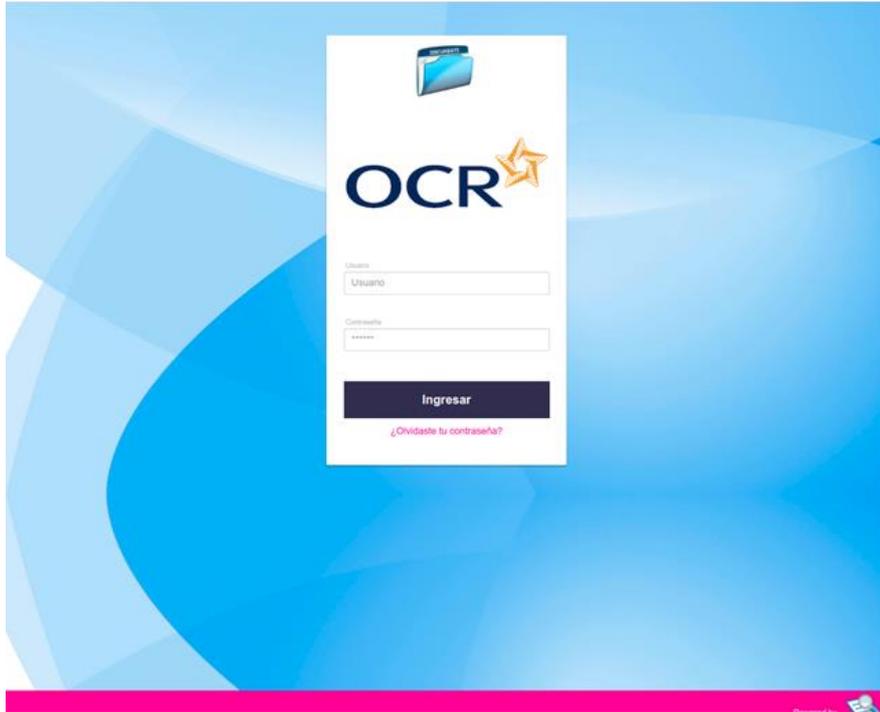


Figura 18: Inicio de Sesión login

2.- Ingreso de tipos de documento.

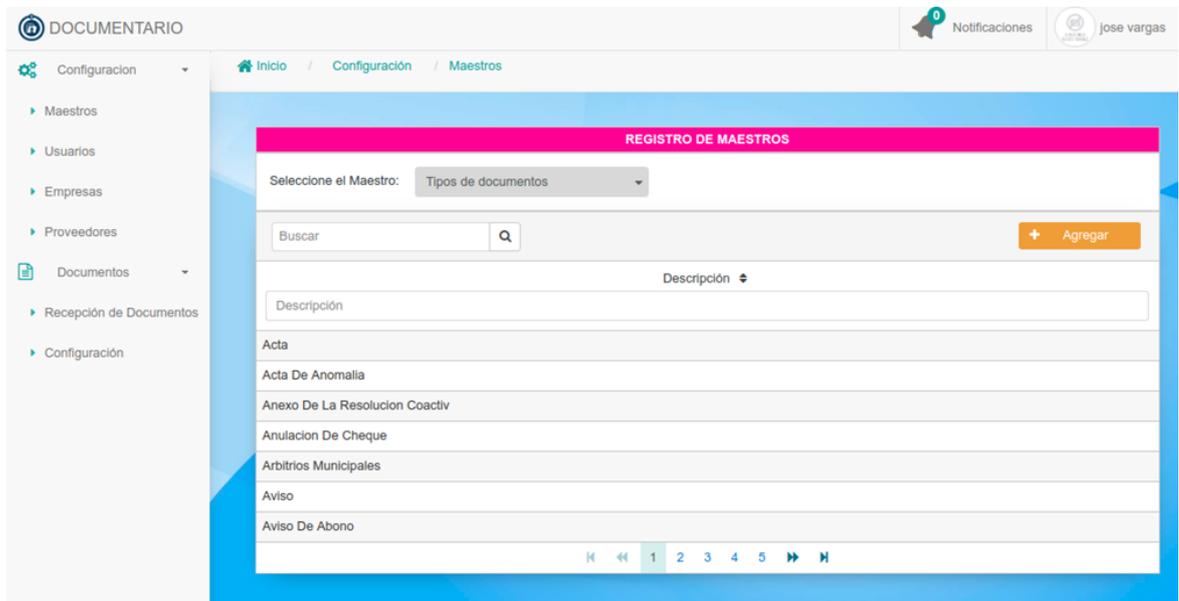


Figura 19: Registro de tipos de documento

3.- Registro de usuario.

REGISTRO DE USUARIOS						
Buscar						+ Agregar
Usuario	Nombre y Apellidos	Empresa	Movil	Anexo	Correo	
jvargas	Filtrar por Nombre	Filtrar por Empresa	Filtrar por Telefono	Filtrar por Anexo	Filtrar por Correo	
jvargas	jose vargas salas	Data System & Global Services S.a.c.			jvargasnet@hotmail.com	

Figura 20: Registro de Usuarios

4.- Registro de Proveedores

REGISTRO DE PROVEEDORES		
Buscar		+ Agregar
Razon	Documento	
tra	Filtrar por Documento	
Wme Contrata S.r.l.	RUC: 20517310183	
Transportes Luciano Flores Cargo Express	RUC: 20523678915	
Facundo Lastra Milder	RUC: 10427751738	
Gruas Y Transportes Beco Sac.	RUC: 20543170853	
Estrategia Marketing Digital Sac.	RUC: 20600552547	
Servic Nac De Adiestram En Trabaj Indust	RUC: 20131376503	
Mape Traducciones Sac.	RUC: 20506486484	

Figura 21: Registro de proveedores

4.- Gestión documental formulario principal

The screenshot shows a web interface titled "TRAMITE DOCUMENTARIO". At the top, there are date filters for "Fecha Inicio" (30/11/2018) and "Fecha Fin" (30/11/2018), and a "Columnas" dropdown set to "8 Items selected". Below this is a search bar and three buttons: "Descargar", "Agregar", and "Clasificar". The main area is a table with the following columns: "N° Documento", "De", "Para", "Tipo Documento", "Fecha de Registro", "Usuario de Registro", "Observaciones", and "OcrContenido". Each column has a corresponding filter button. The table contains 10 rows of data, all with the action "ACTUALIZAR" and the user "jose vargas". The footer of the table shows pagination controls for 10 items, with pages 1 and 2 visible.

N° Documento	De	Para	Tipo Documento	Fecha de Registro	Usuario de Registro	Observaciones	OcrContenido
18365		ACTUALIZAR	Informe	30/11/18	jose vargas		
18363		ACTUALIZAR	Comprobante De Pago	30/11/18	jose vargas		
18362		ACTUALIZAR	Certificado	30/11/18	jose vargas		
18360		ACTUALIZAR	Recibo	30/11/18	jose vargas		
18359		ACTUALIZAR	Carta De Renuncia	30/11/18	jose vargas		
18358		ACTUALIZAR	Carta	30/11/18	jose vargas		
18356		ACTUALIZAR	Comprobante De Pago	30/11/18	jose vargas		
18355		ACTUALIZAR	Recibo	30/11/18	jose vargas		
18354		ACTUALIZAR	Carta Notarial	30/11/18	jose vargas		
18352		ACTUALIZAR	Sesion De La Comision Administ	30/11/18	jose vargas		

Figura 22 Gestión Documental

5.- Clasificar un documento botón "Clasificar".

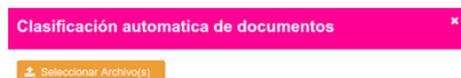


Figura 23 Agregar Documentos

5.- Seleccionamos los documento(s).

 Doc14408_22_7522_1_2018-08-02_16_42_35.pdf	339,7 kB	22 nov
 Doc14456_22_7409_1_2018-08-06_09_40_49.pdf	79,8 kB	22 nov
 Doc14521_22_6848_1_2018-08-08_12_35_54.pdf	1,1 MB	22 nov
 Doc14547_22_8015_1_2018-08-08_12_34_45.pdf	309,0 kB	22 nov

6.- Visualización de documentos



Seleccionar Archivo(s)

Doc14408_22_7522_1_2018-08-02_16_42_35.pdf 339,7 kB 22 nov

Doc14456_22_7409_1_2018-08-06_09_40_49.pdf 79,8 kB 22 nov

Doc14521_22_6848_1_2018-08-08_12_35_54.pdf 1,1 MB 22 nov

Doc14547_22_8015_1_2018-08-08_12_34_45.pdf 309,0 kB 22 nov

CRGas
Lima, 03 de Agosto del 2018
RECEPCION

CFP
CONSORCIO
RECEPCION
02 AGO 2018

SESION DE LA COMISION ADMINISTRADORA
RECEPCION
08 AGO 2018

Enviar al clasificador

Figura 24 Visualización de documentos

7.- Confirmación del envío para que sea clasificado por la IA.

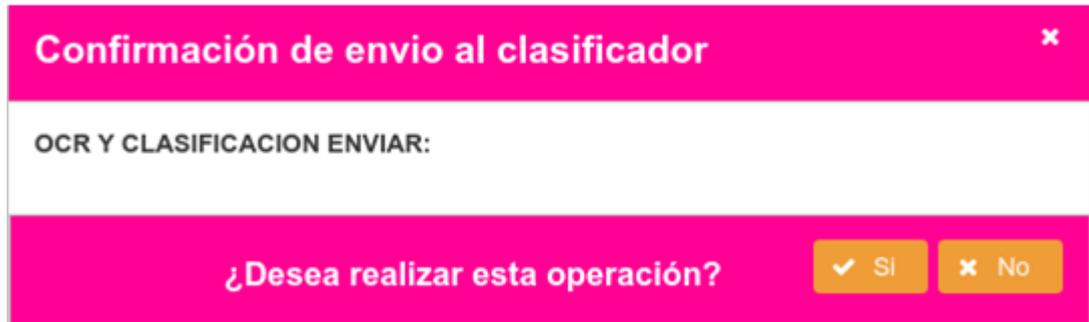


Figura 25 Confirmación del Sistema

8.- Resultado de la clasificación.

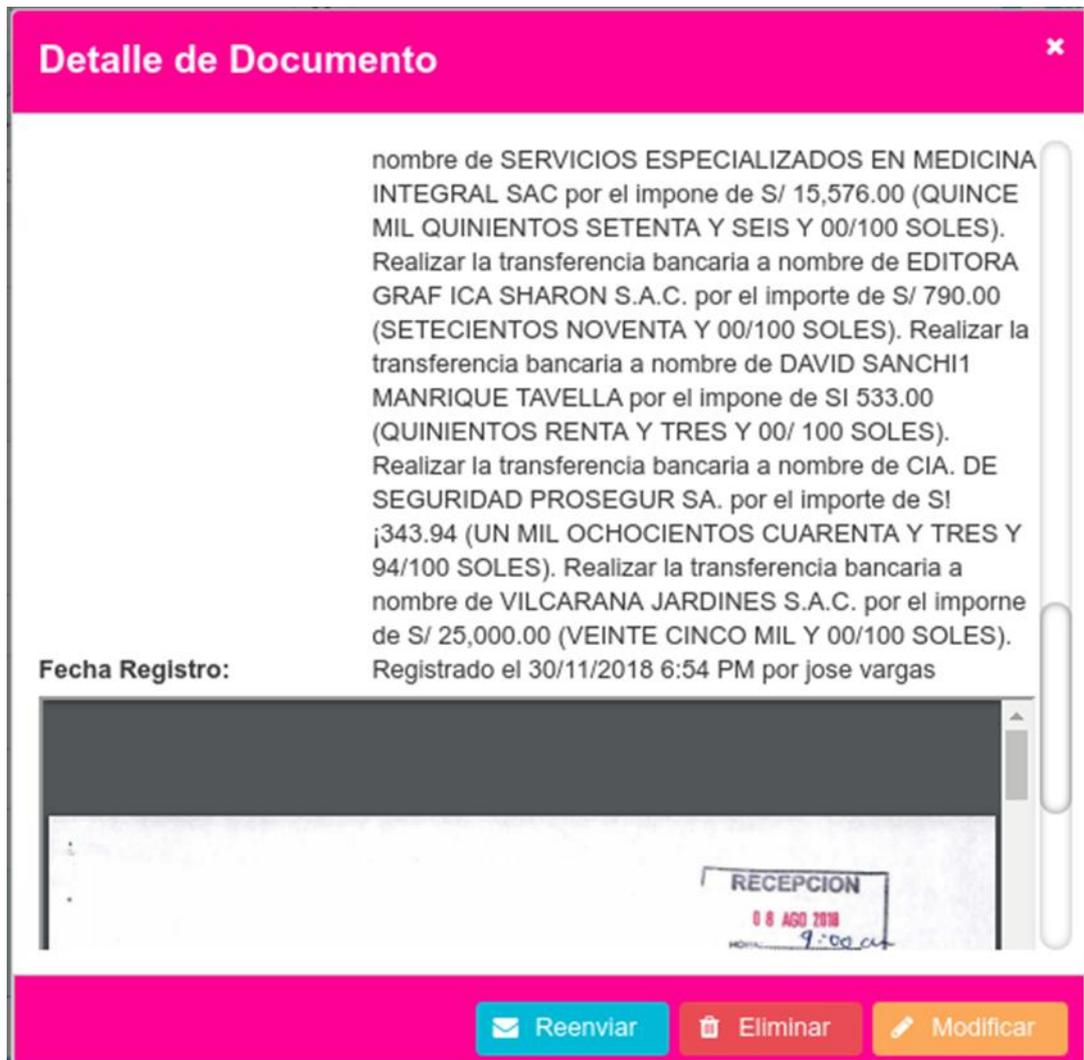


Figura 26 Resultado de la Clasificación

ANEXO 7: Ficha de registro indicadores tiempo pre - test

FICHA DE REGISTRO	
TIEMPO REGISTRO DE DOCUMENTOS	

Cliente : DATA SYSTEM GLOBAL SERVICES

Área/Dpto.: SISTEMAS

Sistema: DOCUMENTACION

Proceso: Tiempo promedio para registrar documentos

Fecha : 27/11/2018

Observador: ALEX CLETO DURAND CORDOVA

Periodo de Observación: 01-11-2018 al 15-11-2018

Ítem	fecha	hora inicio	hora fin	Tiempo de promedio en minutos <i>Segundos</i>
1	04/05/2018	17:21:34	17:34:14	660
2	04/05/2018	17:21:34	17:33:34	720
3	04/05/2018	17:21:34	17:34:33	715
4	04/05/2018	17:21:34	17:33:12	687
5	04/05/2018	17:21:34	17:32:52	669
6	04/05/2018	17:21:34	17:34:27	718
7	04/05/2018	17:24:14	17:36:23	714
8	04/05/2018	17:25:45	17:37:45	678
9	04/05/2018	17:30:17	17:41:48	760
10	05/09/2018	09:16:34	09:28:20	713
11	05/09/2018	09:19:55	09:31:01	769
12	05/09/2018	09:33:17	09:45:06	698
13	05/09/2018	09:35:11	09:46:52	750
14	05/09/2018	09:36:38	09:49:08	717
15	05/09/2018	09:38:20	09:49:38	697
16	05/09/2018	09:43:17	09:54:27	720
17	05/09/2018	09:47:14	09:58:44	749
18	05/09/2018	09:48:24	10:01:19	727
19	05/09/2018	09:54:03	10:06:54	773
20	05/09/2018	09:55:02	10:07:56	770
21	05/10/2018	10:21:47	10:34:15	770
22	05/10/2018	10:24:33	10:36:21	687
23	05/10/2018	10:24:33	10:37:24	773
24	05/10/2018	10:31:01	10:43:54	766
25	05/10/2018	10:31:01	10:43:10	689
26	05/10/2018	10:36:36	10:48:34	745
27	05/10/2018	10:41:34	10:52:42	686
28	05/10/2018	10:41:34	10:54:21	755
29	05/12/2018	12:01:54	12:13:27	662
30	05/12/2018	12:09:18	12:20:49	689
31	05/12/2018	12:09:18	12:21:08	703
32	05/12/2018	12:11:11	12:23:19	748
33	05/12/2018	12:53:38	13:06:34	738
34	05/03/2018	15:19:54	15:31:35	753

35	05/03/2018	15:22:39	15:35:19	763
36	05/03/2018	15:29:59	15:42:32	697
37	05/03/2018	15:48:15	15:59:39	669
38	05/04/2018	16:02:25	16:15:01	750
39	05/04/2018	16:03:48	16:15:30	721
40	05/06/2018	18:35:19	18:47:20	780
41	05/06/2018	18:36:41	18:48:36	698
42	09/09/2018	09:48:54	10:00:17	662
43	09/09/2018	09:54:04	10:06:04	713
44	09/09/2018	09:58:16	10:10:40	683
45	09/10/2018	10:07:02	10:19:43	737
46	09/10/2018	10:16:18	10:28:40	721
47	09/10/2018	10:19:24	10:31:22	768
48	09/10/2018	10:24:55	10:36:01	724
49	09/10/2018	10:34:38	10:46:43	670
50	09/10/2018	10:36:13	10:48:51	758
51	09/10/2018	10:41:15	10:53:31	748
52	09/10/2018	10:48:39	11:01:08	721
53	09/10/2018	10:49:38	11:01:18	678
54	09/10/2018	10:50:30	11:02:58	688
55	09/10/2018	10:51:37	11:04:30	662
56	09/10/2018	10:53:17	11:04:47	686
57	09/10/2018	10:54:25	11:06:05	663
58	09/10/2018	10:55:44	11:06:59	725
59	09/10/2018	10:56:57	11:08:37	663
60	09/10/2018	10:57:51	11:09:08	737
61	09/10/2018	10:59:10	11:11:42	770
62	09/11/2018	11:00:38	11:12:12	741
63	09/11/2018	11:01:37	11:13:39	728
64	09/11/2018	11:02:25	11:14:00	749
65	09/11/2018	11:02:57	11:15:37	772
66	09/11/2018	11:05:35	11:16:58	678
67	09/11/2018	11:08:42	11:20:01	702
68	09/11/2018	11:12:57	11:24:32	707
69	09/11/2018	11:14:28	11:25:40	698
70	09/11/2018	11:14:28	11:26:02	717
71	09/11/2018	11:14:28	11:26:31	683
72	09/11/2018	11:17:00	11:28:47	704
73	09/11/2018	11:20:24	11:32:45	690
74	09/11/2018	11:21:05	11:32:35	719
75	09/11/2018	11:22:00	11:34:26	671
76	09/11/2018	11:25:23	11:37:02	701
77	09/11/2018	11:25:28	11:37:58	745
78	09/11/2018	11:25:28	11:37:05	708
79	09/11/2018	11:25:28	11:36:37	680
80	09/11/2018	11:25:28	11:37:40	723
81	09/11/2018	11:26:50	11:39:26	712

Handwritten signature or mark

82	09/11/2018	11:26:50	11:39:21	718
83	09/11/2018	11:26:50	11:38:09	699
84	09/11/2018	11:28:15	11:39:32	749
85	09/11/2018	11:30:40	11:42:16	690
86	09/11/2018	11:37:33	11:49:17	670
87	09/11/2018	11:40:31	11:52:07	693
88	09/12/2018	12:02:22	12:14:17	717
89	09/12/2018	12:06:23	12:19:21	719
90	09/12/2018	12:22:26	12:34:29	675
91	09/12/2018	12:41:23	12:52:32	779
92	09/12/2018	12:51:54	13:04:23	747
93	09/12/2018	12:55:32	13:07:20	758
94	09/12/2018	12:56:53	13:09:40	661
95	09/02/2018	14:55:18	15:07:04	764
96	09/02/2018	14:56:12	15:07:35	704
97	09/02/2018	14:57:22	15:08:54	685
98	09/02/2018	14:57:22	15:08:52	734
99	09/02/2018	14:59:20	15:11:00	759
100	09/03/2018	15:51:13	16:02:29	682
101	09/03/2018	15:58:03	16:10:04	780
102	09/04/2018	16:02:40	16:14:35	697
103	09/04/2018	16:05:31	16:16:50	763
104	09/04/2018	16:51:21	17:03:59	723
105	09/04/2018	16:55:55	17:07:17	699
106	09/04/2018	16:59:49	17:10:58	707
107	09/05/2018	17:03:30	17:15:58	720
108	09/05/2018	17:07:38	17:19:13	776
109	09/05/2018	17:18:36	17:31:31	768
110	09/05/2018	17:27:12	17:39:26	707
111	10/09/2018	09:28:54	09:40:05	697
112	10/09/2018	09:30:22	09:41:51	697
113	10/09/2018	09:34:29	09:47:06	673
114	10/09/2018	09:40:02	09:51:16	692
115	10/09/2018	09:40:02	09:53:00	673
116	10/09/2018	09:40:02	09:52:14	742
117	10/09/2018	09:42:14	09:54:26	773
118	10/09/2018	09:43:20	09:56:10	751
119	10/10/2018	10:07:08	10:18:11	763
120	10/10/2018	10:07:08	10:18:36	731
121	10/10/2018	10:07:08	10:18:40	727
122	10/10/2018	10:07:08	10:20:06	688
123	10/10/2018	10:07:08	10:18:35	709
124	10/10/2018	10:07:08	10:18:55	748
125	10/10/2018	10:07:08	10:19:08	696
126	10/10/2018	10:07:08	10:20:07	666
127	10/10/2018	10:07:08	10:18:43	695
128	10/10/2018	10:07:08	10:19:21	677

Ab

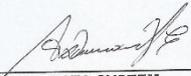
129	10/10/2018	10:07:08	10:19:55	666
130	10/10/2018	10:23:17	10:35:25	742
131	10/10/2018	10:23:17	10:35:41	715
132	10/10/2018	10:23:17	10:34:41	736
133	10/10/2018	10:23:17	10:35:26	776
134	10/10/2018	10:23:17	10:34:28	728
135	10/10/2018	10:23:17	10:35:25	674
136	10/10/2018	10:23:17	10:36:02	665
137	10/10/2018	10:23:17	10:35:27	754
138	10/10/2018	10:23:17	10:34:36	715
139	10/10/2018	10:23:17	10:35:54	741
140	10/10/2018	10:23:17	10:36:09	737
141	10/11/2018	11:26:46	11:38:38	686
142	10/11/2018	11:31:03	11:43:38	699
143	10/11/2018	11:32:13	11:43:42	687
144	10/11/2018	11:33:28	11:45:17	705
145	10/11/2018	11:34:22	11:46:37	663
146	10/12/2018	12:26:17	12:37:47	679
147	10/12/2018	12:29:09	12:40:15	753
148	10/12/2018	12:32:01	12:44:27	694
149	10/12/2018	12:39:43	12:51:12	702
150	10/12/2018	12:48:13	12:59:19	682
151	10/12/2018	12:49:20	13:01:54	705
152	10/12/2018	12:51:18	13:03:20	713
153	10/03/2018	15:20:56	15:33:17	666
154	10/03/2018	15:22:13	15:33:41	662
155	10/03/2018	15:26:55	15:38:39	757
156	10/03/2018	15:44:59	15:57:50	683
157	10/04/2018	16:09:46	16:21:07	697
158	10/04/2018	16:12:07	16:23:08	675
159	10/04/2018	16:38:14	16:50:27	738
160	10/05/2018	17:02:41	17:14:34	692
161	10/05/2018	17:04:56	17:15:56	686
162	10/05/2018	17:55:22	18:06:30	743
163	10/06/2018	18:01:34	18:13:03	678
164	11/09/2018	09:21:24	09:32:28	745
165	11/10/2018	10:54:05	11:05:57	664
166	11/10/2018	10:55:34	11:08:19	695
167	11/11/2018	11:55:51	12:08:30	689
168	11/02/2018	14:54:46	15:07:13	769
169	11/02/2018	14:55:42	15:07:24	666
170	11/02/2018	14:57:17	15:08:41	763
171	11/02/2018	14:57:17	15:09:38	757
172	11/02/2018	14:58:58	15:10:00	739
173	11/03/2018	15:02:31	15:13:58	678
174	11/03/2018	15:03:50	15:15:00	775
175	11/04/2018	16:14:01	16:26:08	770

176	11/04/2018	16:17:25	16:30:11	742
177	11/04/2018	16:23:09	16:35:43	762
178	11/05/2018	17:03:43	17:16:30	771
179	11/05/2018	17:05:44	17:18:39	779
180	11/05/2018	17:09:27	17:20:39	721
181	11/05/2018	17:11:43	17:23:14	753
182	11/05/2018	17:11:43	17:22:54	674
183	11/05/2018	17:11:43	17:23:21	686
184	11/05/2018	17:11:43	17:23:02	674
185	11/05/2018	17:17:10	17:28:25	693
186	11/05/2018	17:17:10	17:28:11	685
187	11/05/2018	17:18:51	17:29:54	719
188	11/05/2018	17:18:51	17:30:36	710
189	11/05/2018	17:21:35	17:34:30	723
190	11/05/2018	17:23:47	17:36:19	688
191	11/05/2018	17:25:20	17:38:07	748
192	11/05/2018	17:26:35	17:37:36	759
193	11/05/2018	17:27:17	17:38:32	678
194	11/05/2018	17:27:59	17:39:41	698
195	11/05/2018	17:28:49	17:40:54	751
196	11/05/2018	17:29:52	17:41:13	729
197	11/05/2018	17:30:37	17:42:19	664
198	11/05/2018	17:31:42	17:42:55	715
199	11/05/2018	17:32:59	17:45:53	705
200	11/05/2018	17:32:59	17:44:04	669
201	11/05/2018	17:36:31	17:47:59	776
202	11/05/2018	17:36:31	17:47:46	747
203	11/05/2018	17:40:47	17:52:15	661
204	11/05/2018	17:40:47	17:52:30	749
205	11/05/2018	17:42:32	17:54:47	771
206	11/05/2018	17:45:39	17:58:08	772
207	11/05/2018	17:46:23	17:58:14	701
208	11/05/2018	17:48:28	18:00:20	676
209	11/05/2018	17:48:28	18:00:13	714
210	11/05/2018	17:50:02	18:01:21	711
211	11/05/2018	17:51:54	18:04:13	661
212	11/05/2018	17:52:56	18:04:04	697
213	11/05/2018	17:53:46	18:05:26	748
214	11/05/2018	17:54:55	18:07:18	690
215	11/05/2018	17:55:48	18:07:10	677
216	11/05/2018	17:57:33	18:08:55	720
217	11/05/2018	17:58:38	18:11:30	679
218	11/05/2018	17:59:26	18:12:24	714
219	11/06/2018	18:01:41	18:13:19	687
220	11/06/2018	18:08:34	18:19:57	692
221	12/09/2018	09:22:22	09:34:53	661
222	12/09/2018	09:23:36	09:36:05	743

223	12/09/2018	09:31:40	09:43:08	669
224	12/09/2018	09:55:24	10:07:48	692
225	12/10/2018	10:00:09	10:11:35	696
226	12/10/2018	10:02:09	10:14:49	692
227	12/10/2018	10:08:02	10:20:43	709
228	12/10/2018	10:10:11	10:22:14	708
229	12/10/2018	10:20:04	10:31:56	774
230	12/10/2018	10:21:14	10:33:09	711
231	12/10/2018	10:22:15	10:34:48	731
232	12/10/2018	10:29:32	10:41:49	712
233	12/10/2018	10:46:50	10:58:17	762
234	12/10/2018	10:47:52	10:59:58	686
235	12/11/2018	11:55:43	12:07:33	715
236	12/12/2018	12:02:52	12:13:54	745
237	12/12/2018	12:05:54	12:17:54	706
238	12/12/2018	12:05:54	12:17:43	768
239	12/12/2018	12:08:03	12:19:34	731
240	12/12/2018	12:19:05	12:30:26	673
241	12/12/2018	12:20:23	12:31:27	758
242	12/12/2018	12:52:03	13:05:00	715
243	12/01/2018	13:00:49	13:12:30	700
244	12/03/2018	15:40:40	15:52:58	689
245	12/03/2018	15:40:40	15:52:11	727
246	12/03/2018	15:43:10	15:54:14	718
247	12/03/2018	15:43:10	15:54:48	674
248	12/03/2018	15:47:37	15:59:55	767
249	12/03/2018	15:48:30	16:00:32	769
250	12/03/2018	15:51:43	16:04:42	689
251	12/04/2018	16:01:45	16:13:11	704
252	12/04/2018	16:03:18	16:14:42	766
253	12/04/2018	16:09:55	16:22:35	719
254	12/04/2018	16:23:45	16:36:43	710
255	12/04/2018	16:45:21	16:56:39	721
256	12/04/2018	16:45:21	16:56:33	776
257	12/04/2018	16:52:44	17:04:47	747
258	12/04/2018	16:56:41	17:07:46	664
259	12/04/2018	16:58:04	17:09:12	684
260	12/04/2018	16:59:10	17:11:49	724
261	12/05/2018	17:00:16	17:11:38	697
262	12/05/2018	17:01:20	17:14:19	664
263	12/05/2018	17:04:37	17:16:02	770
264	12/05/2018	17:33:38	17:46:33	668
265	15/10/2018	10:59:25	11:11:19	664
266	15/12/2018	12:03:44	12:16:23	662
267	15/12/2018	12:05:31	12:17:43	777
268	15/12/2018	12:12:54	12:24:03	713
269	15/12/2018	12:58:20	13:11:16	725

Observaciones:

Este cuadro permitirá observar el tiempo promedio que le toma registrar en el periodo de noviembre del 2018 con el antiguo sistema.


DATA SYSTEM
Alex Durand Córdoba
Gerente General

ANEXO 8: Ficha de registro indicadores tiempo post-test

/	FICHA DE REGISTRO
	TIEMPO REGISTRO DE DOCUMENTOS

Cliente : DATA SYSTEM GLOBAL SERVICES

Área/Dpto.: SISTEMAS

Sistema: DOCUMENTACION

Proceso: Tiempo promedio para registrar documentos

Fecha : 27/11/2018

Observador: ALEX CLETO DURAND CORDOVA

Periodo de Observación: 16-11-2018 al 31-11-2018

Ítem	fecha	hora inicio	hora fin	Tiempo de promedio en minutos <i>Segundos</i>
1	19/05/2018	17:30:36	17:37:42	333
2	19/05/2018	17:30:36	17:38:03	336
3	19/05/2018	17:30:36	17:39:14	323
4	19/05/2018	17:41:57	17:49:56	406
5	19/05/2018	17:41:57	17:50:51	314
6	19/05/2018	17:41:57	17:50:23	328
7	19/05/2018	17:41:57	17:48:59	346
8	19/05/2018	17:41:57	17:50:43	332
9	19/05/2018	17:41:57	17:50:16	362
10	19/05/2018	17:41:57	17:50:08	350
11	19/06/2018	18:14:26	18:22:03	336
12	19/06/2018	18:17:02	18:25:12	420
13	22/10/2018	10:41:18	10:48:48	330
14	22/10/2018	10:43:51	10:51:53	400
15	22/10/2018	10:45:34	10:53:49	374
16	22/10/2018	10:45:34	10:52:59	322
17	22/10/2018	10:49:52	10:57:29	417
18	22/11/2018	11:03:03	11:11:57	401
19	22/11/2018	11:04:30	11:12:11	323
20	22/11/2018	11:11:10	11:20:02	311
21	22/11/2018	11:26:11	11:34:30	420
22	22/11/2018	11:36:17	11:43:21	318
23	22/11/2018	11:38:47	11:47:08	410
24	22/11/2018	11:38:47	11:46:54	303
25	22/11/2018	11:39:56	11:47:51	325
26	22/11/2018	11:41:06	11:49:28	395
27	22/11/2018	11:47:40	11:56:28	317
28	22/11/2018	11:47:40	11:56:40	367
29	22/11/2018	11:47:40	11:56:18	349
30	22/11/2018	11:47:40	11:55:50	384
31	22/11/2018	11:47:40	11:56:11	383
32	22/12/2018	12:09:02	12:16:24	401
33	22/12/2018	12:11:49	12:20:08	393
34	22/12/2018	12:15:11	12:24:00	324

35	22/12/2018	12:19:03	12:26:37	397
36	22/12/2018	12:26:34	12:33:56	360
37	22/12/2018	12:32:12	12:41:04	321
38	22/12/2018	12:32:12	12:39:21	402
39	22/12/2018	12:32:12	12:39:16	371
40	22/12/2018	12:35:35	12:44:22	379
41	22/12/2018	12:36:45	12:45:02	326
42	22/12/2018	12:38:08	12:45:29	325
43	22/12/2018	12:40:12	12:48:15	302
44	22/12/2018	12:42:17	12:50:15	344
45	22/12/2018	12:42:17	12:50:04	403
46	22/12/2018	12:46:11	12:53:26	303
47	22/12/2018	12:47:18	12:55:51	395
48	22/12/2018	12:48:11	12:56:29	404
49	22/12/2018	12:49:07	12:56:52	335
50	22/12/2018	12:51:12	12:58:51	391
51	22/12/2018	12:52:34	13:01:11	392
52	22/12/2018	12:55:00	13:02:49	388
53	22/12/2018	12:58:01	13:05:55	306
54	22/01/2018	13:00:07	13:08:56	347
55	22/01/2018	13:02:18	13:09:45	416
56	22/01/2018	13:04:05	13:11:22	397
57	22/01/2018	13:07:36	13:15:50	381
58	22/03/2018	15:36:41	15:44:43	368
59	22/03/2018	15:36:41	15:44:13	375
60	22/03/2018	15:36:41	15:44:22	401
61	22/03/2018	15:36:41	15:44:03	346
62	22/03/2018	15:36:41	15:44:23	376
63	22/03/2018	15:36:41	15:43:52	366
64	22/03/2018	15:36:41	15:44:40	395
65	22/03/2018	15:36:41	15:44:41	315
66	22/03/2018	15:36:41	15:43:55	323
67	22/03/2018	15:36:41	15:44:57	371
68	22/03/2018	15:47:28	15:54:36	365
69	22/04/2018	16:00:02	16:08:02	408
70	22/04/2018	16:01:24	16:10:23	330
71	22/04/2018	16:02:45	16:10:20	385
72	22/04/2018	16:03:48	16:12:08	326
73	22/04/2018	16:14:39	16:21:50	395
74	22/04/2018	16:28:47	16:37:09	307
75	22/04/2018	16:35:03	16:42:32	305
76	22/04/2018	16:44:07	16:52:03	324
77	22/05/2018	17:04:31	17:12:42	344
78	22/05/2018	17:16:07	17:23:15	326
79	22/05/2018	17:19:58	17:28:46	402
80	22/05/2018	17:24:23	17:32:28	318
81	22/05/2018	17:28:50	17:36:05	322

82	22/06/2018	18:09:13	18:17:16	311
83	22/06/2018	18:09:13	18:18:00	320
84	22/06/2018	18:09:13	18:16:31	332
85	22/06/2018	18:09:13	18:18:00	378
86	22/06/2018	18:09:13	18:17:22	409
87	23/09/2018	09:40:29	09:49:08	348
88	23/09/2018	09:40:29	09:48:32	352
89	23/09/2018	09:40:29	09:48:38	370
90	23/09/2018	09:40:29	09:47:50	317
91	23/09/2018	09:40:29	09:47:51	357
92	23/09/2018	09:40:29	09:49:09	390
93	23/09/2018	09:40:29	09:47:58	414
94	23/09/2018	09:40:29	09:47:33	338
95	23/09/2018	09:40:29	09:48:29	364
96	23/09/2018	09:40:29	09:47:47	322
97	23/09/2018	09:40:29	09:48:22	379
98	23/09/2018	09:40:29	09:49:25	402
99	23/09/2018	09:40:29	09:48:12	329
100	23/09/2018	09:40:29	09:47:47	300
101	23/09/2018	09:40:29	09:48:36	396
102	23/10/2018	10:29:48	10:37:27	329
103	23/10/2018	10:31:25	10:38:49	335
104	23/10/2018	10:35:19	10:44:03	353
105	23/10/2018	10:46:35	10:54:49	389
106	23/10/2018	10:46:35	10:55:21	321
107	23/10/2018	10:56:43	11:04:09	367
108	23/11/2018	11:06:27	11:13:44	304
109	23/11/2018	11:06:27	11:14:55	369
110	23/11/2018	11:11:11	11:19:30	371
111	23/11/2018	11:13:27	11:22:24	314
112	23/11/2018	11:13:27	11:21:45	406
113	23/11/2018	11:26:54	11:34:52	393
114	23/11/2018	11:28:18	11:36:09	393
115	23/11/2018	11:34:16	11:42:30	391
116	23/11/2018	11:37:04	11:45:58	355
117	23/11/2018	11:37:04	11:44:57	398
118	23/11/2018	11:47:17	11:55:52	358
119	23/11/2018	11:49:11	11:56:19	404
120	23/11/2018	11:50:22	11:57:38	308
121	23/11/2018	11:51:29	12:00:22	361
122	23/12/2018	12:05:39	12:14:04	300
123	23/12/2018	12:14:58	12:23:46	360
124	23/12/2018	12:15:48	12:24:24	360
125	23/12/2018	12:29:28	12:36:38	410
126	23/12/2018	12:33:56	12:41:38	417
127	23/12/2018	12:35:18	12:43:59	335
128	23/03/2018	15:31:17	15:40:08	390

AS

129	23/03/2018	15:44:00	15:52:56	370
130	23/03/2018	15:52:14	15:59:17	343
131	23/04/2018	16:00:14	16:08:41	366
132	23/04/2018	16:06:46	16:14:56	333
133	23/04/2018	16:25:45	16:34:00	337
134	23/04/2018	16:29:21	16:37:38	334
135	23/04/2018	16:33:11	16:41:12	382
136	23/04/2018	16:37:09	16:45:55	344
137	23/04/2018	16:39:21	16:46:42	395
138	23/04/2018	16:45:22	16:53:14	393
139	23/04/2018	16:51:34	16:59:43	369
140	24/09/2018	09:17:11	09:24:27	413
141	24/10/2018	10:01:19	10:08:55	380
142	24/10/2018	10:09:04	10:16:58	329
143	24/10/2018	10:13:01	10:21:43	362
144	24/10/2018	10:33:45	10:40:51	382
145	24/10/2018	10:40:00	10:47:33	341
146	24/10/2018	10:45:17	10:54:01	340
147	24/10/2018	10:55:31	11:02:35	323
148	24/11/2018	11:08:52	11:17:34	381
149	24/12/2018	12:44:55	12:53:35	318
150	24/12/2018	12:46:44	12:54:14	395
151	24/12/2018	12:48:20	12:55:44	376
152	24/03/2018	15:37:12	15:45:18	403
153	24/03/2018	15:39:42	15:48:00	382
154	24/04/2018	16:21:33	16:29:30	338
155	24/04/2018	16:24:15	16:31:32	390
156	24/04/2018	16:25:12	16:32:54	357
157	24/04/2018	16:28:47	16:36:29	338
158	24/05/2018	17:22:53	17:30:58	389
159	24/05/2018	17:30:26	17:37:36	324
160	25/09/2018	09:41:51	09:50:20	314
161	25/10/2018	10:24:39	10:32:20	346
162	25/10/2018	10:34:53	10:43:39	333
163	25/10/2018	10:40:17	10:48:41	384
164	25/10/2018	10:41:22	10:49:10	409
165	25/10/2018	10:57:04	11:04:42	406
166	25/11/2018	11:07:28	11:15:22	373
167	25/11/2018	11:10:13	11:18:36	376
168	25/12/2018	12:49:41	12:56:52	369
169	25/12/2018	12:56:05	13:04:15	322
170	25/12/2018	12:59:36	13:06:58	344
171	25/01/2018	13:05:37	13:13:10	335
172	25/04/2018	16:06:45	16:15:00	329
173	25/04/2018	16:08:40	16:16:20	412
174	25/04/2018	16:19:01	16:27:20	361
175	25/04/2018	16:21:06	16:29:14	338

176	25/04/2018	16:29:55	16:38:40	350
177	25/04/2018	16:35:19	16:43:14	304
178	25/05/2018	17:13:19	17:21:54	403
179	25/05/2018	17:18:39	17:27:29	302
180	25/05/2018	17:33:27	17:41:05	366
181	25/06/2018	18:00:28	18:09:04	340
182	26/09/2018	09:51:25	09:58:59	348
183	26/09/2018	09:51:25	09:58:44	369
184	26/09/2018	09:51:25	09:59:14	337
185	26/09/2018	09:51:25	09:59:03	378
186	26/09/2018	09:51:25	09:59:01	367
187	26/09/2018	09:51:25	10:00:13	397
188	26/09/2018	09:51:25	09:59:07	339
189	26/09/2018	09:51:25	09:59:35	412
190	26/09/2018	09:51:25	10:00:16	399
191	26/09/2018	09:51:25	09:59:07	331
192	26/09/2018	09:51:25	09:58:54	355
193	26/09/2018	09:52:59	10:01:05	346
194	26/09/2018	09:52:59	10:00:32	325
195	26/09/2018	09:54:31	10:01:59	366
196	26/09/2018	09:59:13	10:06:18	365
197	26/09/2018	09:59:13	10:07:26	348
198	26/09/2018	09:59:13	10:06:36	391
199	26/09/2018	09:59:13	10:06:37	386
200	26/09/2018	09:59:13	10:08:12	399
201	26/09/2018	09:59:13	10:06:31	333
202	26/09/2018	09:59:13	10:08:03	391
203	26/09/2018	09:59:13	10:06:19	415
204	26/09/2018	09:59:13	10:07:33	354
205	26/09/2018	09:59:13	10:06:44	411
206	26/09/2018	09:59:13	10:07:55	357
207	26/09/2018	09:59:13	10:07:53	390
208	26/09/2018	09:59:13	10:06:40	407
209	26/09/2018	09:59:13	10:07:47	328
210	26/10/2018	10:03:59	10:12:39	354
211	26/10/2018	10:04:46	10:13:16	347
212	26/10/2018	10:05:34	10:14:00	349
213	26/10/2018	10:09:16	10:18:03	330
214	26/10/2018	10:11:30	10:19:36	303
215	26/10/2018	10:12:10	10:20:07	334
216	26/10/2018	10:13:04	10:20:04	319
217	26/10/2018	10:14:06	10:22:41	359
218	26/10/2018	10:14:49	10:21:57	403
219	26/10/2018	10:18:45	10:25:58	415
220	26/10/2018	10:19:24	10:27:18	348
221	26/10/2018	10:19:45	10:28:02	300
222	26/10/2018	10:20:43	10:27:53	351

AB

223	26/10/2018	10:20:49	10:29:34	310
224	26/10/2018	10:22:03	10:30:40	393
225	26/10/2018	10:22:05	10:29:59	413
226	26/10/2018	10:22:54	10:30:37	413
227	26/10/2018	10:23:34	10:31:53	355
228	26/10/2018	10:25:16	10:32:54	327
229	26/10/2018	10:26:21	10:33:41	317
230	26/10/2018	10:26:38	10:34:05	385
231	26/10/2018	10:27:39	10:36:22	320
232	26/10/2018	10:28:29	10:35:59	390
233	26/10/2018	10:28:29	10:35:29	392
234	26/10/2018	10:28:29	10:37:08	398
235	26/10/2018	10:39:12	10:47:27	381
236	26/10/2018	10:44:24	10:52:24	356
237	26/10/2018	10:44:24	10:53:05	397
238	26/10/2018	10:44:24	10:52:25	315
239	26/10/2018	10:45:23	10:52:35	314
240	26/10/2018	10:47:23	10:54:58	313
241	26/10/2018	10:51:24	10:59:45	305
242	26/11/2018	11:04:00	11:11:20	387
243	26/11/2018	11:05:13	11:12:30	361
244	26/11/2018	11:10:40	11:17:55	311
245	26/11/2018	11:12:08	11:19:17	314
246	26/11/2018	11:16:06	11:23:49	354
247	26/11/2018	11:17:34	11:24:52	351
248	26/11/2018	11:19:11	11:27:30	301
249	26/11/2018	11:21:39	11:28:47	338
250	26/11/2018	11:21:59	11:29:46	302
251	26/11/2018	11:32:50	11:41:37	351
252	26/11/2018	11:33:28	11:41:20	408
253	26/11/2018	11:35:57	11:43:21	335
254	26/11/2018	11:36:23	11:45:05	345
255	26/11/2018	11:36:50	11:44:30	367
256	26/11/2018	11:40:40	11:49:12	320
257	26/11/2018	11:43:11	11:51:17	329
258	26/12/2018	12:53:45	13:01:53	310
259	26/03/2018	15:48:17	15:56:46	356
260	26/03/2018	15:50:12	15:57:22	303
261	26/03/2018	15:53:02	16:01:48	341
262	26/03/2018	15:53:02	16:00:11	338
263	26/03/2018	15:55:44	16:03:30	416
264	26/03/2018	15:58:28	16:06:48	351
265	26/04/2018	16:03:42	16:10:58	345
266	26/04/2018	16:05:23	16:13:23	343
267	26/04/2018	16:07:00	16:14:37	355
268	26/04/2018	16:08:42	16:16:26	355
269	26/04/2018	16:10:41	16:18:05	375

A
70

Observaciones:

Este cuadro permitirá observar el tiempo promedio que le toma registrar en el periodo de noviembre del 2018 con el NUEVO sistema.



DATA SYSTEM

Alex Durand Córdova
Gerente General

ANEXO 9: Ficha de porcentaje de precisión

FICHA DE REGISTRO
Porcentaje de precisión del resultado obtenido por la clasificación

Cliente : DATA SYSTEM GLOBAL SERVICES

Área/Dpto.: SISTEMAS

Sistema: DOCUMENTACION

Proceso: Tiempo promedio para registrar documentos

Fecha : 27/11/2018

Observador: ALEX CLETO DURAND CORDOVA

Periodo de Observación: 01-11-2018 al 31-11-018

archivo	contenido	usuario	bayesiano	Pre-test Calsifico Ma(1)	Post-test Clasifico Ma(1)
Doc14308_23_3766_1_2018-08-01_09.4		Comprabante De Pago	Comprabante De Pago	0	0
Doc14309_66_3766_1_2018-08-01_09.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14310_7616_8234_1_2018-06-01_09.5	0001- N7 002120	Comprabante De Pago	Comprabante De Pago	0	0
Doc14311_36_8234_1_2018-06-01_09.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14312_8121_8234_1_2018-06-01_09.5	MOCAS	Comprabante De Pago	Comprabante De Pago	0	0
Doc14313_36_7388_1_2018-06-01_10.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14314_64_3766_1_2018-06-01_10.4	GRUPOCORIL	Comprabante De Pago	Comprabante De Pago	0	0
Doc14315_54_3766_1_2018-06-01_10.4	GRUPO CORIL	Comprabante De Pago	Comprabante De Pago	0	0
Doc14316_83_3766_1_2018-06-01_10.4	GRUPO CORIL	Comprabante De Pago	Comprabante De Pago	0	0
Doc14317_66_8228_1_2018-06-01_10.5	Carta	Comprabante De Pago	Comprabante De Pago	1	0
Doc14318_63_3766_1_2018-06-01_11.4	ELECTRONICA	Comprabante De Pago	Comprabante De Pago	0	0
Doc14319_45_3766_1_2018-06-01_11.5	ELECCIÓN'S c1'cA	Comprabante De Pago	Comprabante De Pago	0	0
Doc14320_57_3766_1_2018-06-01_11.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14321_130_3766_1_2018-06-01_11.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14322_25_3766_1_2018-06-01_11.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14323_63_3766_1_2018-06-01_11.5		Comprabante De Pago	Comprabante De Pago	0	0
Doc14324_27_3766_1_2018-06-01_12.0	ELE'grE'fbs	Comprabante De Pago	Comprabante De Pago	0	0
Doc14325_36_8110_1_2018-06-01_12.0	Calle Luvs Anas	Comprabante De Pago	Comprabante De Pago	0	0
Doc14326_58_8235_1_2018-06-01_12.1	JULIA R.U.C.	Comprabante De Pago	Recibo	1	0
Doc14327_26_1450_1_2018-06-01_16.2	20259492965 mam	Comprabante De Pago	Comprabante De Pago	0	0
Doc14328_57_1450_1_2018-06-01_16.2	Jr 5m AgumN N' ms	Comprabante De Pago	Comprabante De Pago	0	0
Doc14329_45_1450_1_2018-06-01_16.3	zuzmzma mnx. zasn	Comprabante De Pago	Comprabante De Pago	0	0
Doc14330_63_1450_1_2018-06-01_16.3	zazmzsemmm-mquqm	Comprabante De Pago	Comprabante De Pago	0	0
Doc14331_66_1450_1_2018-06-01_16.3	2a255mzu5TELErAx	Comprabante De Pago	Comprabante De Pago	0	0
Doc14332_22_1450_1_2018-06-01_16.3	Jr 5m AgumN N' was um	Comprabante De Pago	Comprabante De Pago	0	0
Doc14333_22_6850_1_2018-06-01_17.1		Requerimiento	Comprabante De Pago	1	0
Doc14334_22_6850_1_2018-06-01_17.1		Requerimiento	Comprabante De Pago	1	0
Doc14335_21_1450_1_2018-06-01_17.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14336_25_1450_1_2018-06-01_17.2	Jr in AgumN N' in Um.	Comprabante De Pago	Comprabante De Pago	0	0
Doc14337_130_1450_1_2018-06-01_17.2	mam 2mm MJY -	Comprabante De Pago	Comprabante De Pago	0	0
Doc14338_23_7987_1_2018-06-01_17.2	Impresion 2: Página 1	Comprabante De Pago	Comprabante De Pago	0	0
Doc14339_66_7992_1_2018-06-01_17.3		Carta	Carta	0	0
Doc14340_7616_8238_1_2018-06-01_17.3	JOSE R.U.C.	Comprabante De Pago	Recibo	1	0
Doc14341_21_6679_1_2018-06-01_18.0	AL:	Comprabante De Pago	Comprabante De Pago	0	0
Doc14342_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14343_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14344_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14345_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14346_21_6679_1_2018-06-01_18.0	DIRECCION :	Comprabante De Pago	Comprabante De Pago	0	0
Doc14347_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14348_21_6679_1_2018-06-01_18.0		Comprabante De Pago	Comprabante De Pago	0	0
Doc14349_130_3486_1_2018-06-01_18.0	Formulario	Comprabante De Pago	Comprabante De Pago	1	0
Doc14350_21_6679_1_2018-06-01_18.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14351_21_6679_1_2018-06-01_18.2	FACTURA	Comprabante De Pago	Comprabante De Pago	0	0
Doc14352_21_6679_1_2018-06-01_18.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14353_21_6679_1_2018-06-01_18.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14354_21_6679_1_2018-06-01_18.2	SA. cane Montarosa N'	Comprabante De Pago	Comprabante De Pago	0	0
Doc14355_21_6679_1_2018-06-01_18.2		Comprabante De Pago	Comprabante De Pago	0	0
Doc14356_21_6679_1_2018-06-01_18.2		Comprabante De Pago	Comprabante De Pago	0	0

Doc14357_21_6679_1_2018-08-01_18:2		Comprobante De Pago	Comprobante De Pago	0	0
Doc14358_21_6679_1_2018-08-01_18:3	. W	Comprobante De Pago	Comprobante De Pago	0	0
Doc14359_21_6679_1_2018-08-01_18:3	o 0&0 o	Comprobante De Pago	Comprobante De Pago	0	0
Doc14360_21_6679_1_2018-08-01_18:3	E1	Comprobante De Pago	Comprobante De Pago	0	0
Doc14361_21_6679_1_2018-08-01_18:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14362_21_6679_1_2018-08-01_18:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14363_21_6679_1_2018-08-01_18:3	5.A Cans Mamansa N'	Comprobante De Pago	Comprobante De Pago	0	0
Doc14364_21_6679_1_2018-08-01_18:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14365_21_6679_1_2018-08-01_18:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14366_21_6679_1_2018-08-01_18:3	5A Cane Montemsa N'	Comprobante De Pago	Comprobante De Pago	0	0
Doc14367_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14368_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14369_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14370_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14371_21_6679_1_2018-08-01_18:4	UmChacantladelEsiarqu	Comprobante De Pago	Comprobante De Pago	0	0
Doc14372_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14373_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14374_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14375_21_6679_1_2018-08-01_18:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14376_21_6679_1_2018-08-01_18:4	h-ARKE'55,A~ Caus	Comprobante De Pago	Comprobante De Pago	0	0
Doc14377_21_6679_1_2018-08-01_18:5	. W	Comprobante De Pago	Comprobante De Pago	0	0
Doc14378_21_6679_1_2018-08-01_18:5	LA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14379_21_6679_1_2018-08-01_18:5	5,A~ Caila Maniensa N'	Comprobante De Pago	Comprobante De Pago	0	0
Doc14380_21_6679_1_2018-08-01_18:5	5A Caila Mnmanusa N'	Comprobante De Pago	Comprobante De Pago	0	0
Doc14381_21_6679_1_2018-08-01_18:5	Lilb.ChauarUladelEsi3n	Comprobante De Pago	Comprobante De Pago	0	0
Doc14382_21_6679_1_2018-08-01_18:5	FACTURA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14383_21_6679_1_2018-08-01_18:5	0.59..	Comprobante De Pago	Comprobante De Pago	0	0
Doc14384_21_6679_1_2018-08-01_18:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14385_21_6679_1_2018-08-01_18:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14386_45_3766_1_2018-08-02_09:0	EL'E'égEéñ?"	Comprobante De Pago	Liquidacion De Retribucio	1	0
Doc14387_21_345_1_2018-08-02_09:09	.Impwm -	Comprobante De Pago	Comprobante De Pago	0	0
Doc14388_31_1352_1_2018-08-02_09:2	HONORARIOS	Comprobante De Pago	Recibo	1	0
Doc14389_7816_6679_1_2018-08-02_10		Comprobante De Pago	Comprobante De Pago	0	0
Doc14390_7816_6679_1_2018-08-02_10	DXRECCION :	Comprobante De Pago	Comprobante De Pago	0	0
Doc14391_66_7992_1_2018-08-02_10:0		Carta	Informe	1	0
Doc14392_22_7360_1_2018-08-02_10:4		Certificado	Comprobante De Pago	1	0
Doc14393_66_390_1_2018-08-02_10:50	Impresion Página 1 de 1	Comprobante De Pago	Comprobante De Pago	0	0
Doc14394_21_4124_1_2018-08-02_11:3		Carta	Carta	0	0
Doc14395_7764_1178_1_2018-08-02_12		Resolucion Coactiva	Comprobante De Pago	1	0
Doc14396_94_6873_1_2018-08-02_12:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14397_94_6873_1_2018-08-02_12:1	Aguilna Hidalhnnda	Comprobante De Pago	Comprobante De Pago	0	0
Doc14398_38_302_1_2018-08-02_12:25	R.u.c. 10272862929	Comprobante De Pago	Recibo	1	0
Doc14399_38_302_1_2018-08-02_12:57		Comprobante De Pago	Recibo	1	0
Doc14400_66_8045_1_2018-08-02_13:0		Carta	Comprobante De Pago	1	0
Doc14401_21_7949_1_2018-08-02_15:4	SE m mu r. 70543n7mm	Comprobante De Pago	Comprobante De Pago	0	0
Doc14402_21_624_1_2018-08-02_15:45	Elac nomða - Imprasio ::	Comprobante De Pago	Comprobante De Pago	0	0
Doc14403_22_7331_1_2018-08-02_16:1	2018	Carta	Liquidacion De Retribucio	1	0
Doc14404_23_3758_1_2018-08-02_16:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14405_8121_8182_1_2018-08-02_16	RONAL DAVID R.U.C.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14406_7816_8182_1_2018-08-02_16		Comprobante De Pago	Recibo	1	0
Doc14407_21_3693_1_2018-08-02_16:3	O CORPORACION DE	Carta	Liquidacion De Retribucio	1	0
Doc14408_22_7522_1_2018-08-02_16:4	flz MSG (me	Carta	Carta	0	0
Doc14409_21_1099_1_2018-08-02_16:5	IUIK UISKW	Comprobante De Pago	Comprobante De Pago	0	0
Doc14410_21_1099_1_2018-08-02_16:5	013m	Comprobante De Pago	Comprobante De Pago	0	0
Doc14411_67_3282_1_2018-08-02_17:1		Comprobante De Pago	Comprobante De Pago	0	0

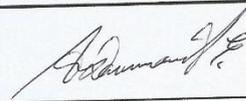
Doc14412_67_3262_1_2018-06-02_17:1	i	Comprobante De Pago	Comprobante De Pago	0	0
Doc14413_21_3706_1_2018-06-02_17:5	Run 20515543604 5:07!	Comprobante De Pago	Comprobante De Pago	0	0
Doc14414_25_3706_1_2018-06-02_18:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14415_23_3706_1_2018-06-02_18:0	Procesos Empresariales	Comprobante De Pago	Comprobante De Pago	0	0
Doc14416_22_3706_1_2018-06-02_18:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14417_57_3706_1_2018-06-02_18:0	DIRECCION	Comprobante De Pago	Comprobante De Pago	0	0
Doc14418_25_3766_1_2018-06-03_09:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14419_63_3766_1_2018-06-03_09:0	GRUPO CORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14420_23_3766_1_2018-06-03_09:0	GRUPO CORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14421_23_3766_1_2018-06-03_09:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14422_66_3766_1_2018-06-03_09:0	ELÉGGNGA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14423_66_3766_1_2018-06-03_09:1	GRUPOCORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14424_130_566_1_2018-06-03_09:1	, vmsuna PERU SAC.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14425_66_8241_1_2018-06-03_09:5	n3 AGUZMB	Carta Notarial	Carta	1	0
Doc14426_66_8221_1_2018-06-03_11:0		Carta	Certificado	1	0
Doc14427_45_6738_1_2018-06-03_11:3	zma275msr1	Comprobante De Pago	Recibo	1	0
Doc14428_45_6738_1_2018-06-03_11:3	207392370551 227637	Comprobante De Pago	Recibo	1	0
Doc14429_66_6738_1_2018-06-03_11:3	2a3027sm55T mam	Comprobante De Pago	Recibo	1	0
Doc14430_130_6738_1_2018-06-03_11:4	momna T amen Sims	Comprobante De Pago	Recibo	1	0
Doc14431_21_6738_1_2018-06-03_11:4	2asna7a7oss r:17 um	Comprobante De Pago	Recibo	1	0
Doc14432_21_6738_1_2018-06-03_11:4	2aau2757us\$ r mam	Comprobante De Pago	Recibo	1	0
Doc14433_25_6738_1_2018-06-03_11:4	PROPIETARIOS	Comprobante De Pago	Comprobante De Pago	0	0
Doc14434_26_6738_1_2018-06-03_11:4	20302757955 r mew	Comprobante De Pago	Recibo	1	0
Doc14435_63_749_1_2018-06-03_11:59	Multisectorial	Comprobante De Pago	Carta	1	0
Doc14436_21_8243_1_2018-06-03_12:0	Pisa 12 um, Chasar'la	Comprobante De Pago	Carta	1	0
Doc14437_21_749_1_2018-06-03_12:07	Mulfisectorial GCST-	Comprobante De Pago	Carta	1	0
Doc14438_21_8146_1_2018-06-03_12:0	E'ecubmca - Impresion -	Comprobante De Pago	Comprobante De Pago	0	0
Doc14439_63_8146_1_2018-06-03_12:1	Impresion :1.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14440_59_7388_1_2018-06-03_12:1	RECEPCION	Carta	Carta	0	0
Doc14441_59_7388_1_2018-06-03_12:3	RECEPCION	Carta	Carta	0	0
Doc14442_59_7388_1_2018-06-03_12:3	RECEPCION	Carta	Carta	0	0
Doc14443_91_7239_1_2018-06-03_12:4	i	Comprobante De Pago	Comprobante De Pago	0	0
Doc14444_91_7239_1_2018-06-03_12:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14445_116_3950_1_2018-06-03_13:0	INVERSION -	Comprobante De Pago	Comprobante De Pago	0	0
Doc14446_59_7388_1_2018-06-03_13:0		Carta	Carta	0	0
Doc14447_26_7876_1_2018-06-03_15:3	Electkanica - Impres'ion :	Comprobante De Pago	Comprobante De Pago	0	0
Doc14448_66_8244_1_2018-06-03_15:3	F RECEPCION	Carta	Carta Notarial	1	0
Doc14449_7816_8215_1_2018-06-03_15:3	' J FACTURA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14450_22_3208_1_2018-06-03_16:1		Carta	Carta De Renuncia	1	0
Doc14451_22_647_1_2018-06-03_16:22	SOLSTICIO S.A.C.	Carta	Carta	0	0
Doc14452_23_3770_1_2018-06-03_16:5	icwsns DESDE: woe/2m	Transferencia	Comprobante De Pago	1	0
Doc14453_25_3282_1_2018-06-03_17:2	OOOZCO5OO	Comprobante De Pago	Comprobante De Pago	0	0
Doc14454_23_3732_1_2018-06-06_09:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14455_22_7414_1_2018-06-06_09:3		Carta	Estados Financieros	1	0
Doc14456_22_7409_1_2018-06-06_09:4		Carta	Estados Financieros	1	0
Doc14457_22_6849_1_2018-06-06_09:4		Carta	Estados Financieros	1	0
Doc14458_25_6849_1_2018-06-06_09:4		Carta	Estados Financieros	1	0
Doc14459_25_7414_1_2018-06-06_09:4		Carta	Estados Financieros	1	0
Doc14460_25_7409_1_2018-06-06_09:4		Carta	Estados Financieros	1	0
Doc14461_25_3726_1_2018-06-06_10:4	S...AC	Comprobante De Pago	Comprobante De Pago	0	0
Doc14462_23_3726_1_2018-06-06_10:4	(\21 Wari Impresores	Comprobante De Pago	Comprobante De Pago	0	0
Doc14463_25_3726_1_2018-06-06_10:4	FACTURA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14464_7816_8245_1_2018-06-06_11:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14465_7816_8245_1_2018-06-06_11:0	20515802650	Comprobante De Pago	Comprobante De Pago	0	0
Doc14466_38_8234_1_2018-06-06_11:0	Mili*	Comprobante De Pago	Comprobante De Pago	0	0

Doc14467_7816_8234_1_2018-06-06_11:1	* 1 311	Comprobante De Pago	Liquidacion De Retribucio	1	0
Doc14468_38_8234_1_2018-06-06_11:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14469_7816_8234_1_2018-06-06_11:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14470_66_3706_1_2018-06-06_11:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14471_72_3464_1_2018-06-06_11:2	171139111516	Comprobante De Pago	Comprobante De Pago	0	0
Doc14472_21_8194_1_2018-06-06_11:2		Comprobante De Pago	Comprobante De Pago	0	0
Doc14473_21_8246_1_2018-06-06_11:2	Fideicomiso.	Comprobante De Pago	Carta	1	0
Doc14474_21_8194_1_2018-06-06_11:2	en Fideicomiso.	Comprobante De Pago	Carta	1	0
Doc14475_21_6679_1_2018-06-06_11:2		Comprobante De Pago	Comprobante De Pago	0	0
Doc14476_7816_8231_1_2018-06-06_11:1	VALENTINA R.U.C.	Comprobante De Pago	Recibo	1	0
Doc14477_7816_8247_1_2018-06-06_11:1	BRUNELLA VERONICA	Comprobante De Pago	Informe	1	0
Doc14478_66_8228_1_2018-06-06_11:3		Carta	Comprobante De Pago	1	0
Doc14479_22_6850_1_2018-06-06_12:3	Carla Norria!	Carta	Carta Notarial	1	0
Doc14480_63_4124_1_2018-06-06_12:5		Carta	Informe	1	0
Doc14481_21_8248_1_2018-06-06_15:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14482_141_3706_1_2018-06-06_15:1	Procesos EmBresariales	Comprobante De Pago	Comprobante De Pago	0	0
Doc14483_23_3770_1_2018-06-06_15:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14484_25_7955_1_2018-06-06_15:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14485_7816_8249_1_2018-06-06_15:1	MARIA VIRGINIA R.U.C.	Comprobante De Pago	Recibo	1	0
Doc14486_26_450_1_2018-06-06_16:00	gadondevwir	Comprobante De Pago	Comprobante De Pago	0	0
Doc14487_21_8250_1_2018-06-06_16:4	2018	Carta	Liquidacion De Retribucio	1	0
Doc14488_27_8251_1_2018-06-06_17:5	MOTOR MECHANIC SAC.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14489_7816_8215_1_2018-06-07_11:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14490_22_193_1_2018-06-07_18:17	P ERUANA DE	Carta	Comprobante De Pago	1	0
Doc14491_21_8199_1_2018-06-08_09:3	LOSTAUNAU	Comprobante De Pago	Comprobante De Pago	0	0
Doc14492_22_299_1_2018-06-08_09:37	R.U.C. 20600592646	Comprobante De Pago	Comprobante De Pago	0	0
Doc14493_8121_8099_1_2018-06-08_09:04	S.A.C.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14494_45_3766_1_2018-06-08_09:3	GRUPO CORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14495_63_3766_1_2018-06-08_09:4	ELECTRONICA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14496_63_3766_1_2018-06-08_09:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14497_63_3766_1_2018-06-08_09:4	GRUPO CORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14498_27_3766_1_2018-06-08_09:4	ELECTRONICA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14499_63_3766_1_2018-06-08_09:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14500_25_3766_1_2018-06-08_09:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14501_45_3766_1_2018-06-08_09:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14502_22_6850_1_2018-06-08_10:2		Requerimiento	Comprobante De Pago	0	1
Doc14503_22_6850_1_2018-06-08_10:3		Requerimiento	Comprobante De Pago	0	1
Doc14504_23_3770_1_2018-06-08_10:4		Transferencia	Comprobante De Pago	0	1
Doc14505_23_3770_1_2018-06-08_10:5		Transferencia	Comprobante De Pago	0	1
Doc14506_23_3770_1_2018-06-08_10:5		Circular	Liquidacion De Retribucio	1	0
Doc14507_59_7388_1_2018-06-08_11:1		Carta	Carta	0	0
Doc14508_38_7860_1_2018-06-08_11:2	Electronica ~ Impresion	Comprobante De Pago	Comprobante De Pago	0	0
Doc14509_22_6849_1_2018-06-08_11:3	2018	Carta	Carta De Renuncia	1	0
Doc14510_79_3766_1_2018-06-08_11:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14511_79_3766_1_2018-06-08_11:4	GRUPO CORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14512_79_3766_1_2018-06-08_11:4	GRUPOCORIL	Comprobante De Pago	Comprobante De Pago	0	0
Doc14513_29_3766_1_2018-06-08_11:4	ELECTRONICA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14514_64_3766_1_2018-06-08_11:4	ELEFE'Éfb	Comprobante De Pago	Comprobante De Pago	0	0
Doc14515_26_8180_1_2018-06-08_11:5	20548042888	Comprobante De Pago	Comprobante De Pago	0	0
Doc14516_38_8180_1_2018-06-08_11:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14517_8121_8099_1_2018-06-08_12:0	S.A.C.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14518_38_8099_1_2018-06-08_12:0	S.A.C.	Comprobante De Pago	Comprobante De Pago	0	0
Doc14519_45_8253_1_2018-06-08_12:2		Carta De Renuncia	Carta De Renuncia	0	0
Doc14520_38_3124_1_2018-06-08_12:2	de 2019	Carta	Carta De Renuncia	1	0
Doc14521_22_6848_1_2018-06-08_12:3	I R:CEPCION	Sesion De La Comision Adm	Carta	1	0

Doc14522_66_1070_1_2018-08-08_12:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14523_66_770_1_2018-08-08_12:44		Comprobante De Pago	Comprobante De Pago	0	0
Doc14524_7764_7407_1_2018-08-08_12	ELECTRONICA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14525_57_3766_1_2018-08-08_12:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14526_141_3766_1_2018-08-08_12	EL'E'égÉÉTH'CA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14527_25_3766_1_2018-08-08_12:5	ELECTRONICA	Comprobante De Pago	Comprobante De Pago	0	0
Doc14528_130_3766_1_2018-08-08_12		Comprobante De Pago	Comprobante De Pago	0	0
Doc14529_21_7222_1_2018-08-08_12:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14530_21_1674_1_2018-08-08_12:5	PERU S.A.	Comprobante De Pago	Informe	1	0
Doc14531_7816_6254_1_2018-08-08_13	HUMBERTO R.U.C.	Comprobante De Pago	Recibo	1	0
Doc14532_23_3766_1_2018-08-08_13:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14533_23_3766_1_2018-08-08_13:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14534_22_6679_1_2018-08-08_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14535_22_6679_1_2018-08-08_15:5	A	Comprobante De Pago	Comprobante De Pago	0	0
Doc14536_22_6679_1_2018-08-08_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14537_22_6679_1_2018-08-08_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14538_22_6679_1_2018-08-08_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14539_22_6679_1_2018-08-08_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14540_23_8255_1_2018-08-08_16:1	2018.	Carta	Carta De Renuncia	1	0
Doc14541_23_7762_1_2018-08-08_16:4		Oficio	Informe	1	0
Doc14542_146_647_1_2018-08-08_16:4	INMOBILIARIAS	Comprobante De Pago	Comprobante De Pago	0	0
Doc14543_21_8256_1_2018-08-08_17:3	R 8 A60 lei	Carta Notarial	Carta De Renuncia	1	0
Doc14544_23_3732_1_2018-08-08_17:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14545_21_3788_1_2018-08-09_10:0	ImnwchIASADQVEam	Comprobante De Pago	Comprobante De Pago	0	0
Doc14546_66_3766_1_2018-08-09_10:0		Carta	Carta Notarial	1	0
Doc14547_22_8016_1_2018-08-09_12:2		Carta	Carta	0	0
Doc14548_25_784_1_2018-08-09_12:27	COMPROBANTE DE	Comprobante De Pago	Comprobante De Pago	0	0
Doc14549_25_3706_1_2018-08-09_12:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14550_25_3706_1_2018-08-09_12:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14551_25_3706_1_2018-08-09_12:3	Procesos Empresariales	Comprobante De Pago	Comprobante De Pago	0	0
Doc14552_25_3706_1_2018-08-09_12:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14553_23_3706_1_2018-08-09_12:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14554_23_3706_1_2018-08-09_12:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14555_23_3706_1_2018-08-09_12:4		Comprobante De Pago	Comprobante De Pago	0	0
Doc14556_57_3706_1_2018-08-09_12:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14557_57_3706_1_2018-08-09_12:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14558_57_3706_1_2018-08-09_12:5	2	Comprobante De Pago	Comprobante De Pago	0	0
Doc14559_21_3706_1_2018-08-09_13:0	A	Comprobante De Pago	Comprobante De Pago	0	0
Doc14560_21_3706_1_2018-08-09_13:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14561_21_3706_1_2018-08-09_13:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14562_83_3706_1_2018-08-09_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14563_83_3706_1_2018-08-09_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14564_22_3706_1_2018-08-09_15:5		Comprobante De Pago	Comprobante De Pago	0	0
Doc14565_22_3706_1_2018-08-09_16:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14566_22_3706_1_2018-08-09_16:0		Comprobante De Pago	Comprobante De Pago	0	0
Doc14567_63_3706_1_2018-08-09_16:0	R.U.C.: 20543163130	Comprobante De Pago	Comprobante De Pago	0	0
Doc14568_29_3706_1_2018-08-09_16:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14569_79_3706_1_2018-08-09_16:1	Procesos Empresariales	Comprobante De Pago	Comprobante De Pago	0	0
Doc14570_66_3706_1_2018-08-09_16:1		Comprobante De Pago	Comprobante De Pago	0	0
Doc14571_22_6850_1_2018-08-09_16:1	Requerimiento	Comprobante De Pago	1	0
Doc14572_22_193_1_2018-08-09_16:19	EERUANA DE	Carta	Comprobante De Pago	1	0
Doc14573_23_3726_1_2018-08-09_16:3	(N7.) Wari Impresores s	Comprobante De Pago	Comprobante De Pago	0	0
Doc14574_130_7246_1_2018-08-09_16:3		Comprobante De Pago	Comprobante De Pago	0	0
Doc14575_45_7246_1_2018-08-09_16:3	SERVICIO DE	Comprobante De Pago	Comprobante De Pago	0	0
Doc14576_23_3704_1_2018-08-09_16:4	ENCUADERNACIONES	Comprobante De Pago	Comprobante De Pago	0	0

Observaciones:

Este cuadro permitirá observar el porcentaje de asertividad del algoritmo bayesiano en el periodo de noviembre del 2018 con el antiguo sistema.



DATA SYSTEM
Alex Durand Córdova
Gerente General

ANEXO 11: Acta de aprobación de Originalidad de tesis

 UCV UNIVERSIDAD CÉSAR VALLEJO	ACTA DE APROBACIÓN DE ORIGINALIDAD DE TESIS	Código : F06-PP-PR-02.02 Versión : 09 Fecha : 23-03-2018 Página : 1 de 1
--	---	---

Yo, **DR. HILARIO FALCON MANUEL**, docente de la Facultad Ingeniería y Escuela Profesional Ingeniería de Sistemas de la Universidad César Vallejo Sede Lima Este, revisor (a) de la tesis titulada

“Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril”, del (de la) estudiante **VARGAS SALAS JOSE FERNANDO**, constató que la investigación tiene un índice de similitud de 19% verificable en el reporte de originalidad del programa Turnitin.

El/la suscrito (a) analizó dicho reporte y concluyó 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.

Lima, San Juan de Lurigancho 06 de diciembre del 2018



DR. HILARIO FALCON MANUEL

DNI: 10132075

			
Elabora	Dirección de Investigación	Revisó	Vicerectorado de Investigación

ANEXO 12: Autorización de la versión final del trabajo de investigación



UNIVERSIDAD CÉSAR VALLEJO

AUTORIZACIÓN DE LA VERSIÓN FINAL DEL TRABAJO DE INVESTIGACIÓN

CONSTE POR EL PRESENTE EL VISTO BUENO QUE OTORGA EL ENCARGADO DE INVESTIGACIÓN DE

Mg. María Acuña Meléndez

A LA VERSIÓN FINAL DEL TRABAJO DE INVESTIGACIÓN QUE PRESENTA:

Vargas Salas Jose Fernando

INFORME TÍTULADO:

Implantación de un modelo para Clasificación Automática de Textos en Grupo Coril

PARA OBTENER EL TÍTULO O GRADO DE:

INGENIERO DE SISTEMAS

SUSTENTADO EN FECHA: 06 DE DICIEMBRE DEL 2018

NOTA O MENCIÓN: (12) (DOCE)

Mg. María Acuña Meléndez
CP de Ingeniería de Sistemas campus Lima Este

	Elabora	Dirección de Investigación	Revisó			Vicerrectorado de Investigación
--	---------	----------------------------	--------	--	--	---------------------------------

ANEXO 13: Porcentaje del Turnitin

Feedback Studio - Google Chrome
 https://ev.turnitin.com/app/carta/ev?u=1066256785&o=1100524643&session-id=4f4c8ec083b8986b6b1b81482494420633d5=&lang=es&student_user=1

feedback studio

jose fernando vargas salas Implantación de un modelo para Clasificación Automática de Textos en Grupo Corti

UNIVERSIDAD CÉSAR VALLEJO
 FACULTAD DE INGENIERIA DE SISTEMAS
 ESCUELA PROFESIONAL DE INGENIEROS DE SISTEMAS

"Implantación de un modelo para Clasificación Automática de Textos en Grupo Corti"

TESIS PARA OBTENER EL TITULO PROFESIONAL DE INGENIERIA DE SISTEMAS

AUTOR:
 Vargas Salas José Fernando

ASESOR:
 Mg. MANUEL HILARIO FALCÓN

LÍNEA DE INVESTIGACIÓN:
 SISTEMAS DE INFORMACIÓN TRANSACCIONALES

LIMA - PERÚ
 (2018)

Página: 1 de 79 Número de palabras: 15757

Resumen de coincidencias **19%**

1	repositorio.ucv.edu.pe Fuente de internet	4%
2	Entregado a Universidad... Trabajo del estudiante	3%
3	tesis.pucp.edu.pe Fuente de internet	1%
4	Entregado a Pontificia... Trabajo del estudiante	1%
5	oa.upm.es Fuente de internet	1%
6	Entregado a Universidad... Trabajo del estudiante	1%
7	gredos.usal.es Fuente de internet	<1%
8	Entregado a Universidad... Trabajo del estudiante	<1%
9	opac.pucv.cl Fuente de internet	<1%
10	ruja.ulb.es Fuente de internet	<1%
11	alicia.concytec.gob.pe Fuente de internet	<1%

Text-only Report | Turnitin Classic | High Resolution **Activado**