

**UNIVERSIDAD NACIONAL DE TRUJILLO**

**Facultad de Ciencias Físicas y Matemáticas**

**Escuela Profesional de Informática**



**Aplicación de Sistemas Inteligentes para la Clasificación  
Automática de Documentos**

**TESIS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO  
INFORMÁTICO**

**AUTORES:** Contreras Pumamango Luis Anthony  
Sanchez Rodriguez Paul Denis

**ASESORA:** Mg. Sifuentes Díaz Yenny Milagritos

**TRUJILLO - PERÚ**

**2019**

## Dedico esta tesis a:

*Mis padres, por el gran apoyo recibido en todo momento, por sus consejos, por la motivación constante que me ha permitido ser una persona de bien, pero sobre todo por su confianza.*

*Danna Stephanía, por ser la personita que alegra mis días y por quien debo esforzarme para ser mejor en este largo camino.*

*Mi tío Antonio, siendo para mí un gran ejemplo a seguir, en lo profesional y en la calidad de ser humano.*

*Mis profesores, por su gran apoyo y motivación para la culminación de nuestros estudios profesionales y para la elaboración de esta tesis; a la profesora Yenny Milagritos Sifuentes Díaz por su apoyo y paciencia ofrecido en este trabajo.*

*Contreras Pumamango Luis Anthony*

## Dedico esta tesis a:

*Mis padres, por las grandes enseñanzas de la vida y su incondicional apoyo,  
por sus ánimos de verme cumplir mis metas.*

*Mi esposa la Sra. Alexandra Anastacio Avalos e hijo Francisco, por ser el  
motivo de querer superarme profesionalmente hacia un futuro mejor.*

*Mis profesores, los cuales me enseñaron los conocimientos científicos ad-  
quiridos en la Universidad Nacional de Trujillo y enseñanzas de la vida.*

*Sanchez Rodriguez Paul Denis*

## Agradecimientos

### Agradezco a:

*Mis padres, por el apoyo moral y los consejos dados. A Danna por sus ocupaciones y hacer divertidos mis días.*

*Mi asesora, Mg. Yenny Milagritos Sifuentes Díaz, por la base fundamental en nuestro desarrollo de esta tesis.*

*Nuestra casa de estudios, Universidad Nacional de Trujillo, en la cual junto a grandes profesores se desarrolla la labor de inculcar conocimientos y valores para futuros profesionales.*

*Contreras Pumamango Luis Anthony*

## Agradezco a:

*Mis padres, por el gran apoyo que me dan cada día, por sus buenos consejos de vida.*

*Mis profesores del Departamento de Informática, de los cuales recibí una gran cantidad de conocimientos y enseñanzas de la vida misma.*

*Mi asesora Mg. Yenny Milagritos Sifuentes Díaz, que siempre se mostró disponible e interesada en ayudarme.*

*Mi compañero de tesis el Sr. Luis Anthony Contreras Pumamango, por la compañía e interés en el desarrollo de la presente tesis.*

*Sánchez Rodríguez Paul Denis*

UNIVERSIDAD NACIONAL DE TRUJILLO  
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS  
ESCUELA PROFESIONAL DE INFORMATICA

ACTA DE SUSTENTACION DEL TRABAJO DE TESIS

En la Universidad Nacional de La Libertad – Trujillo, Facultad de Ciencias Físicas y Matemáticas, siendo las 10:00 a.m. del día miércoles 20 de noviembre del 2019, se reunieron los señores Profesores Ricardo Guevara Ruíz (Presidente), Liz Pedro Huamán (Secretario) y Yenny Sifuentes Díaz (Vocal-Asesor), integrantes del Jurado nombrados para revisar, estudiar y dictaminar el Trabajo de Tesis presentado por los Bachilleres CONTRERAS PUMAMANGO LUIS ANTHONY y SANCHEZ RODRIGUEZ PAUL DENIS, aspirantes al Título de INGENIERO INFORMATICO. Con la concurrencia del público, el Señor Presidente del Jurado dio por abierta la actuación.

Acto seguido, los aspirantes sustentaron el Trabajo de Tesis intitulado “APLICACIÓN DE SISTEMAS INTELIGENTES PARA LA CLASIFICACION AUTOMATICA DE DOCUMENTOS” respondiendo.....SATISFACTORIAMENTE..... las preguntas formuladas por los Señores Miembros del Jurado.

Luego procedió a la evaluación en votación secreta y escrita, constatándose que CONTRERAS PUMAMANGO LUIS ANTHONY y SANCHEZ RODRIGUEZ PAUL DENIS, han sido.....APROBADOS POR MAYORIA....., con lo que el Señor Presidente del Jurado dio por terminada la actuación, sentándose la presente acta que firmaron para constancia, en la ciudad de Trujillo a los veinte días del mes de noviembre del año dos mil diecinueve.

  
Prof. Ricardo Guevara Ruiz  
Presidente

  
Prof. Liz Pedro Huamán  
Secretario

  
Prof. Yenny Sifuentes Díaz  
Vocal

# Resumen

Actualmente la importancia que le dan las empresas e instituciones a la clasificación de sus documentos, se vuelve uno de los principales temas debido a la laboriosidad y el tiempo demás para dicha ejecución, pues el gran aumento de información digital y la necesidad de buscar y ubicar la información en el menor tiempo, lo convierte en un tema esencial del cual deben preocuparse las empresas y centros administrativos. La presente tesis busca resolver el problema, mostrando una forma de clasificación automática de documentos. Existen distintas formas de clasificar y diversos conocimientos que ayudarían a mejorar la clasificación de estos, lo cual es necesario investigar sobre los sistemas inteligentes. Se diseñó un sistema de clasificación usando como herramienta principal las redes neuronales. Implementamos una aplicación para nuestro problema, optimizando el tiempo de búsqueda de documentos en más de 760 segundos, logramos mostrar una sensibilidad de 94.6 %, una precisión del 100 % y una exactitud de 94.7 % para el criterio de subir con éxito y una sensibilidad de 92.42 %, una precisión de 99.18 % y una exactitud de 91.7 % para el criterio documento sin error ortográfico. Finalmente analizando los resultados obtenidos de nuestra aplicación se contrastó la hipótesis planteada, y se logró los objetivos propuestos de sensibilidad, precisión, exactitud y tiempo.

**Palabras claves:** redes neuronales, clasificación de documentos, sistemas inteligentes.

## Abstract

Currently the importance that companies and institutions give to the classification of their documents, is once again one of the main issues due to the hard work and time for this execution, since the great increase in digital information and the need to search and locate the information in the shortest time, which becomes an essential issue of which the companies and administrative centers. This thesis seeks to solve the problem, showing a way of automatic classification of documents. There are different forms of classification and diverse knowledge that help to improve the classification of these, which is necessary to investigate on intelligent systems. A classification system was designed using neural networks as the main tool. We implemented an application for our problem, optimizing the document search time in more than 760 seconds, we managed to show a sensitivity of 94.6 %, an accuracy of 100 % and an accuracy of 94.7 % for the criterion of uploading successfully and a sensitivity of 92.42 %, an accuracy of 99.18 % and an accuracy of 91.7 % for the document criterion without spelling error. Finally, analyzing the results obtained from our application, the proposed hypothesis is contrasted, and the proposed objectives of sensitivity, precision, precision and time are found.

**Keywords:** neural networks, document classification, intelligent systems.

# Índice de figuras

2.1.	Modelo definidos por KADS. . . . .	7
2.2.	Red neuronal artificial. . . . .	12
2.3.	Gráfico de la función sigmoide. . . . .	14
2.4.	Representación de un red multicapa. . . . .	17
2.5.	Ecuaciones del backpropagation. . . . .	19
2.6.	Partes del oficio. . . . .	23
2.7.	Partes del oficio múltiple. . . . .	24
2.8.	Partes del informe. . . . .	25
2.9.	Solicitud. . . . .	26
2.10.	Resolución rectoral. . . . .	27
2.11.	Partes del memorandum. . . . .	28
2.12.	Editor Joone. . . . .	30
2.13.	Diagrama de tipos de documentos. . . . .	33
2.14.	Diagrama de tipos de documentos. . . . .	34
3.1.	Requerimiento funcional de investigación. . . . .	37
3.2.	Modelo de diagrama de clases para la investigación de la tesis. . . . .	38
3.3.	Modelo de estados para la investigación de la tesis. . . . .	39
3.4.	Modelo de casos de uso para la investigación de la tesis. . . . .	39

3.5. Diagrama de componentes de nuestra investigación. . . . .	40
3.6. Modelo de secuencia para clasificar automáticamente documentos. . . . .	41
3.7. Modelo de secuencia para reporte de documentos. . . . .	41
3.8. Diagrama de la base de datos de entidad relación. . . . .	42
3.9. Login de la aplicación. . . . .	42
3.10. Interfaz principal. . . . .	43
3.11. Interfaz de la carga de los documentos. . . . .	43
3.12. Interfaz de la búsqueda múltiple de documentos. . . . .	44
3.13. Interfaz de reporte de los documentos. . . . .	44
3.14. Diagrama de flujo para el desarrollo de la investigación. . . . .	45
3.15. Extracción de datos. . . . .	46
3.16. Descripción del patrón. . . . .	47
3.17. Arquitectura de la aplicación. . . . .	47
3.18. Red neuronal en Joone. . . . .	48
3.19. Entrenamiento de las 5000 épocas. . . . .	48
3.20. Pesos de salida. . . . .	49
3.21. Pesos de salida. . . . .	50
3.22. Bias. . . . .	50
3.23. Inicio de vector característico. . . . .	51
3.24. Activando neuronas en la capa de entrada. . . . .	51
3.25. Activando neuronas en la capa oculta. . . . .	52
3.26. Activando neuronas en la capa de salida. . . . .	52
3.27. Resultado de la función sigmoide en la aplicación. . . . .	53

4.1.	Resultado detallado para la búsqueda, segun criterio subir con éxito . . . . .	55
4.2.	Tabla cruzada: subir con exito vs. encontró el doc. . . . .	55
4.3.	Resultado detallado para la búsqueda segun criterio sin error ortográfico. . .	57
4.4.	Tabla cruzada: sin error ortográfico vs. encontró el doc. . . . .	57
4.5.	Trabajadores antes de usar la aplicación. . . . .	59
4.6.	Trabajadores usando la aplicación. . . . .	59
A.1.	Inicio de la interfaz de la aplicación. . . . .	72
A.2.	Interfaz de administración de los usuarios. . . . .	72
A.3.	Interfaz de registro de actividades. . . . .	73
A.4.	Destino de los documentos. . . . .	73
A.5.	Tipos de documentos. . . . .	73
A.6.	Función para cargar muchos documentos a la vez. . . . .	74
A.7.	La librería POI permite ingresar los tipos de documentos: .doc, .docx y .pdf.	74
A.8.	Evaluación del arreglo. . . . .	75
A.9.	Se muestra los valores de la red neuronal. . . . .	76
A.10.	Se muestra los nombres de la variables del usuario. . . . .	76

# Índice de tablas

2.1. Componentes del perceptrón. . . . .	16
2.2. Indicadores para en nuestra investigación. . . . .	32
2.3. Datos obtenidos para la investigación. . . . .	33
2.4. Tipos de documentos y sus cantidades. . . . .	34
3.1. Requerimiento no funcional de nuestra investigación. . . . .	38
3.2. Datos del entrenamiento . . . . .	49
3.3. Vector de aprendizaje. . . . .	49
4.1. Tabla de valores predictivos. . . . .	55
4.2. Tabla base de comparación de tiempos . . . . .	58
4.3. Resultados con tiempos, sin usar la aplicación y usando la aplicación . . . . .	59
A.1. Tabla con los documentos de nuestra población . . . . .	65
A.2. Encuesta a trabajadores sin la aplicación. . . . .	69
A.3. Encuesta a trabajadores con la aplicación. . . . .	70
A.4. Destino de los documentos. . . . .	71

# Índice general

<b>Dedicatoria</b>	I
<b>Dedicatoria</b>	II
<b>Agradecimientos</b>	III
<b>Resumen</b>	VI
<b>Abstract</b>	VII
<b>Índice de Figuras</b>	VIII
<b>Índice de Tablas</b>	XI
<b>1. Introducción</b>	1
1.1. Justificación de la investigación . . . . .	2
1.2. Formulación del problema . . . . .	2
1.3. Hipótesis . . . . .	3
1.4. Objetivos . . . . .	3
1.4.1. Generales . . . . .	3
1.4.2. Específicos . . . . .	3
1.5. Estructura de la tesis . . . . .	3
	XII

<b>2. Materiales y Métodos</b>	<b>5</b>
2.1. Marco teórico . . . . .	5
2.1.1. Metodología Common Kads . . . . .	5
2.1.1.1. Descripción . . . . .	6
2.1.2. Inteligencia Artificial (IA) . . . . .	7
2.1.3. Sistemas inteligentes . . . . .	8
2.1.3.1. Capacidades requeridas . . . . .	9
2.1.4. Red Neuronal Artificial (RNA) . . . . .	10
2.1.5. Tipos de aprendizaje . . . . .	11
2.1.5.1. Aprendizaje supervisado . . . . .	11
2.1.6. Elementos básicos . . . . .	11
2.1.6.1. Función de entrada (input function) . . . . .	12
2.1.6.2. Función de activación (activation function) . . . . .	13
2.1.6.3. Función de salida (output function) . . . . .	14
2.1.7. Perceptrón . . . . .	15
2.1.7.1. Perceptrón simple . . . . .	15
2.1.7.2. Perceptrón Multicapa . . . . .	16
2.1.8. Backpropagation . . . . .	17
2.1.9. Clasificación automática de documentos . . . . .	19
2.1.9.1. Métodos de clasificación automática de documentos . . . . .	20
2.1.9.2. Representación vectorial . . . . .	20
2.1.9.3. Centroide . . . . .	21
2.1.9.4. Similitud . . . . .	21
2.1.10. Codificación de los datos de entrada. . . . .	22

2.1.11.	Descripción documentos en la investigación . . . . .	23
2.1.12.	Mysql Worbench . . . . .	28
2.1.13.	Netbeans . . . . .	29
2.1.14.	Java . . . . .	29
2.1.15.	Joone . . . . .	29
	2.1.15.1. Características . . . . .	30
2.2.	Método de la investigación . . . . .	31
2.2.1.	Tipo de investigación . . . . .	31
2.2.2.	Variables de la Investigación . . . . .	31
	2.2.2.1. Variable Dependiente . . . . .	31
	2.2.2.2. Variable Independiente . . . . .	31
	2.2.3. Operacionalización de la variable . . . . .	31
2.3.	Recolección de datos para la elaboración del modelo . . . . .	32
2.3.1.	Técnica . . . . .	32
2.3.2.	Población . . . . .	32
2.3.3.	Muestra . . . . .	33
	2.3.3.1. Muestra por conveniencia . . . . .	34
2.4.	Etapas de la investigación . . . . .	34
<b>3.</b>	<b>Sistema inteligente para la clasificación automática de documentos</b>	<b>36</b>
3.1.	Análisis . . . . .	36
	3.1.1. Requerimiento funcional . . . . .	36
	3.1.2. Requerimiento no funcional . . . . .	37
3.2.	Diseño . . . . .	38

3.2.1. Diagrama de clases . . . . .	38
3.2.2. Diagrama de modelos de estados . . . . .	39
3.2.3. Diagrama de casos de uso . . . . .	39
3.2.4. Diagrama de componentes . . . . .	40
3.2.5. Diagrama de secuencia . . . . .	41
3.2.6. Diagrama de la base de datos entidad relación . . . . .	42
3.2.7. Diseño de interfaces . . . . .	42
3.2.8. Diseño de algoritmo . . . . .	45
3.2.9. Extracción de datos . . . . .	46
3.2.10. Arquitectura de la aplicación . . . . .	47
3.3. Implementación . . . . .	47
3.4. Entrenamiento . . . . .	48
3.5. Funcionamiento de las redes neuronales . . . . .	50
3.6. Función activación . . . . .	53
<b>4. Resultados y discusión de la tesis</b>	<b>54</b>
4.1. Resultados computacionales . . . . .	54
4.1.1. Resultado según criterio subir con éxito . . . . .	54
4.1.2. Resultado según criterio sin error ortográfico . . . . .	56
4.2. Tiempo . . . . .	58
<b>5. Consideraciones finales</b>	<b>60</b>
5.1. Conclusiones . . . . .	60
5.2. Trabajos futuros . . . . .	61
<b>Referencias bibliográficas</b>	<b>62</b>

**A. Código de la aplicación**

**65**

BIBLIOTECA DE CIENCIAS  
Y MATEMÁTICAS FÍSICAS

XVI

# Capítulo 1

## Introducción

Al tener un trabajo de oficina y estar constantemente moviendo documentos para realizar algún trámite o por simple protocolo, nos damos cuenta que, esto nos consume mucho tiempo, que se puede utilizar en la realización de otras actividades o pendientes de mayor importancia, y no suficiente con eso tenemos que tener en cuenta el tiempo que utilizamos para buscar y peor aun cuando no se recuerda el nombre con el cual fue guardado en la computadora. Este es un problema típico que se padece en muchas empresas y/o instituciones públicas ya que, a pesar de contar con un software para la derivación del documento, llámese oficio, solicitud, informe, resolución rectoral, memorándum, etc., se necesita mover el oficio en físico y por ende tener una copia en tu computadora.

Pensando en ello y en el objetivo de optimizar ese proceso, presentamos una aplicación para la clasificación de automática de documentos, la cual se basa en la búsqueda y la clasificación del contenido para obtener como resultado el tipo de documento o el destino, hacia dónde va dirigido, se tomaron estas dos opciones de búsqueda para facilitar al usuario teniendo en cuenta que la memoria puede ser muy frágil; y estas dos opciones, tipo de documento y destino, son las palabras clave que se debe tener presente para empezar su búsqueda, con esta aplicación se muestra la eficiencia de nuestra tesis y el gran uso que se le puede dar.

Teniendo como interrogante ¿Qué es lo que hace especial a nuestra aplicación?, al aplicar

sistemas inteligentes, las redes neuronales nos permite tener resultados favorables y auto regulables, ya que al manejar redes neuronales nuestro sistema va aprendiendo y a mayor entrenamiento el grado de certeza aumenta. Este sistema está pensado para un fácil manejo, utilizando interfaces muy amigables, teniendo en cuenta que los principales usuarios serán personal administrativo, y como lugar de aplicación la Universidad Nacional de Trujillo, siendo por este motivo que se realiza la aplicación teniendo como datos de ejemplo documentos de la Universidad Nacional de Trujillo, los cuales será nuestra población.

## **1.1. Justificación de la investigación**

La importancia de esta investigación desde un punto de vista informático se justifica académicamente, como un aporte para la investigación y una motivación para profundizar en inteligencia artificial, redes neuronales.

Este proyecto se justifica socialmente, ya que no solo beneficiará al personal administrativo que maneje el sistema, sino a todos los solicitantes, reduciendo el tiempo de espera y facilidad para encontrar cualquier tipo de documento sin importar si se recuerda o no el nombre con el cual se guardó, y al ser realizado con sistemas inteligentes la aplicación irá mejorando a medida que se vaya usando, ya que tiene la principal ventaja de ir aprendiendo.

## **1.2. Formulación del problema**

El problema de la clasificación de documentos, esta tarea es laboriosa, encontrar un grupo de documentos y clasificarlo tomaría mucho tiempo ya que primero debes saber el tipo de documento, para esto primero se tendría que leer para saber su significado y luego asignarle una categoría. Este es un problema latente la cual genera tiempo perdido. Este problema afecta a cualquier organización que no cuenta con un sistema de clasificación automática de

documentos, lo cual produce incomodidad y demora a los encargados de esta labor. Analizando nuestro problema, nos nace la siguiente pregunta:

¿Cómo clasificar de manera automática los documentos en una oficina administrativa de la Universidad Nacional de Trujillo?

### **1.3. Hipótesis**

La aplicación de sistemas inteligentes permitirá la clasificación automática de documentos en una oficina administrativa de la Universidad Nacional de Trujillo.

### **1.4. Objetivos**

#### **1.4.1. Generales**

Desarrollar una aplicación de sistema inteligente que permita la clasificación automática de documentos.

#### **1.4.2. Específicos**

- Investigar sobre sistemas inteligentes para la clasificación automática.
- Analizar las características más destacadas de los sistemas inteligentes que hay en la actualidad.
- Implementar un prototipo para la clasificación automática de documentos en una oficina administrativa de la Universidad Nacional de Trujillo.
- Analizar los resultados obtenidos para comprobar la validez de la hipótesis planteada.

### **1.5. Estructura de la tesis**

Nuestro trabajo de investigación tiene los siguientes capítulos:

- En el primer capítulo se desarrolla los aspectos generales así como la justificación, formulación del problema, hipótesis de la investigación y objetivos del presente trabajo de investigación.
- En el segundo capítulo se realiza la recopilación de diferentes conceptos y definiciones que es de mucha importancia a lo largo de toda la investigación, como la metodología de investigación, inteligencia artificial, sistemas inteligentes, redes neuronales, perceptrón, perceptrón multicapa, backpropagation., también se define definiciones de clasificación automática de documentos, así como su representación vectorial y el centroide del conjunto de palabras, similitud aplicando métodos de clasificación automática de documentos, también describimos los tipos de documentos y por último como conceptos finales, consideramos a los gestores, plataformas y lenguajes que utilizaremos así como MySQL Workbench 8.0, NeatBeans IDE 8.2, Java y Joone.
- En el tercer capítulo se analiza, comprende, implementa y optimiza las soluciones de los sistemas inteligentes para la clasificación automática de documentos. Empezando por algunos problemas de algoritmos actuales. También la prueba a seguir de redes neuronales, se diseñan los modelos propuestos y se explica la implementación del sistema.
- En el cuarto capítulo se contrasta los resultados con el prototipo de la aplicación de sistemas inteligentes para la clasificación automática de documentos.
- En el quinto capítulo se presenta las conclusiones de nuestro trabajo de investigación, demostrando los objetivos cumplidos.

# Capítulo 2

## Materiales y Métodos

### 2.1. Marco teórico

El siguiente capítulo presenta conceptos importantes que utilizamos como parte fundamental para el desarrollo de nuestra tesis, este material bibliográfico fue investigado por la relación que tiene con los temas de nuestra investigación. La investigación se tornó muy amplia, principalmente en los conceptos claves para nuestro conocimiento científico, el cual ayudó a desarrollar la tesis, inteligencia artificial, sistemas inteligentes, redes neuronales, perceptrón, backpropagation, conceptos importantes ya que sin ellos sería imposible el modelamiento y desarrollo.

#### 2.1.1. Metodología Common Kads

La metodología elegida para el desarrollo del software es la metodología ágil *Common KADS*. Es una metodología diseñada para el análisis y la construcción de sistemas basados en conocimiento (SBC) de forma análoga a los métodos empleados en ingeniería de software. Fue propuesta y desarrollada por un grupo de investigadores pertenecientes a diversos países de la comunidad Europea, a través del programa ESPRIT para la innovación y la aplicación de tecnología informática avanzada. Lo que se pretendía era crear un estándar para la

ingeniería del conocimiento y sistemas de conocimiento con el cual se pudieran construir sistemas industriales de calidad a gran escala, en una forma estructurada y controlada. (Lagos, 2002).

### **2.1.1.1. Descripción**

En *CommonKADS* podemos ver reflejadas algunas ideas que han emergido, no solo de la experiencia en la ingeniería del conocimiento, sino también en el campo de la ingeniería del software en general. Estas ideas se pueden concretar en tres conceptos: modelado, reutilización y gestión del riesgo.

El modelado se refiere la configuración ideal que representa de manera simplificada una teoría, la cual ayuda a comprender las teorías y leyes.

La reutilización se centra en utilizar nuevamente productos ya creados anteriormente una o varias veces las cuales sean necesarias.

Gestión del riesgo se refiere al proceso de identificar, analizar y responder a factores de riesgo a lo largo de la vida de un proyecto, beneficiando los objetivos propuestos.

La metodología *CommonKADS* abarca todo el ciclo de desarrollo de software mediante un número de modelos interrelacionados que capturan los principales rasgos del sistema y de su entorno. El proceso de desarrollo de SBC consiste en llenar un conjunto de plantillas de los modelos. Asociados a las plantillas, *CommonKADS* define estados de los modelos que caracterizan hitos en el desarrollo de cada modelo. Estos estados permiten la gestión del proyecto, cuyo desarrollo se realiza de una forma cíclica dirigida por los riesgos. (Lagos, 2002).

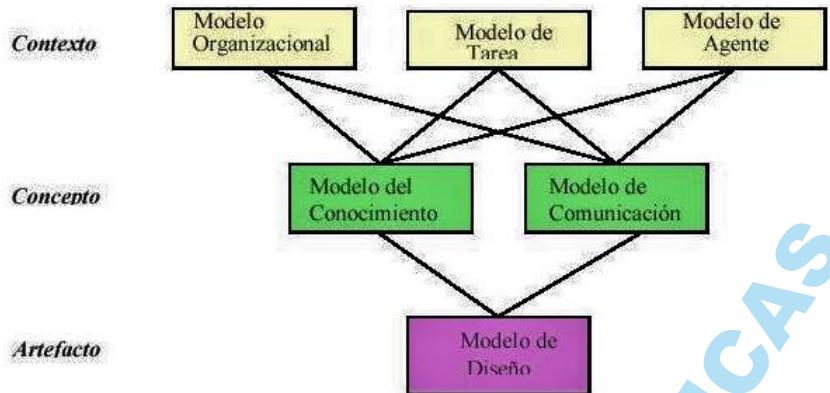


Figura 2.1: Modelo definidos por KADS.

Fuente: Lagos (2002).

La metodología *CommonKads* tiene tres etapas: el contexto, concepto y artefacto; en cada etapa contienen modelos, en la etapa de contexto tiene modelo, organizacional, modelo de tarea y modelo de agente; en concepto tiene modelo de conocimiento y modelo de comunicación; y en artefacto tiene modelo de diseño.

La relación que fija con UML son los siguientes diagramas: diagrama de clases, diagrama de estados, diagrama de actividades y diagrama de casos de uso. (Lagos, 2002).

### 2.1.2. Inteligencia Artificial (IA)

Empecemos esto con una pregunta clave: ¿Qué es la inteligencia artificial?. Aunque es un término de compleja definición, esta puede ser catalogada como una rama de la ciencia que se dedica al estudio de la forma en que se desarrolla el proceso del pensamiento humano para reproducirlo en un ente artificial. Es por ello que para poder lograrlo, la I.A. centra su estudio en dos áreas: el cuerpo humano (estudio de la inteligencia) y el ordenador electrónico. Teniendo en cuenta que para poder reproducir el pensamiento es necesario conocerlo, como primera meta para la investigación de los científicos fue entender los procesos cognoscitivos de la mente humana. De esta manera definir su modelo de inteligencia, para luego

proceder a programarlo en ordenadores, simulando cada uno de los procesos y comprobando los resultados de sus teorías. (Prince, 2009).

Según McCarthy (1989), acuñó la expresión “inteligencia artificial”, y la definió como: “la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes”.

### **2.1.3. Sistemas inteligentes**

Los sistemas inteligentes radican en un conjunto de herramientas y aplicaciones que agrupados llevan a cabo la recopilación, extracción y formato de información obtenida de distintas fuentes con el único propósito de crear medios inteligentes y artificiales para múltiples usos. Frecuentemente se usa para el soporte en la toma de decisiones, sin embargo, esta no es su única funcionalidad.

Son sistemas que, durante su existencia, aprenden y almacenan para luego actuar continuamente de forma interna o externa, de forma que pueda alcanzar su propósito superándose cada vez. Tiene su propio objetivo esencial y para alcanzarlo, escoge una acción tomada de las experiencias que ha almacenado en su memoria.

Para poder hablar de un sistema inteligente es necesario que exista un entorno con el cual el sistema pueda interactuar y, además, el sistema inteligente debe contener “sentidos” que le permitan recibir comunicaciones de dicho entorno y con ello poder transmitir información. El sistema actúa continuamente y cuenta con una memoria para archivar el resultado de sus acciones. Como se dijo anteriormente, este sistema tiene un objetivo, por lo cual, el conseguirlo significaría el hecho de seleccionar la respuesta adecuada ante cada estímulo. También se caracteriza porque a través de su memoria, durante su existencia, aprende de sus experiencias, logrando así, mejorar tanto su rendimiento como su eficiencia. Por último, es

necesario indicar que este sistema consume energía, la cual utiliza en sus procesos internos y su actuar. (Cruz, 2012).

#### 2.1.3.1. Capacidades requeridas

Según Cruz (2012), un sistema inteligente podrá ser considerado completo, si se incluye dentro de estas diversas funcionalidades, tales como :

- **Inteligencia:** Es la capacidad de alcanzar nuestros objetivos. La inteligencia incluye la capacidad de aprender a lograrlo.
- **Conceptualización:** Un concepto es el elemento primordial del pensamiento. Es el almacenamiento físico, material de información (en neuronas o electrones). Todos los conceptos de la memoria están interrelacionados en red. La capacidad de poder conceptualizar implica el desarrollo de distintos niveles de abstracción.
- **Reglas de actuación:** Una regla de actuación es la consecuencia de una experiencia o el resultado de interpretar la propia memoria, es decir, relaciona situaciones y efectos de la acción.
- **Memoria:** La memoria es un almacenamiento físico de conceptos y reglas de actuación, que incluye la experiencia del sistema.
- **Aprendizaje:** Posiblemente, es la capacidad más importante de un sistema inteligente, ya que este aprende conceptos a partir de la información obtenida a través de los sentidos. Aprende también reglas de actuación a base de su experiencia, la cual en ocasiones, la actuación se almacena con su valor. Una regla de actuación aumenta en valor si permitió el logro de un objetivo. El aprendizaje incluye la inserción de conceptos

abstractos, a base de ejemplos concretos y la creación de conceptos compuestos que contienen los conceptos de partes de un objeto. Entonces el aprendizaje también se define como la capacidad de detectar relaciones (patrones) entre la parte (“situación la parte”, “situación futura”) de una regla de actuación.

#### **2.1.4. Red Neuronal Artificial (RNA)**

Una red neuronal es un sistema de procesadores paralelos conectados entre sí en forma de grafo dirigido. Esquemáticamente cada elemento de procesamiento (neuronas) de la red se representa como un nodo. Estas conexiones establecen una estructura jerárquica que tratando de emular la fisiología del cerebro busca nuevos modelos de procesamiento para solucionar problemas concretos del mundo real. Lo importante en el desarrollo de la técnica de las RNA es su útil comportamiento al aprender, reconocer y aplicar relaciones entre objetos y tramas de objetos propios del mundo real. (Freeman and Skapura, 1991)

Las ventajas de la red neuronal son:

1. Aprendizaje adaptativo: Son adaptables debido a la capacidad de autoajuste de los elementos procesales (neuronas) que componen el sistema.
2. Auto organización: La auto organización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico. Cuando las redes neuronales se usan para reconocer ciertas clases de patrones, ellas auto organizan la información usada.
3. Tolerancia a fallos: Las redes neuronales pueden aprender a reconocer patrones con ruido, distorsionado o incompletos. Esta es una tolerancia a fallos respecto a los datos. Las redes pueden seguir realizando su función (con cierta degradación) aunque se destruya parte de la red.

4. Operación en tiempo real: Las redes neuronales se adaptan bien a esto debido a su implementación paralela.
5. Fácil inserción dentro de la tecnología existente: Con las herramientas computacionales existentes (no del tipo computadora), una red puede ser rápidamente entrenada, comprobada, verificada y trasladada a una implementación hardware de bajo coste.

### **2.1.5. Tipos de aprendizaje**

Hay dos métodos de aprendizaje importantes, que son: aprendizaje supervisado y no supervisado, la diferencia entre ambos tipos depende en la existencia o no de un agente externo que controle todo el proceso. (Matich, 2001):

#### **2.1.5.1. Aprendizaje supervisado**

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. En este tipo de aprendizaje se suelen considerar, a su vez, tres formas de llevarlo a cabo, que dan lugar a los siguientes aprendizajes supervisados: Aprendizaje por corrección de error, aprendizaje por refuerzo y aprendizaje estocástico. (Matich, 2001).

### **2.1.6. Elementos básicos**

A continuación se puede ver, en la Figura 2.2, un esquema de una red neuronal:

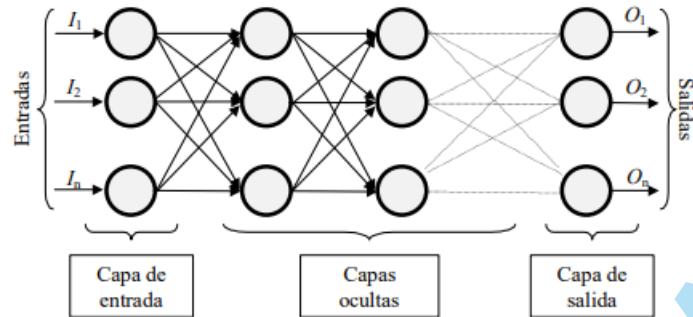


Figura 2.2: Red neuronal artificial.

Fuente: Erlinb (2017).

Esta red neuronal está constituida por neuronas interconectadas y arregladas en tres capas. Los datos ingresan por medio de la “capa de entrada”, pasan a través de la “capa oculta” y salen por la “capa de salida”. Cabe mencionar que la capa oculta puede estar constituida por varias capas.

#### 2.1.6.1. Función de entrada (input function)

La neurona trata a muchos valores de entrada como si fueran uno solo; esto recibe el nombre de entrada global. Por lo tanto, ahora nos enfrentamos al problema de cómo se pueden combinar estas simples entradas ( $in_{i1}, in_{i2}, \dots$ ) dentro de la entrada global,  $gin_i$ . Esto se logra a través de la función de entrada, la cual se calcula a partir del vector entrada. La función de entrada puede describirse como sigue:

$$input_i = (in_{i1} \bullet w_{i1}) * (in_{i2} \bullet w_{i2} * \dots * (in_{in} \bullet w_{in})) \quad (2.1)$$

donde:  $*$  representa al operador apropiado (por ejemplo: máximo, sumatoria, productoria, etc.),  $n$  al número de entradas a la neurona  $n_i$  y  $w_i$  al peso. (Matich, 2001).

### 2.1.6.2. Función de activación (activation function)

Una neurona biológica puede estar activa (excitada) o inactiva (no excitada); es decir, que tiene un “estado de activación”. Las neuronas artificiales también tienen diferentes estados de activación; algunas de ellas solamente dos, al igual que las biológicas, pero otras pueden tomar cualquier valor dentro de un conjunto determinado. (Matich, 2001).

La función activación calcula el estado de actividad de una neurona; transformando la entrada global (menos el umbral,  $\sigma_i$ ) en un valor (estado) de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1). Esto es así, porque una neurona puede estar totalmente inactiva (0 o -1) o activa (1). La función activación, es una función de la entrada global ( $gin_i$ ) menos el umbral ( $\sigma_i$ ). Las funciones de activación más comúnmente utilizadas se detallan a continuación:

- Función Lineal

$$f(x) = \begin{cases} -1 & x \leq -1/a \\ a * x & -1/a < x < 1/a \\ 1 & x \geq 1/a \end{cases} \quad (2.2)$$

con  $x = gin_i - \sigma$ , y  $a > 0$ .

Los valores de salida obtenidos por medio de esta función de activación serán:  $a \cdot (gin_i - \sigma_i)$ , cuando el argumento de  $(gin_i - \sigma_i)$  esté comprendido dentro del rango  $(-1/a, 1/a)$ . Por encima o por debajo de esta zona se fija la salida en 1 o -1, respectivamente. Cuando  $a = 1$  (siendo que la misma afecta la pendiente de la gráfica), la salida es igual a la entrada. (Matich, 2001).

- Función Sigmoidea

$$f(x) = \frac{1}{1 + e^{-gx}}, \text{ con } x = gin_i - \sigma, \quad (2.3)$$

Los valores de salida que proporciona esta función están comprendidos dentro de un rango que va de 0 a 1. Al modificar el valor de  $g$  se ve afectada la pendiente de la función de activación.

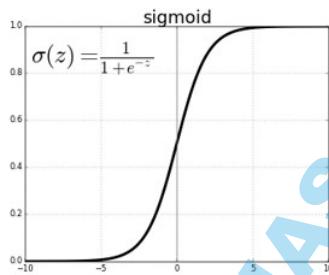


Figura 2.3: Gráfico de la función sigmoide.

Fuente: Elaboración propia.

- Función Tangente hiperbólica

$$f(x) = \frac{e^{gx} - e^{-gx}}{e^{gx} + e^{-gx}}, \text{ con } x = gin_i - \sigma, \quad (2.4)$$

Los valores de salida de la función tangente hiperbólica están comprendidos dentro de un rango que va de -1 a 1. Al modificar el valor de  $g$  se ve afectada la pendiente de la función de activación.

#### 2.1.6.3. Función de salida (output function)

El último componente que una neurona necesita es la función de salida. El valor resultante de esta función es la salida de la neurona  $i$  ( $out_i$ ); por ende, la función de salida determina

que valor se transfiere a las neuronas vinculadas. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona subsiguiente. Normalmente, no cualquier valor es permitido como una entrada para una neurona, por lo tanto, los valores de salida están comprendidos en el rango  $[0, 1]$  o  $[-1, 1]$ . También pueden ser binarios 0, 1 o  $-1, 1$ . (Matich, 2001).

Dos de las funciones de salida más comunes son cuando: la salida es la misma que la entrada (función identidad), o binaria, que devuelve uno cuando el  $act_i$  es mayor o igual que el umbral  $\epsilon_i$ , caso contrario es cero.

## 2.1.7. Perceptrón

El perceptrón lee los valores de entrada, después suma todas las entradas teniendo en cuenta los pesos y por último el resultado lo introduce en una función de activación que nos genera el resultado final. (Espitia Corredor, Daniel Eduardo, 2017).

### 2.1.7.1. Perceptrón simple

Se determina los pesos sinápticos y el umbral que proporcione el óptimo ajuste de la entrada con la salida, estas variables, para determinar estas variables se sigue un proceso adaptativo, el cual comienza con valores aleatorios y se van modificando según la diferencia entre los valores deseados y los calculados por la red.

Recordar que el perceptrón sólo es capaz de representar funciones lineales, dado que no dispone de capas ocultas, para esto existen los perceptrones multicapa.

Tabla 2.1: Componentes del perceptrón.

Componentes del perceptrón	Definición
Entradas	Es la información que recibe el perceptrón.
Pesos	Son los valores numéricos que se encargan de establecer la influencia de una entrada en la salida deseada.
Bias	Es un parámetro que tienen algunos modelos de redes neuronales el cual permite encontrar fácilmente la separación entre posibilidades de salida de una red neuronal.
Función de activación	Se encarga de determinar un valor de salida una vez se han procesado cada una de las entradas.

Fuente: Elaboración propia según Espitia Corredor, Daniel Eduardo (2017).

En la Tabla 2.1 se representan los componentes del perceptrón simple.

El entrenamiento del perceptrón es un proceso iterativo y sigue los siguientes pasos, hasta lograr reducir el error:

- Paso 1: Inicializar los pesos y el bias.
- Paso 2: Calcular las salidas (net) con los pesos y el bias.
- Paso 3: Obtener la salida utilizando la función de activación y calcular cada valor del error.
- Paso 4: Corregir el Bias y los pesos.

### 2.1.7.2. Perceptrón Multicapa

El Perceptrón multicapa es una red de alimentación hacia delante compuesta por una capa de N neuronas de entrada (sensores), otra capa formada por M neuronas de salida y un número determinado de capas ocultas. (Ver Figura 2.4). El tamaño de éstas dependerán de la dificultad de la correspondencia a implementar. (Munt, 2018).

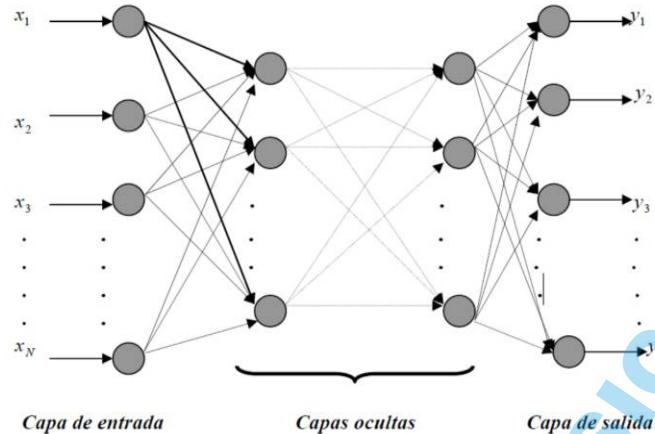


Figura 2.4: Representación de un red multicapa.

Fuente: Alba Morera (2018).

El objetivo que se busca con este tipo de red es el mismo, establecer una correspondencia entre un conjunto de entrada y un conjunto de salidas deseadas, de manera que:

$$(x_1, \dots, x_n) \in \mathbb{R}^N \rightarrow (y_1, \dots, y_n) \in \mathbb{R}^M \quad (2.5)$$

Para ello se dispone de un conjunto de p pares de entrenamiento de manera que sabemos perfectamente que al patrón de entrada  $(x_1^k, \dots, x_N^k)$  le corresponde la salida  $(y_1^k, \dots, y_M^k)$ ,  $k = 1, \dots, p$ . Así, nuestro conjunto de entrenamiento es:

$$(x_1^k, \dots, x_N^k) \rightarrow (y_1^k, \dots, y_M^k), k=1, \dots, p \quad (2.6)$$

### 2.1.8. Backpropagation

La red neuronal propaga la señal de los datos de entrada hacia adelante a través de sus parámetros hacia el momento de la decisión, y luego propaga hacia atrás la información sobre el error, a la inversa a través de la red, para que pueda alterar los parámetros. Esto sucede

paso a paso:

- La red adivina los datos, utilizando sus parámetros.
- La red se mide con una función de pérdida.
- El error se propaga hacia atrás para ajustar los parámetros equivocados.

La propagación hacia atrás de errores o retropropagación (del inglés backpropagation) es un método de cálculo del gradiente utilizado en algoritmos de aprendizaje supervisado utilizados para entrenar redes neuronales artificiales. El método emplea un ciclo propagación (adaptación de dos fases). Una vez que se ha aplicado un patrón a la entrada de la red como estímulo, este se propaga desde la primera capa a través de las capas siguientes de la red, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas.

Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo, las neuronas de la capa oculta sólo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del espacio total de entrada. Después del entrenamiento, cuando se les presente un patrón arbitrario de entrada que contenga ruido o que esté incompleto, las neuronas de la capa oculta de la red responderán con una salida activa si

la nueva entrada contiene un patrón que se asemeje a aquella característica que las neuronas individuales hayan aprendido a reconocer durante su entrenamiento.

$$\delta^L = \nabla_a C \odot \sigma'(z^L) \quad (\text{BP1})$$

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l) \quad (\text{BP2})$$

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l \quad (\text{BP3})$$

$$\frac{\partial C}{\partial w_{j,k}^l} = a_k^{l-1} \delta_j^l \quad (\text{BP4})$$

Figura 2.5: Ecuaciones del backpropagation.

Fuente: Michael A. Nielsen (2015).

Donde: BP1 es para el error en la capa de salida, BP2 es para el error en términos del error en la siguiente capa, BP3 es para la tasa de cambio del costo con respecto a cualquier sesgo en la red y BP4 es para la tasa de cambio del costo con respecto a cualquier peso en la red.

### 2.1.9. Clasificación automática de documentos

Es una acción que realizamos de manera habitual en nuestro día a día, muchas veces este comportamiento se demuestra de manera espontánea, pues desde que se inicia el día se empieza a clasificar objetos, todo ello para una rápida ubicación y fácil acceso. En la clasificación de documentos hay pasos a seguir comunes que se encuentran en la mayoría de documentos al momento de clasificar puesto que requiere de un análisis, selección, etc. de los documentos o lo contenido de ellos. Esta clasificación llevada a un ámbito laboral, llámese empresa o alguna institución educativa, se vuelve el problema más común para ellos, teniendo como relación que mientras mayores niveles de producción, mayor se hace la base de datos y por ende mayor el tiempo que se necesita para encontrar algún documento. Es

por ello que se necesitan sistemas más sofisticados para la clasificación de documentos.

En la actualidad existen técnicas con las cuales podemos satisfacer estas necesidades, la informática de la mano con la inteligencia artificial son las herramientas del futuro.

#### **2.1.9.1. Métodos de clasificación automática de documentos**

Los métodos para la clasificación automática de documento son:

- Análisis de la estructura gráfica del documento.
- Análisis de la estructura gráfica del documento junto con palabras claves.
- Análisis y procesamiento de texto.

#### **2.1.9.2. Representación vectorial**

En la rama de recuperación de la información, los investigadores optaron por la representación vectorial para el manejo de documentos, así a cada documento se le asigna un vector de en espacio de términos (grupo de palabras que se repiten en los documentos de la colección). (Steinbach et al., 2000).

Se puede representar a cada documento por su vector de frecuencia de términos:

$d_{ft} = (ft_1, ft_2, \dots, ft_m)$ , donde  $ft_i$  es la frecuencia de término  $i$  en el documento. (Steinbach et al., 2000).

Para refinar lo anterior mencionado, el resultado es el producto de la frecuencia de término y la frecuencia de documento inversa. Ésta técnica, denominada “ $ft - fdi$ ”, permite que si un término aparece en la mayor parte de los documentos, no se le toma mucha importancia. Y su representación sería  $d_{ft-fdi} = (ft_1 \log(N/df_1), ft_2 \log(N/df_2), \dots, ft_m \log(N/df_m))$ , donde

$N$  es la cantidad total de documentos de la colección y  $df_i$  es la cantidad de documentos que contienen al término  $i$ . (Faloutsos and Oard, 1998).

#### 2.1.9.3. Centroide

Se tiene un grupo de elementos  $S$ , que tiene  $h$  elementos  $S_i$ , se define su centroide  $C_s$  como el promedio de los vectores que componen el grupo:  $C_s = \frac{\sum_{i=1}^h S_i}{h}$ . La propiedad más importante es que su distancia promedio de cualquier punto hasta cada elemento del grupo es equivalente a la distancia entre ese punto y el centroide del grupo.(Steinbach et al., 2000).

#### 2.1.9.4. Similitud

Cada grupo de documentos tienen semejanzas entre sí, para poder evaluar lo mencionado es necesario contar con una medida cuantitativa de la similitud entre documentos. (Steinbach et al., 2000).

Una representación más sencilla, es en el que los atributos del vector esta en binario, definiendo  $|v|$  como la cantidad de atributos de  $v$ :

-Coeficiente de Jaccard

$$\left| \frac{d_1 \cap d_2}{d_1 \cup d_2} \right| \quad (2.7)$$

-Coeficiente de coseno

$$\frac{|d_1 \cap d_2|}{\sqrt{|d_1|} * \sqrt{|d_2|}} \quad (2.8)$$

Los coeficientes anteriores logran definir la semejanza de los documentos y la cantidad de términos que tienen en común con la relación de tamaño de los documentos.

Existe un modelo de presentación la cual es muy utilizada, en donde se calculan los vectores

de frecuencia o de frecuencia inversa, y  $||v||$  es la longitud (norma) del vector  $v$ , la medidas más comunes son:

-coeficiente del coseno extendido

$$\cos(d_1, d_2) = \frac{d_1 \cdot d_2}{||d_1|| * ||d_2||} \quad (2.9)$$

La ecuación tiene la propiedad de independencia del tamaño de los documentos, puesto que  $\cos(d_1, d_2) = \cos(a \cdot d_1, d_2)$  para  $a > 0$ . Por lo que, se normalizan los documentos para que sean de longitud unitaria, entonces,  $\cos(d_1, d_2) = d_1 \cdot d_2$ .

### 2.1.10. Codificación de los datos de entrada.

Los datos tienen que ser codificados, o sea, deben hallarse valores apropiados para representar las características simbólicas (alto, bajo, adecuado, etc.). (Matich, 2001).

Se distinguen dos tipo de variables a ser codificadas:

- Variables o atributos numéricos (frecuentemente llamadas continuas).
- Variables o atributos simbólicos (frecuentemente llamados discretos).

Un atributo numérico es aquel que puede tomar cualquier valor dentro de un cierto intervalo  $[a, b]$ ; donde  $a$  puede ser  $-\infty$  (menos infinito) y  $b, \infty$  (infinito). Por ejemplo el peso puede medirse en libras; entonces cualquier valor entre  $[0, \infty)$  está permitido. Ahora si los pesos son dados por un cierto número de términos, semejantes a: alto o bajo; entonces el atributo se denomina simbólico. Por lo tanto, dividiendo el intervalo  $[a, b]$  de una variable numérica dentro de subintervalos, podemos confeccionar un atributo continuo pseudodiscreto.

## 2.1.11. Descripción documentos en la investigación

Definición de los documentos que utilizaremos como elementos a clasificar.

**Oficio:** es el tipo de documento más formal y el de mayor uso por la Universidad Nacional de Trujillo, el cual se encarga de comunicar y llevar a cabo gestiones vinculadas con acuerdos, invitaciones y/o solicitar algo.

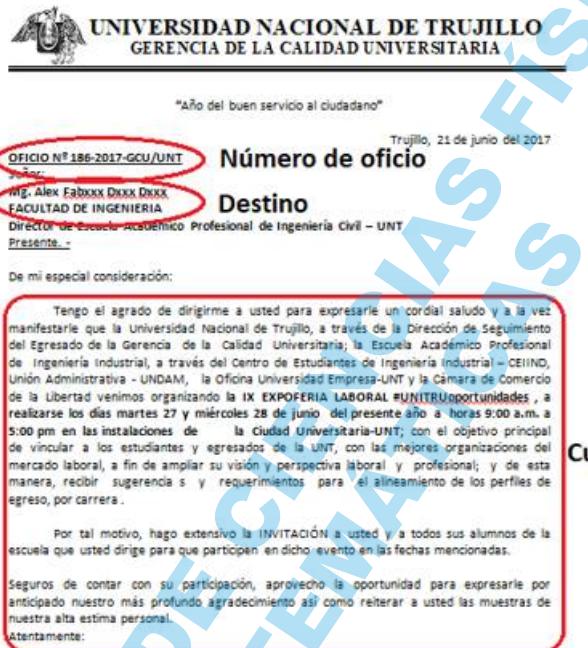


Figura 2.6: Partes del oficio.

Fuente: Elaboración propia.

**Oficio múltiple:** este tipo de documento se usa cuando un mismo tema o asunto va dirigido a más de un destino, a escuelas, direcciones o unidades dentro y fuera de la Universidad Nacional de Trujillo. Un punto a tener en cuenta en este tipo de oficio es que el número de oficio será el mismo para todos los ejemplares.



"Año del buen servicio al ciudadano"

Trujillo, 15 de Marzo de 2017

OFICIO MÚLTIPLE N° 034-2017-GCU/UNT

Señora:

KELXXX BAIXXX

EMPRESA - CARTAVIO S.A.A.

### Tipo y número de oficio

### Destino

Presento:-

De mi especial consideración:

Siendo el agrado de dirigirme a usted para expresarle un cordial saludo y a la vez manifestarle que la Universidad Nacional de Trujillo, a través de la Dirección de Seguimiento del Egresado de la Gerencia de la Calidad Universitaria; la Escuela Académico Profesional de Ingeniería Industrial, a través del Centro de Estudiantes de Ingeniería Industrial – CEINO, la Oficina Universidad-Empresa-UNT y la Cámara de Comercio de La Libertad venimos organizando la IX EXPOFERIA LABORAL #UNITRUopportunidades, a realizarse los días jueves 11 y viernes 12 de Mayo del presente año a horas 9:00 am. a 5:00pm en las instalaciones de la Ciudad Universitaria-UNT; con el objetivo principal de vincular a los estudiantes y egresados de la UNT, con las mejores organizaciones del mercado laboral, a fin de ampliar su visión y perspectiva laboral y profesional; y de esta manera, recibir sugerencias y requerimientos para el alineamiento de los perfiles de egreso, por carrera.

Por tal motivo, hago extensivo la INVITACIÓN a vuestra organización a participar de la "IX EXPOFERIA LABORAL 2017 # UNITRUopportunidades", y para confirmar su asistencia y demás coordinaciones ponemos a su disposición el teléfono: 044-220317(oicina de Gerencia de Calidad Universitaria, con la Sra. Thalia Meza) en el horario de 8:00 am - 2:00 pm, 4945696407 (Sr. Alexander Pérez Villanueva) y al correo electrónico [unitruopportunidades@unitru.edu.pe](mailto:unitruopportunidades@unitru.edu.pe). Asimismo podrá inscribirse a través del formulario que le anotamos a su correo.

Seguros de contar con su participación, aprovecho la oportunidad para expresarle por anticipado nuestro más profundo agradecimiento así como reiterar a usted las muestras de nuestra alta estima personal.

Atentamente:

DEA. AMIXKA CUTIXKA KXAK  
GERENTE DE CALIDAD UNIVERSITARIA

### Remitente

Figura 2.7: Partes del oficio múltiple.

Fuente: Elaboración propia.

**Informe:** definimos informe como el documento que tiene por función informar algún acontecimiento en la Universidad Nacional de Trujillo, la cual es la consecuencia de algún hecho ocurrido ocurrido.

*"Año de la Consolidación del mar de Gran"*

INFORME N° 289 - 2016-CPD/DP

De:	Dra. <b>Amparo Gómez-Basa</b> , Gerente de Calidad Universitaria		
A:	<b>Dra. Chávez, F. Edilma</b> , Directora de Planeamiento y Desarrollo		
Asunto:	Solicitud revisión del Ing. Chávez Ulises Wilson Tlalaño.		
Ref:	Expediente N° 301316053 E. Régimen N° 324016053.		
Fecha:	Tacna, 18 de Octubre del 2016		
Ran:			
<p>Es pleno respeto, a su designado en Asunto al Expte. N° 301316053 E. Regimen N° 324016053 Oficio N°034-2016-DPC-FAC-ING. Señalando el Diccionario de la Facultad de Ingeniería, De Humanas, Universidad Nacional Mayor de San Marcos, dice que es el momento del Ing. Segundo Reunión Años Largo, y en su reemplazo continúa al Ing. Edmundo Chávez Ulises, en la categoría de Profesor Asociado jefe de Postura T. C., 40 horas para el año académico 2016-II, sucesivo, para el Departamento Académico de Ingeniería de Sistemas de la Sección Jaénquicopas, al mencionado docente nacido en la Ciudad de Guadalajara, le rinde agradecimiento el normal desarrollo de las actividades lectivas y la permanencia del docente. Siguen documentos adjuntos:</p>			
FACULTAD	NOMBRE	Nº PLAZA	A.
Facultad de Ingeniería – Departamento Académico de Ingeniería de Sistemas.	Ing. Chávez Ulises Wilson Edilmerio (del 22-08-16 al 16-12-2016)	783	Jef. de Postura T.C. 40 hrs

Al respecto esta Dirección informa:

Que según Oficio N°070-2016-DPTO-SISTE-FAC-ING. Del departamento Académico de Ingeniería de Sistemas, firmado por el Director de Departamento, advirtiendo que anuncia su término para el año académico 2016-II sucesivo, hasta que se celebre por convocatoria pública la plena correspondencia.

Que según Informe N° 403-2016-Dec.Fac. Ing. De la Facultad de Ingeniería, indica el término y su sucesión del Ing. Chávez Ulises Wilson Tlalaño, quien asumirá la carga lectiva en tanto se precise el sorteo para publicar su sucesión dentro plazo.

Que según Informe N° 120-2016-CHPER/Acad. De la Gerencia de Recursos Humanos, considerando el acuerdo devolvencia pendiente el económico destino preparado, con cargo al financiamiento de la plaza N° 773 de Profesor Titular TC, debiendo regularmente la reforma de dicha medida remunerativa. Peticionar a la Universidad, pendiente en atención de las observaciones del Organismo de Control Interno. Desde el 22 de agosto al 16 de diciembre del 2016 a hasta que se celebre por convocatoria pública.

Para la atención de lo señalado debe de recordar que la competencia leyes de la Oficina de Asuntos Jurídicos y donde el punto de vista de la renombrada legislación, se deberá tener en cuenta lo establecido en la Ley N° 30372, Ley de Presupuesto del Sector Público para el año fiscal 2016, Artículo 8, inciso 8.1. literal d) "La contratación para el cumplimiento de estos servicios, así como la prestación del personal o para la ejecución temporal de los servicios del sector público, en tanto se implemente la Ley 30053, Ley del Servicio Civil, en los casos que corresponda. En el caso de los empleados permanentes del personal, que corresponde al caso que se establece precedido a partir del año 2013, debiéndole tener en cuenta que al ingresar a la administración pública se tienen necesariamente por convenio público de monto y sujetos a los desembolsos de costos respectivos. En el caso del asenso o posesión del personal las entidades deben tener en cuenta previamente a la ejecución de dicha suerte de personal, la establecida en el literal b) de la norma dispuesta anteriormente en la Ley 28431, Ley General del Sistema Nacional de Presupuesto. En el caso de suplemento de personal, para ver finalizada la labor para la cual fue nombrada la persona, los asunciones sujetas en quedan establecidas anteriormente".

Por lo que se acuerda por Resolución Rectoral: El presente puesto será ejercido de acuerdo a lo establecido en la legislación de personal, dentro de los límites establecidos en la legislación de la UNMSM.

**FUENTE DE FINANCIAMIENTO:** Recursos Ordinarios.

**CÁDENA FUNCIONAL:** PROFESIONAL TITULAR.

**0046 3400001 5881353 22 048 8109 4974136** - Docente de la educación universitaria de postgrado. Oficial de Ingeniería y Departamento Académico de Ingeniería de Sistemas.

**CLASIFICACIÓN DEL GASTO:**

2.1.1.1.1.2. Personal remunerado

En cuento informar a Unidad.

## Remitente Destino

*BIBLIOTECA UNMSM*

## Cuerpo

### Remitente

Figura 2.8: Partes del informe.

Fuente: Elaboración propia.

**Solicitud:** es un documento en el cuál se pide el goce de un derecho, beneficio o servicio que se considera que puede ser atendido por las direcciones, unidades y/o rector dentro de la Universidad Nacional de Trujillo . El derecho a petición en caso de Perú está amparado por el Art.2º inciso.20.



UNIVERSIDAD NACIONAL DE TRUJILLO  
DIRECCIÓN DE PLANEACIÓN Y DESARROLLO  
DIRECCIÓN DE PRESUPUESTO

*"Año de la Consolidación del mar de Grau"*

SOLICITUD N° 348 - 2014-GPD-DP Número de Solicitud

Econ. Albaix Luis Caxx Dirección de Presupuesto (e)	Remitente
Mr. Aranguiz Fierros Rodríguez Axx Dirección de Planeación y Desarrollo	Destino
Asunto : Solicitud pago por curso de nivelación 2014-III al docente Darío Emiliano Medina Centro de la Sede de Santiago de Chuco.	Asunto
Ref. : Expediente N° 129716539 E Registro N° 152316339	
Fecha : Trujillo, 16 de agosto de 2016	Registro N°
<p>Este gasto diligenciamos en su despacho en relación al Expediente N° 129716539 E y Registro N° 152316339, Oficio N° 081-2014-UNT-a-5CH-D. Con la finalidad de atender lo solicitado por el Director General de la Sede de Santiago de Chuco Emírito Felipe Zavala Benites, solicitando el pago por curso de nivelación. A favor del Docente Darío Emiliano Medina Centro, a cargo del curso "Fitopatología General" dictado a los estudiantes del VI Ciclo de la Escuela de Agronomía en el ciclo de Nivelación 2014-III. Según documento adjunto.</p> <p>Al respecto esta Dirección informa:</p> <p>Lo solicitado se atenderá previo informe favorable de la Gerencia de Recursos Humanos - Oficina de Personal Docente y lo solicitado por el Director General de la Sede Santiago de Chuco Emírito Felipe Zavala Benites, el pago a favor del Docente Darío Emiliano Medina Centro, a cargo del curso "Fitopatología General" dictado a los estudiantes del VI Ciclo de la Escuela de Agronomía en el ciclo de Nivelación 2014-III, se atenderá con el presupuesto asignado a la Sede de Santiago de Chuco, por lo que de contar con disponibilidad presupuestal y de ser autorizado, mediante Resolución Rectoral, el presente gasto será ejecutado de acuerdo a la clasificación funcional programática, fuente de financiamiento y específica del gasto 2016.</p>	

FUENTE DE FINANCIAMIENTO: R.D.R.

CADENA FUNCIONAL PROGRAMÁTICA 003  
0066 3000001 5001363 22 048 0109 0070136 Desarrollo de la Educación Universitaria de Postgrado. (Sede de Santiago de Chuco).

CLASIFICACIÓN DEL GASTO: Según sea el caso:  
2.3.2.7.3.2 Realizado por personas naturales  
2.1.1.9.3.4 Asignación por encuesta

Es cuando informo a Usted.

Econ. Albaix Luis Caxx Caxx  
Dirección de Presupuesto (e) **Firma**

**Cuerpo**

Figura 2.9: Solicitud.

Fuente: Elaboración propia.

**Resolución rectoral:** La resolución rectoral dentro de la Universidad Nacional de Trujillo, son documentos muy importantes ya que son emitidas por el rector universitario, quien es la máxima autoridad en la Universidad Nacional de Trujillo.

RESOLUCIÓN RECTORAL N° 260 -2014- Número de resolución

Trujillo, 10 de Setiembre del 2014

Viaje el expediente N° 301, con 55 folios presentado por el Oficina de la Oficina de Planificación y Desarrollo de la Universidad Nacional de Trujillo, sobre aprobación de MODIFICACIONES PRESUPUESTARIAS en el Presupuesto de la Universidad Nacional de Trujillo para el Año Fiscal 2014.

CONSIDERANDO:

Que, mediante Ley N° 28411, se aprobó el Presupuesto del Sector Público para el Año Fiscal 2014, en el que está comprendida la Universidad Nacional de Trujillo;

Que la presente Resolución se sustenta en las "Notas para Modificaciones Presupuestarias", emitidas por la Unidad Ejecutora durante el periodo mensual de Agosto del 2014;

De conformidad con lo establecido en la Directiva N° 005-2010-EF-1601, modificada mediante Resolución Directoral N° 022-2011-EF-00-01, "Directiva para la Ejecución Presupuestaria", en el Capítulo IV - Disposiciones Especiales para la Ejecución Presupuestaria del Gobierno Nacional, Gobiernos Regionales y Gobiernos Locales, Artículo 20°, Modificaciones Presupuestarias en el Nivel Funcional Programático, numeral 20.1, Literal b) ... se formaliza por Resolución del Titular del Pliego... deberá permitir asegurar el cumplimiento de los Álamos Presupuestarios, de acuerdo a la oportunidad de su ejecución, según la priorización de gastos aprobada por el Titular del Pliego presupuestario.

Que para efecto de las Modificaciones Presupuestarias en el Nivel Funcional Programático, los Pliegos deben tomar en cuenta lo establecido en el artículo 40° de la Ley General del Sistema Nacional de Presupuesto - Ley N° 28411.

Estando a lo solicitado por la Unidad Ejecutora y la Oficina de Planificación y Desarrollo, y de conformidad con lo previsto en el artículo 47 de la Ley General del Sistema Nacional de Presupuesto, Ley N° 28411, y en uso de las atribuciones conferidas al Sr. Rector en el Artículo 17º numeral c) del Estatuto de la Universidad Nacional de Trujillo.

SE RESUELVE:

Artículo 1° Formalizar las modificaciones presupuestarias efectuadas en el Nivel Funcional Programático dentro de Unidades Ejecutoras, conforme al Anexo que se adjunta a la presente Resolución, en el marco de lo dispuesto en el Artículo 40° de la Ley N° 28411, Ley General del Sistema Nacional de Presupuesto.

Artículo 2° La presente Resolución se sustenta en las "Notas para Modificaciones Presupuestarias" emitidas por la Unidad Ejecutora durante el mes de Agosto del Año Fiscal 2014.

Artículo 3° Copia de la presente Resolución se presenta, dentro de los cinco (05) días de aprobada, a la Dirección Nacional del Presupuesto Público del Ministerio de Economía y Finanzas.

REGISTRESE, COMUNIQUESE Y ARCHÍVENSE

Cuerpo

Resolución

Remitente

Distribución

DR. J. ANXXXX ALXXXX UXXX DXXX  
PROFESOR SECRETARIO GENERAL (s)  
DR. O. ORXXXX VELXXX BENXXX  
RECTOR

Quintuplicado,  
Vicerrectoría Académica  
Asistente Interna  
Oficina Técnica de Contabilidad  
Directorio Nacional del Presupuesto Público  
Archivo (s)

Vicerrectoría Administrativa  
Oficina de Recursos  
División de Presupuesto

Oficina de Planificación y Desarrollo  
Contaduría Pública de la Nación

Figura 2.10: Resolución rectoral.

Fuente: Elaboración propia.

**Memorandum:** es un informe donde se expone algo que se debe tener en cuenta para un determinado asunto o acción, en la Universidad Nacional de Trujillo se puede usar por ejemplo para cambiar a un trabajador de oficina o para sancionar.

MEMORANDUM\_Nº 003-2016-VRA Encabezado y número

PARA :	Lic. Manxxx Sáxxx Sanxxx Unidad De Logística	<b>Destino</b>
CPC. Eugxxx Sxxx Pxox Lxxx	Jefe de la Unidad de Presupuesto	
Dra. Ampxxx Gutxxx Rroox	Gerente de Calidad Universitaria	
DE :	Sra. Flor Manxxx Lxxx Vioxxx Mrox	<b>Remitente</b>
	Vice Rectorado Administrativo - UNT	
ASUNTO :	Certificación de Crédito Presupuestario Anualizada N° 00000005 apertura del Fondo Fijo Para Caja Chica de la DRSTCU; por el importe de S/. 72,000.00	
REF. :	OFICIO N° 012-2011-GRU-DRSTC-DR-OPP.	
FECHA :	Trujillo, 11 de marzo del 2016.	

---

Para su conocimiento y fines correspondientes remito a usted el Oficio de la referencia mediante el cual el Director de la Oficina de Planificación y Presupuesto da a conocer la Certificación de Crédito Presupuestario Anualizado N° 00000005 Apertura del Fondo Fijo para Caja Chica de la DRSTCU, aprobada por R.D.R.S. N° 001-2011-GRU-DRSTC-DR; distribuido por metas con un Presupuesto Anual de S/. 72,000.00.

El Director de la Oficina de Planificación y Presupuesto debe corregir el error material en que ha incurrido al haber omitido la Fuente de Financiamiento 2. Rubro 09. Recursos Directamente Recaudados.

**Firma**

Atentamente,

Figura 2.11: Partes del memorandum.

Fuente: Elaboración propia.

### 2.1.12. Mysql Worbench

MySQL Workbench es un software creado por la empresa Sun Microsystems, esta herramienta permite modelar diagramas de Entidad-Relación para bases de datos MySQL. Con esta herramienta se puede elaborar una representación visual de las tablas, vistas, procedimientos almacenados y claves foráneas de la base de datos. Además, es capaz de sincronizar el modelo en desarrollo con la base de datos real. Se puede realizar una ingeniería directa e ingeniería inversa para exportar e importar el esquema de una base de datos ya existente el cual haya sido guardado o hecho copia de seguridad con MySQL Administrador. MySQL Workbench puede generar también el guión necesario para crear la base de datos que se ha dibujado en el esquema; es compatible con los modelos de base de datos de DBDesigner 4 y

soporta las novedades incorporadas en MySQL 5.x. (MySQL Workbench, 2019).

### **2.1.13. Netbeans**

Es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. NetBeans IDE es un producto libre y gratuito sin restricciones de uso, es un proyecto de código abierto de gran éxito con una gran base de usuarios, una comunidad en constante crecimiento. Sun Micro Systems fundó el proyecto de código abierto NetBeans en junio de 2000 y continúa siendo el patrocinador principal de los proyectos; actualmente, Sun Microsystems es administrado por Oracle Corporation. (Netbeans, 2019).

### **2.1.14. Java**

Java es un lenguaje de programación y una plataforma informática comercializada por primera vez en 1995 por Sun Microsystems. Hay muchas aplicaciones y sitios web que no funcionarán a menos que tenga Java instalado y cada día se crean más. Java es rápido, seguro y fiable. Desde portátiles hasta centros de datos, desde consolas para juegos hasta súper computadoras, desde teléfonos móviles hasta Internet, Java está en todas partes.(Java, 2019).

### **2.1.15. Joone**

Joone es un framework de red neuronal gratuito para crear, entrenar y probar redes neuronales artificiales, está compuesto por un motor central que es el punto de apoyo de todas las aplicaciones que se desarrollan con Joone. Las aplicaciones de Joone están compuestas de componentes, que son módulos de código conectables, reutilizables y persistentes. Los expertos y diseñadores de inteligencia artificial pueden crear aplicaciones pegando compo-

nentes con un editor gráfico y controlando la lógica con scripts. (Marrone, 2007).

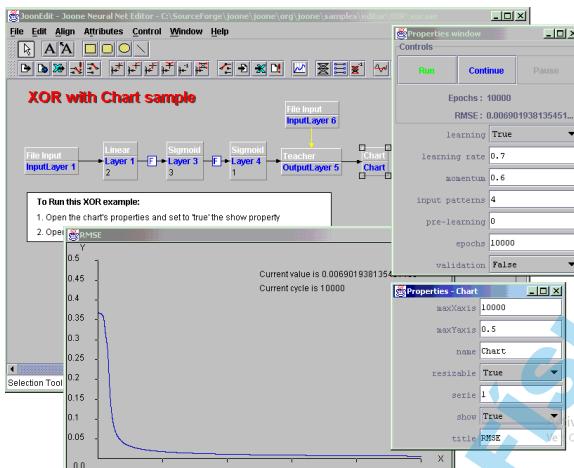


Figura 2.12: Editor Joone.

Fuente: Marrone (2007).

### 2.1.15.1. Características

Las características de Joone según Marrone (2007):

**Arquitectura**, el framework de Joone está construido con una arquitectura modular: el “motor central” está separado de la interfaz visual y permite implementar fácilmente cualquier aplicación nueva basada en él.

**Portabilidad**, Joone es portátil, está escrito en Java puro y puede ejecutarse en cualquier entorno, desde grandes máquinas multiprocesador hasta pequeños dispositivos de bolsillo.

**Expansión**, el framework es expandible con más componentes para implementar nuevos algoritmos de aprendizaje o nuevas arquitecturas. Con Joone es posible implementar cualquier tipo de optimización. Existen dos métodos principales para encontrar la mejor solución para un problema determinado (es decir, para encontrar la mejor red neuronal): optimización local y técnicas de optimización global.

**Multiproceso y escalabilidad**, el motor central de Joone se basa en un motor multiproce-

so, capaz de escalar utilizando todos los recursos informáticos disponibles. Joone proporciona a los usuarios profesionales un entorno distribuido para entrenar muchas redes neuronales en paralelo en varias máquinas.

## **2.2. Método de la investigación**

### **2.2.1. Tipo de investigación**

La presente tesis es del tipo experimental.

### **2.2.2. Variables de la Investigación**

#### **2.2.2.1. Variable Dependiente**

Clasificación automática de documentos.

#### **2.2.2.2. Variable Independiente**

Aplicación de sistemas inteligentes.

### **2.2.3. Operacionalización de la variable**

Nos permiten realizar mediciones y determinar la validez de la hipótesis que fue planteada en la investigación.

Tabla 2.2: Indicadores para en nuestra investigación.

Variable dependiente	Descripción	Indicador	
Clasificación automática de documentos	Técnica inteligente para ordenar un conjunto de documentos por clase o categoría.	Sensibilidad: Precisión: Exactitud: Tiempo:	$\frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{falsos negativos}}$ $\frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{falsos positivos}}$ $\frac{\text{Verdaderos positivos} + \text{falsos positivos}}{\text{Verdaderos positivos} + \text{Verdaderos negativos}}$ $\frac{\text{Total de documentos}}{\text{Representado en segundos}}$

Fuente: Elaboración propia.

## 2.3. Recolección de datos para la elaboración del modelo

### 2.3.1. Técnica

Observación, fue la técnica aplicada para la recolección de datos, específicamente la observación estructurada. Indicando como únicas respuestas 1 = SI, 0 = NO (ver Tabla: A.1).

### 2.3.2. Población

La población para nuestra investigación son todos los posibles tipos de documentos de las diferentes oficinas de la Universidad nacional de Trujillo (ver Tabla: A.4), la cual es una población infinita.

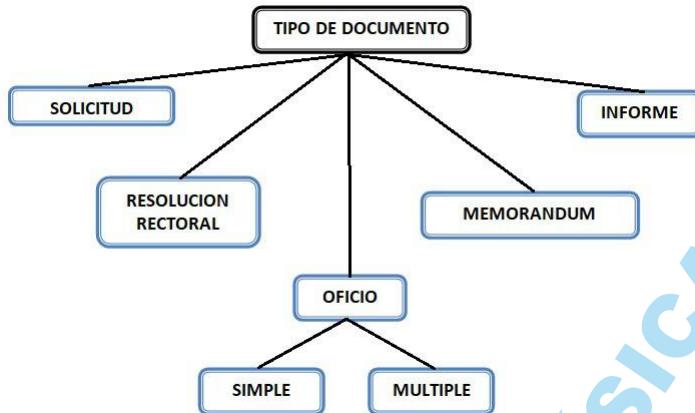


Figura 2.13: Diagrama de tipos de documentos.

Fuente: Elaboración propia.

### 2.3.3. Muestra

Este conjunto de documentos está conformado por una población de 134 documentos, los cuales fueron datos tomados de la recepción en una oficina al azar de la Universidad Nacional de Trujillo, Dirección de Calidad Universitaria, estos documentos fueron editados de tal forma que no permita dar a conocer nombres o datos que podrían ser utilizados con otros fines que no sean de estudio o investigación. Los datos obtenidos fueron:

Tabla 2.3: Datos obtenidos para la investigación.

Día	Cantidad de documentos
Lunes 26	28
Martes 27	31
Miércoles 28	22
Jueves 29	32
Viernes 30	21
Total	134

Fuente: Elaboración propia.

Tabla 2.4: Tipos de documentos y sus cantidades.

Tipo de documento	Cantidad
Oficio	54
Oficio múltiple	14
Informe	24
Solicitud	22
Resolución rectoral	14
Memorandum	6
<b>Total</b>	<b>134</b>

Fuente: Elaboración propia.



Figura 2.14: Diagrama de tipos de documentos.

Fuente: Elaboración propia.

Estos datos fueron recogidos de la oficina y tienen lugar en el tiempo en la última semana de agosto del 2019.

### 2.3.3.1. Muestra por conveniencia

Permite seleccionar aquellos casos accesibles que acepten ser incluidos. Esto, fundamentado en la conveniente accesibilidad y proximidad de los sujetos para el investigador. (Otzen and Manterola, 2017).

## 2.4. Etapas de la investigación

Nuestra investigación comprenderá las siguientes etapas:

- Documentación e investigación bibliográfica de los principales y necesarios conceptos para la elaboración de nuestra investigación, como redes neuronales, capas ocultas, entrenamiento, por mencionar algunos, entre otros.
- Recolección de nuestra población, documentos, los cuales se utilizó en el desarrollo de la investigación.
- Análisis, de las principales técnicas de clasificación de documentos para después aplicarlas en nuestro modelo.
- Implementación de dichas técnicas en nuestro modelo propuesto de investigación.
- Teniendo en cuenta que se necesita entrenar a las redes neuronales, puesto que son que las redes neuronales son heurísticas se requieren un constante ajuste para obtener un óptimo resultado.
- Resultados; para el análisis de resultados, se usa una Matriz de confusión, en la cual se pueden conocer los indicadores, estos son: efectividad, precisión, tasa de acierto y tiempo, los cuales son el sistema de medición utilizado para la evaluación de la calidad del modelo clasificador de documentos.

## Capítulo 3

# Sistema inteligente para la clasificación automática de documentos

### 3.1. Análisis

#### 3.1.1. Requerimiento funcional

La siguiente imagen (ver Figura: 3.1) hace referencia y describen las actividades de nuestra aplicación.



Figura 3.1: Requerimiento funcional de investigación.

Fuente: Elaboración propia.

Usuario-administrador:

- Cargar documento y subir los documentos a clasificar.
- Reportar documentos clasificados, por tipo y por destino.
- Mostrar gráfico de porcentaje de la cantidad de documentos clasificados, por tipo y por destino.

Administrador:

- Administrar permisos y ver los accesos que tiene los usuarios.
- Registrar, nuevo usuario para que tenga acceso a la aplicación.

### 3.1.2. Requerimiento no funcional

Describe otras prestaciones, características y/o limitaciones de la aplicación.

Tabla 3.1: Requerimiento no funcional de nuestra investigación.

Requerimiento no funcional	Descripción
<b>Usabilidad</b>	Aplicación de fácil manejo, con interfaces gráficas simples y con iconos predictivos.
<b>Eficiencia</b>	Aplicación eficiente, ligera y con respuesta rápida, en 300 milésimas de segundo aproximadamente por documento.
<b>Dependibilidad</b>	Aplicación disponible. Aplicación confiable. Aplicación integral, sin alteraciones. Fácil mantenimiento de la aplicación.
<b>Seguridad</b>	Tiene control para el acceso de información.

Fuente: Elaboración propia.

## 3.2. Diseño

### 3.2.1. Diagrama de clases

Modelo el cual seguiremos en el desarrollo del sistema.

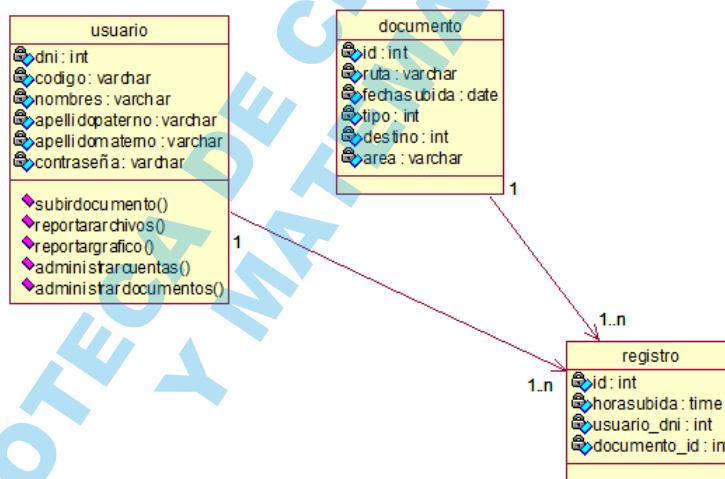


Figura 3.2: Modelo de diagrama de clases para la investigación de la tesis.

Fuente: Elaboración propia.

### 3.2.2. Diagrama de modelos de estados

Modelo de estados el cual seguiremos en el desarrollo del sistema.

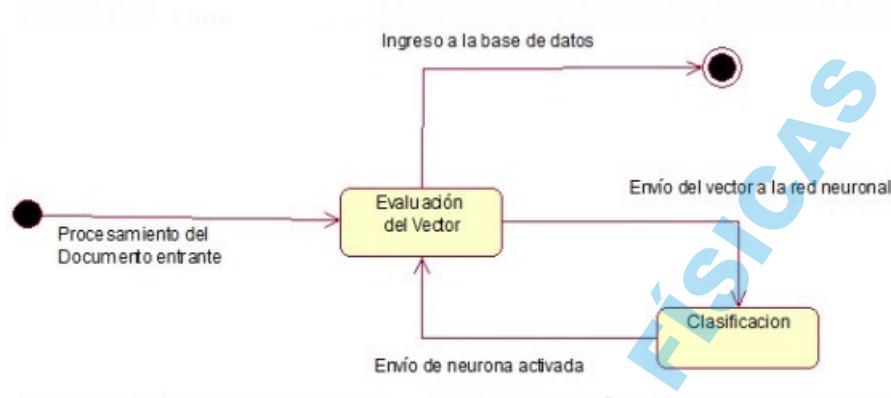


Figura 3.3: Modelo de estados para la investigación de la tesis.

Fuente: Elaboración propia.

### 3.2.3. Diagrama de casos de uso

Modelo de casos de uso el cual seguiremos en el desarrollo del sistema.

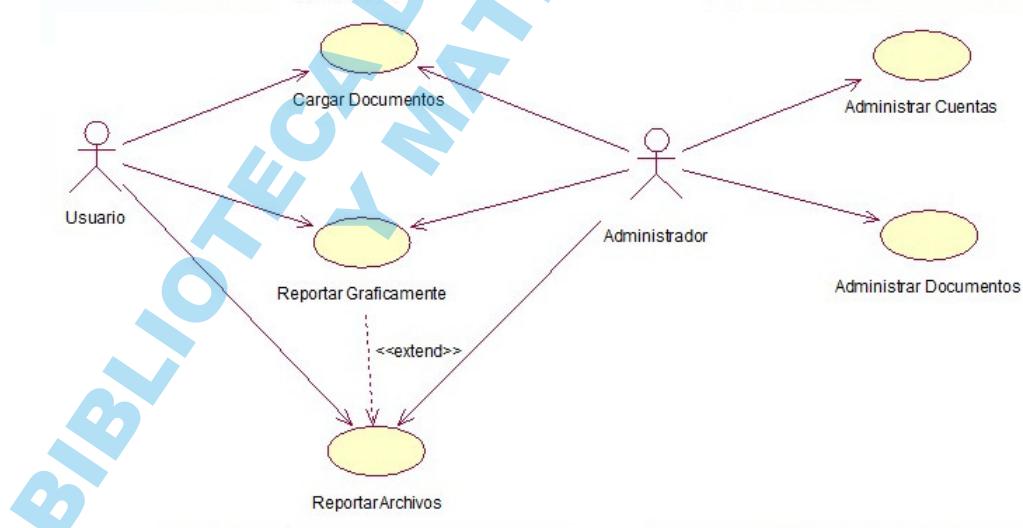


Figura 3.4: Modelo de casos de uso para la investigación de la tesis.

Fuente: Elaboración propia.

- Cargar documentos, permite subir documentos digitalizados a la aplicación.
- Administrar cuentas, permite al administrador administrar las cuentas así como ver el registro de cada usuario y crear usuarios nuevos.
- Administrar documentos, permite al administrador ver la hora y fecha de la carga de documentos por usuario.
- Reporte archivos, permite solicitar los documentos que se subieron a la aplicación.

### 3.2.4. Diagrama de componentes

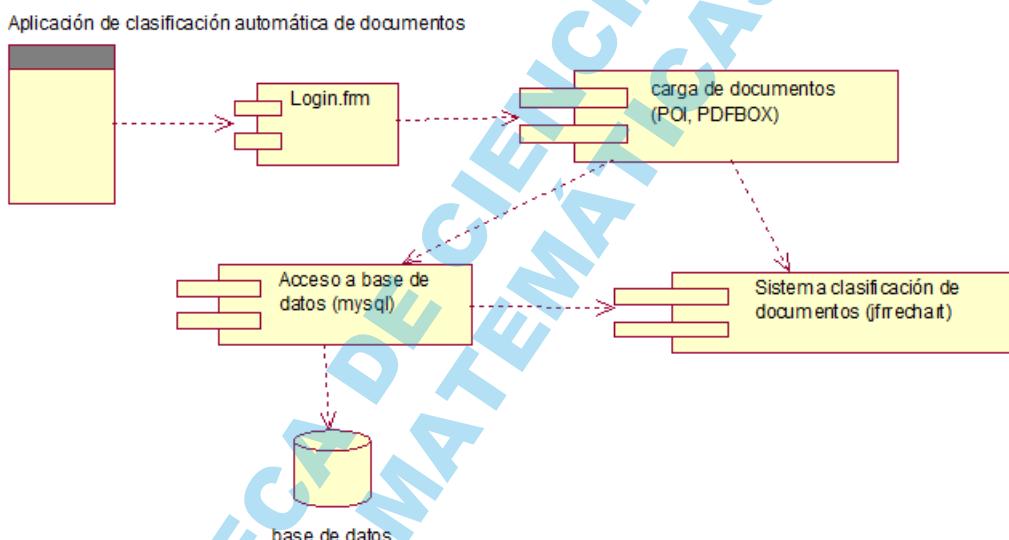


Figura 3.5: Diagrama de componentes de nuestra investigación.

Fuente: Elaboración propia.

### 3.2.5. Diagrama de secuencia

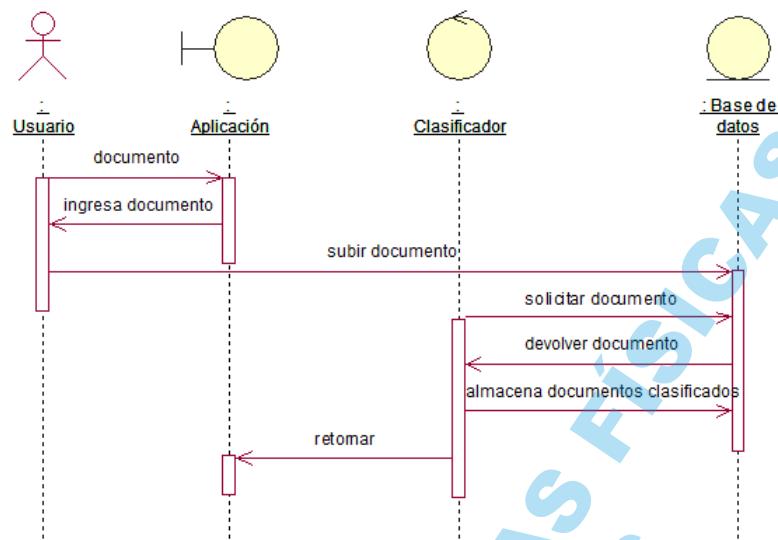


Figura 3.6: Modelo de secuencia para clasificar automáticamente documentos.

Fuente: Elaboración propia.

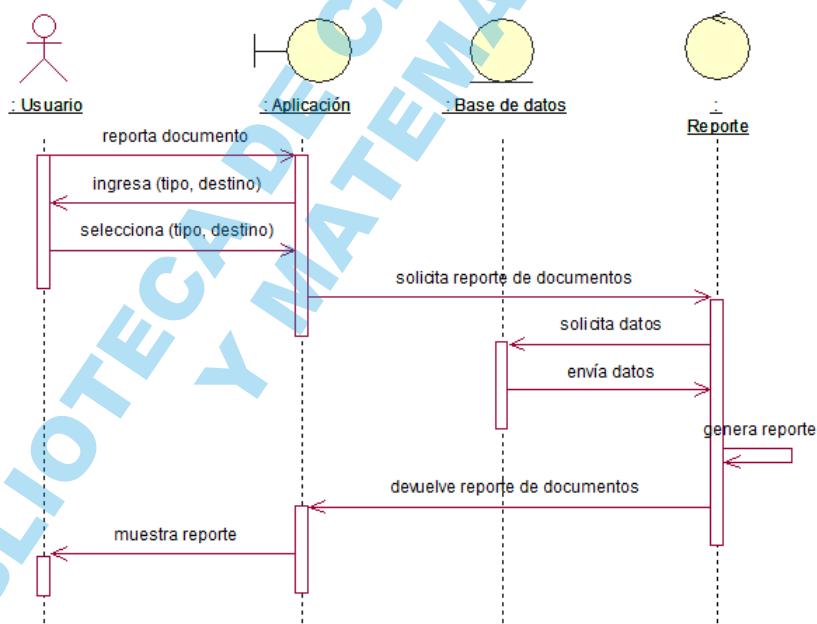


Figura 3.7: Modelo de secuencia para reporte de documentos.

Fuente: Elaboración propia.

### 3.2.6. Diagrama de la base de datos entidad relación

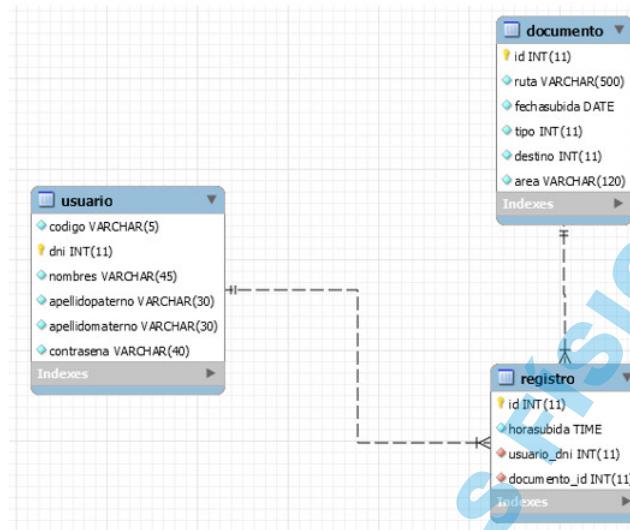


Figura 3.8: Diagrama de la base de datos de entidad relación.

Fuente: Elaboración propia.

### 3.2.7. Diseño de interfaces

A continuación se muestran las interfaces de la aplicación:

En la Figura 3.9, se muestra el login tanto para el administrador como para los usuarios.



Figura 3.9: Login de la aplicación.

Fuente: Elaboración propia.

En la Figura 3.10, se muestra todas las opciones que tiene el administrador después de haber ingresado su login.



Figura 3.10: Interfaz principal.

Fuente: Elaboración propia.

En la Figura 3.11, se muestra la interfaz de cargar los documentos luego de dar click en la opción de cargar documento de la Figura 3.10.



Figura 3.11: Interfaz de la carga de los documentos.

Fuente: Elaboración propia.

En la Figura 3.12, se muestra una interfaz para seleccionar los multiples documentos a subir.

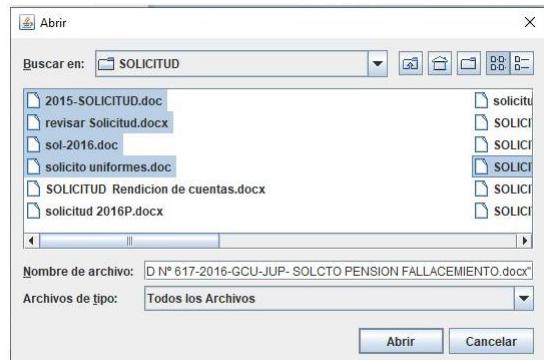
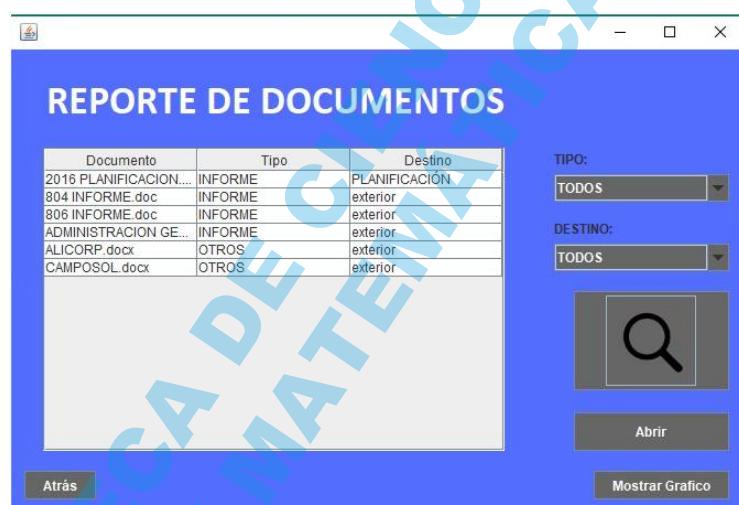


Figura 3.12: Interfaz de la búsqueda múltiple de documentos.

Fuente: Elaboración propia.

En la Figura 3.13, se muestra el reporte de los documentos cargados en la aplicación, listos para su selección y muestra en cualquiera de los formatos (.PDF, .DOCX o .DOC).



Documento	Tipo	Destino
2016 PLANIFICACION...	INFORME	PLANIFICACIÓN
804 INFORME.doc	INFORME	exterior
806 INFORME.doc	INFORME	exterior
ADMINISTRACIÓN GE...	INFORME	exterior
ALICORP.docx	OTROS	exterior
CAMPOSOL.docx	OTROS	exterior

TIPO:

DESTINO:



Abrir

Mostrar Gráfico

Figura 3.13: Interfaz de reporte de los documentos.

Fuente: Elaboración propia.

### 3.2.8. Diseño de algoritmo

El siguiente diagrama de flujo en el cual mostramos el proceso de implementación computacional que seguimos para el desarrollo.

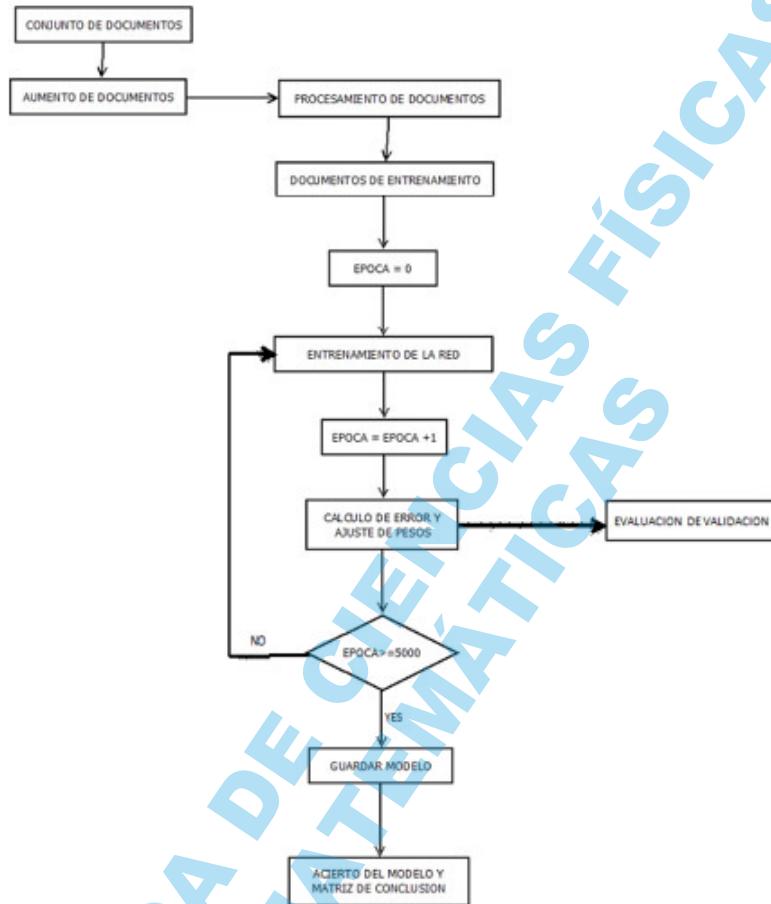


Figura 3.14: Diagrama de flujo para el desarrollo de la investigación.

Fuente: Elaboración propia.

### 3.2.9. Extracción de datos

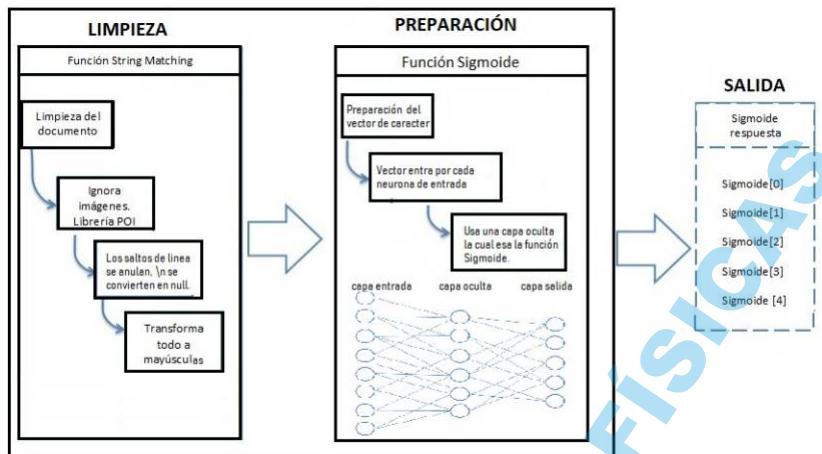


Figura 3.15: Extracción de datos.

Fuente: Elaboración propia.

Limpieza: Extracción de los caracteres necesarios para formar el vector. Con la librería POI eliminamos todas las imágenes y encabezados (marca de agua) del documento, para poder llegar al texto a utilizar. El algoritmo convierte el salto de línea en null, así evitamos espacios en blanco. El algoritmo convierte todas las letras a mayúsculas para una mejor lectura.

Preparación: Se prepara el vector de carácter, el vector entra por cada neurona de entrada. Usa una capa oculta, la cual con la función sigmoide nos da una respuesta y la más aproximada a uno nos dará como respuesta el tipo de documento.

Salida: La función sigmoide nos da como respuesta un valor, para sigmoid[0] el tipo de documento asignado es informe, para sigmoid[1] el tipo de documento asignado es solicitud, para sigmoid[2] el tipo de documento asignado es memorandum, para sigmoid[3] el tipo de documento asignado es oficio y para sigmoid[4] el tipo de documento asignado es resolución rectoral.

Informe	Solicitud	Memorandum	Oficio	Resolución Rectoral	Múltiple	Pedido	Reclamo
1	0	0	0	0	0	1	0

Figura 3.16: Descripción del patrón.

Fuente: Elaboración propia.

### 3.2.10. Arquitectura de la aplicación

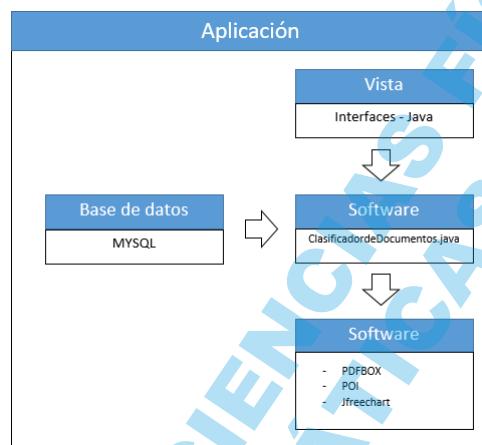


Figura 3.17: Arquitectura de la aplicación.

Fuente: Elaboración propia.

## 3.3. Implementación

Nos enfocamos en la resolución de la tesis y el desarrollo del sistema inteligente, un software realizado en java en la plataforma de Netbeans 8.1, como gestor de datos utilizaremos Workbench MySQL y Joone como diseñador de la red neuronal, el cual nos mostrará los resultados óptimos de nuestra clasificación de documentos a través de nuestra aplicación. Detallamos cómo y qué conceptos tuvimos en cuenta para el desarrollo de la aplicación, mostramos también los procesos de implementación computacional más importantes que fueron

realizados durante el desarrollo de la investigación.

### 3.4. Entrenamiento

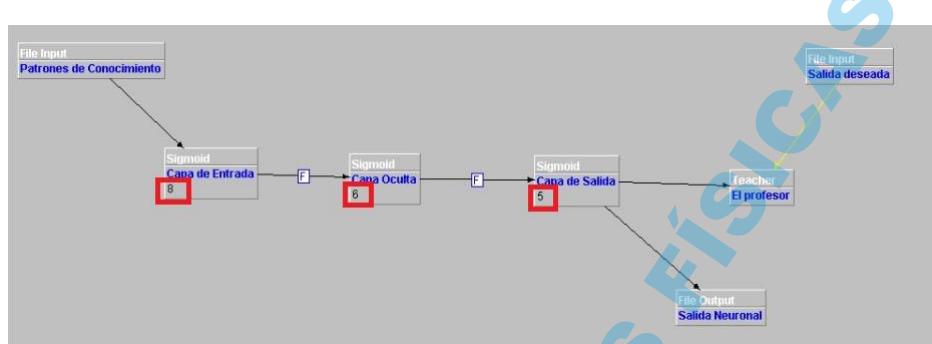


Figura 3.18: Red neuronal en Joone.

Fuente: Elaboración propia.

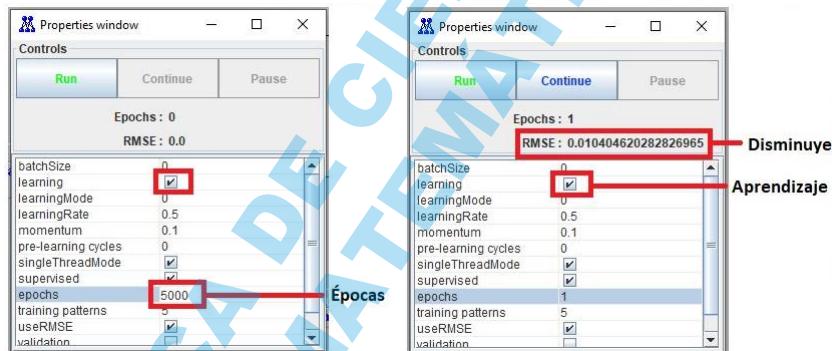


Figura 3.19: Entrenamiento de las 5000 épocas.

Fuente: Elaboración propia.

El cual se hizo con un total de 5000 épocas.

Los datos utilizados fueron extraídos de diversas oficinas de la Universidad Nacional de Trujillo, los cuales conforman la Dataset.

Tabla 3.2: Datos del entrenamiento

Función	Cantidad	Tipo
Entrenamiento	400 documentos	No especificados, tomados solo para esta función
Validación	100 documentos	Oficio = 20 Solicitud = 20 Memorandum = 20 Informe = 20 Resolución rectoral = 20
Muestra	134 documentos	Oficio = 68 Solicitud = 22 Memorandum = 6 Informe = 24 Resolución rectoral = 14

Fuente: Elaboración propia.

Las siguientes tablas muestran nuestras pruebas en el desarrollo de la red.

5.98110098	-4.9181375	-3.2990561	-0.0048796	0.17507329
-3.8951318	-0.0406367	-1.9321498	4.70906128	-1.0464711
-0.9862924	-0.1359934	3.87899577	-6.1465544	-1.0262063
-2.0577522	6.15774414	-3.6682604	-3.2249468	0.40149572
-1.6702981	1.41010313	5.65123345	-3.076806	-4.8624826
6.23790148	-1.151774	-4.9715872	-1.6116612	-0.1031716
-1.3900638	-6.9478075	-0.4391444	-1.4193106	4.76703362
0.18448666	0.6588261	2.09502072	3.47731552	-7.1811308
-5.7543816	4.1316186	-5.6743162	0.85131516	3.56747955
-0.9523734	-1.2967779	0.40931399	3.95046364	-3.5268163

Figura 3.20: Pesos de salida.

Fuente: Elaboración propia.

Tabla 3.3: Vector de aprendizaje.

Vectores de aprendizaje								
Informe	Solicitud	Memorandum	Oficio	Resolución	Múltiple	Pedido	Reclamo	
1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0

Fuente: Elaboración propia.

5.98110098	-4.9181375	-3.2990561	-0.0048796	0.17507329
-3.8951318	-0.0406367	-1.9321498	4.70906128	-1.0464711
-0.9862924	-0.1359934	3.87899577	-6.1465544	-1.0262063
-2.0577522	6.15774414	-3.6682604	-3.2249468	0.40149572
-1.6702981	1.41010313	5.65123345	-3.076806	-4.8624826
6.23790148	-1.151774	-4.9715872	-1.6116612	-0.1031716
-1.3900638	-6.9478075	-0.4391444	-1.4193106	4.76703362
0.18448666	-0.6588261	2.09502072	3.47731552	-7.1811308
-5.7543816	4.1316186	-5.6743162	0.85131516	3.56747955
-0.9523734	-1.2967779	0.40931399	3.95046364	-3.5268163

Figura 3.21: Pesos de salida.

Fuente: Elaboración propia.

-0.4386899
-0.5184561
-0.5417869
-0.0894253
-0.4165299
-0.6130218
-0.1370983
-0.0689692
-0.2013986
-0.5464507
0.4939044
-0.5663579

Figura 3.22: Bias.

Fuente: Elaboración propia.

### 3.5. Funcionamiento de las redes neuronales

Al cargar un documento se va ingresando el vector característico y se interpreta en la red neuronal. Este es un caso de cómo la red neuronal va a ir funcionando con este documento. Los vectores característicos van a responder cuando se activen y estos van a dar lugar a activar la neurona.

#### Funciones:

Una de las principales funciones es “CargarDocumentos.java” la cual usa un componente que le permite una selección múltiple, una ventaja del software es que recibe documentos que estén guardados en diferentes formatos, con la librería POIFS nos permite reconocer las extensiones .doc, .docx., .pdf, de la misma forma algún documento que se ingrese en otro tipo de formato o ya sea como imagen (.jpg, .jpeg, .png, .gif) lo leerá como invalido.

La función BuscarDocumentos, se desarrolló buscando las características que tiene un documento para clasificarlo como un tipo específico.

En este caso el documento por evaluar será oficio, la neurona en la capa de entrada se activa cuando una parte contiene la palabra solicitud y al mismo tiempo petición.

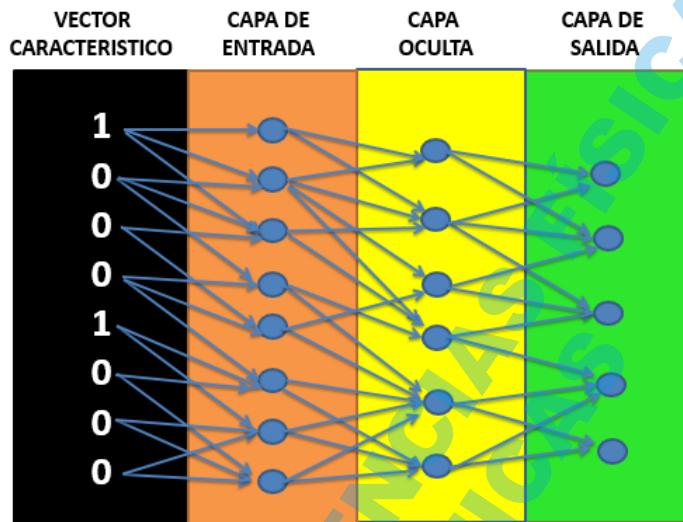


Figura 3.23: Inicio de vector característico.

Fuente: Elaboración propia.

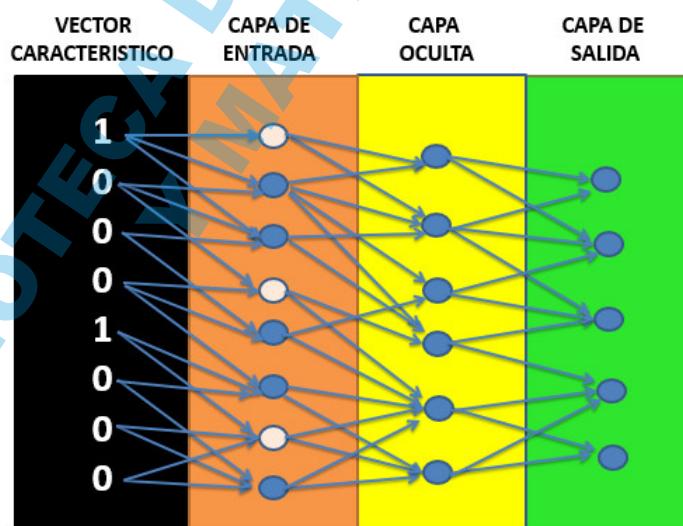


Figura 3.24: Activando neuronas en la capa de entrada.

Fuente: Elaboración propia.

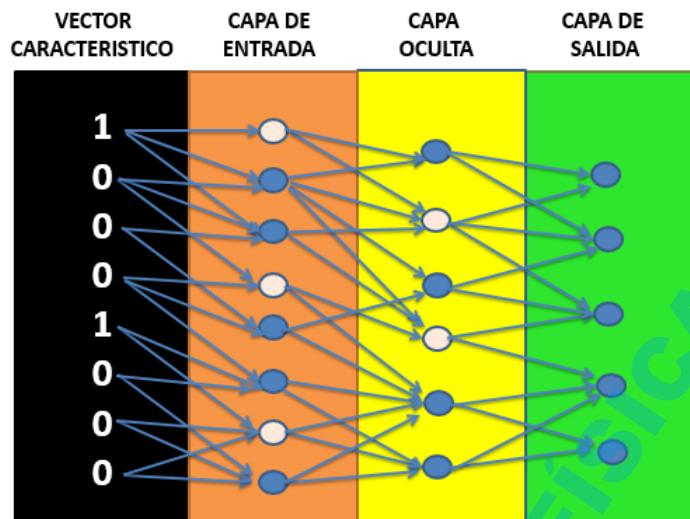


Figura 3.25: Activando neuronas en la capa oculta.

Fuente: Elaboración propia.

Si se activan esos unos, se activan en la capa de entrada esas neuronas.

Y al final en la capa de salida se activa solo una y da la respuesta, cada neurona de salida nos da una respuesta diferente, en este caso la segunda neurona nos da como respuesta una solicitud.

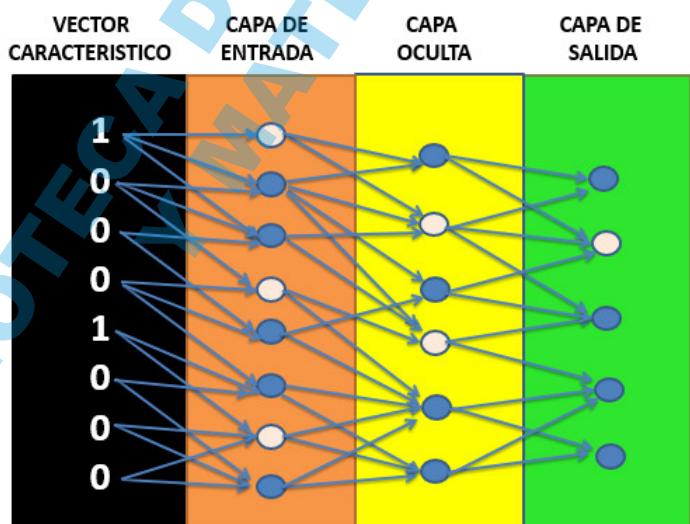


Figura 3.26: Activando neuronas en la capa de salida.

Fuente: Elaboración propia.

### 3.6. Función activación

La neurona no sólo transmite la entrada que recibe, la función de activación utiliza la suma ponderada de la entrada anterior y la transforma como salida. La función sigmoide denotada por `funcionSigmoidea`, donde se denota la constante exponencial, que es aproximadamente 2.71828.

En nuestra aplicación la función sigmoide nos mostrará el resultado de nuestra búsqueda dando numéricamente la opción que más se acerca al resultado que queremos llegar. En este caso “Solicitud” es el resultado que más se aproxima a 1.

```
Sigmoidal [0]- 0.012531487731386662
Sigmoidal [1]- 0.9959445628103717 Solicitud
Sigmoidal [2]- 0.0020733062955850773
Sigmoidal [3]- 0.007473006424699547
Sigmoidal [4]- 0.0015958380984010267
SOLICITUD
INSERT INTO documento(ruta, fechasubida, tipo, destino, area) VALUES ('F:/ANTHONY-2020-NO-BORRAR/')
```

Figura 3.27: Resultado de la función sigmoide en la aplicación.

Fuente: Elaboración propia.

En la Tabla 3.3 nos muestra cuales son los vectores.

## Capítulo 4

# Resultados y discusión de la tesis

En la aplicación del sistema inteligente conseguimos resultados esperados que van de acuerdo con la hipótesis planteada.

### 4.1. Resultados computacionales

Con los indicadores de la Tabla 2.2, se obtuvo los siguientes resultados:

- **Verdadero positivo:** documentos correctos encontrados.
- **Falso positivo:** documentos incorrectos encontrados.
- **Verdadero negativo:** documentos correctos no encontrados.
- **Falso negativo:** documentos incorrectos no encontrados.

#### 4.1.1. Resultado según criterio subir con éxito

A continuación se muestra una tabla:

Tabla 4.1: Tabla de valores predictivos.

	Verdadero (V)	Falso (F)
Positivo (P)	VP	FP
Negativo (N)	VN	FN

Fuente: Elaboración propia.

Donde:

Verdadero, es subir con éxito.

Positivo, es encontró el documento.

Falso, es no subir con éxito.

Negativo, es no encontró el documento.

	OFICIO	INFORME	MEMORANDUM	RESOLUCIÓN RECTORAL	SOLICITUD	OTROS	TOTAL
OFICIO	66	0	0	0	0	0	66
INFORME	0	23	0	0	0	1	24
MEMORANDUM	0	0	4	0	0	0	4
RESOLUCIÓN RECTORAL	0	0	0	11	0	3	14
SOLICITUD	0	0	0	0	19	3	22
OTROS	0	0	0	0	0	4	4
TOTAL	66	23	4	11	19	11	134

Figura 4.1: Resultado detallado para la búsqueda, segun criterio subir con éxito

Fuente: Elaboración propia.

Tabla cruzada Subir con éxito \*Encontró el doc.

Recuento		Encontró el doc.		Total
		0	1	
Subir con éxito	0	7	0	7
	1	4	123	127
Total		11	123	134

Figura 4.2: Tabla cruzada: subir con exito vs. encontró el doc.

Fuente: Elaboración propia.

$$\frac{VP}{VP + FN} = \frac{123}{123 + 7} = \frac{123}{130} = 0,946, \quad 94,6\% \text{ de sensibilidad.} \quad (4.1)$$

$$\frac{VP}{VP + FP} = \frac{123}{123 + 0} = \frac{123}{123} = 1, \quad 100\% \text{ de precisión.} \quad (4.2)$$

$$\frac{VP + VN}{TOTAL} = \frac{123 + 4}{134} = \frac{127}{134} = 0,947, \quad 94,7\% \text{ de exactitud.} \quad (4.3)$$

En los resultados de criterio: Subir con éxito el documento, se obtuvo en el indicador de sensibilidad un 94.6 %, de precisión un 100 % y en exactitud un 94.7 %. Por lo tanto, consideramos nuestra hipótesis como válida.

#### **4.1.2. Resultado según criterio sin error ortográfico**

En la tabla de valores predictivos (ver Tabla: 4.1), se detalla a continuación los valores para el criterio sin error ortográfico:

Verdadero, es subir con éxito.

Positivo, es encontró el documento.

Falso, es no subir con éxito.

Negativo, es no encontró el documento.

	OFICIO	INFORME	MEMORANDUM	RESOLUCIÓN RECTORAL	SOLICITUD	OTROS	TOTAL
OFICIO	67	0	0	0	0	0	67
INFORME	0	23	0	0	0	1	24
MEMORANDUM	0	0	4	0	0	2	6
RESOLUCIÓN RECTORAL	0	0	0	11	0	3	14
SOLICITUD	0	0	0	0	19	3	22
OTROS	0	0	0	0	0	1	1
TOTAL	66	23	4	11	19	11	134

Figura 4.3: Resultado detallado para la búsqueda segun criterio sin error ortográfico.

Fuente: Elaboración propia.

Tabla cruzada Doc. Sin error ortográfico\*Encontró el doc.

Recuento

Doc. Sin error ortográfico	Encontró el doc.		Total
	0	1	
0	10	1	11
1	1	122	123
Total	11	123	134

Figura 4.4: Tabla cruzada: sin error ortográfico vs. encontró el doc.

Fuente: Elaboración propia.

$$\frac{VP}{VP + FN} = \frac{122}{122 + 10} = \frac{122}{132} = 0,9242, \quad 92,42\% \text{ de sensibilidad.} \quad (4.4)$$

$$\frac{VP}{VP + FP} = \frac{122}{122 + 1} = \frac{122}{123} = 0,9918, \quad 99,18\% \text{ de precisión.} \quad (4.5)$$

$$\frac{VP + VN}{TOTAL} = \frac{122 + 1}{134} = \frac{123}{134} = 0,917, \quad 91,7\% \text{ de exactitud.} \quad (4.6)$$

En los resultados del criterio: Sin error ortográfico, se obtuvo en el indicador de sensibilidad un 92.42 %, de precisión un 99.18 % y en el de exactitud un 91.7 %. Por lo tanto consideramos nuestra hipótesis como valida.

## 4.2. Tiempo

En esta parte se miden los resultados que obtuvimos con un grupo de 30 trabajadores, antes y después de utilizar la aplicación.

Tabla 4.2: Tabla base de comparación de tiempos

Tiempo de búsqueda sin la aplicación		Tiempo de búsqueda con la aplicación	
Pasos	Tiempo (segundos)	Pasos	Tiempo (segundos)
Ingresar a (MiPC o Este Equipo)	60	Abrir la aplicación	2.5
Buscar documentos	120	Cargar documentos	2
Abrir el supuesto documento que necesitas	60	Abrir el reporte y ver documento buscado	2
Repetir paso 3 hasta encontrar el documento buscado	300 a más	Abrir documento	0.3

Fuente: Elaboración propia.

Con la Tabla 4.2 se obtuvieron los resultados (ver Tabla: A.2) de los cuales se sacó una media aritmética de aproximadamente 760.2 segundos, de los cuales 22 trabajadores demoraban aproximadamente 600 segundos en buscar un documento y 8 trabajadores demoraban aproximadamente 1200 segundos, lo que hace un total de 22800 segundos para 30 trabajadores

Se aplicó la Tabla: 4.2 y los resultados (ver Tabla: A.3) fueron muy favorables con un total de 16 personas que tardaron aproximadamente 5 segundos en la búsqueda de documentos y 14 personas que tardaron aproximadamente 10 segundos, dándonos un total de 220 segundos para 30 trabajadores, lo cual nos da un media de 7.33 segundos, con este resultado mostramos la optimización de tiempo al usar la aplicación.

Tabla cruzada APROX 10 MIN*APROX 20 MIN				
Recuento		APROX 20 MIN		
		0	1	Total
APROX 10 MIN	0	0	8	8
	1	22	0	22
Total		22	8	30

Figura 4.5: Trabajadores antes de usar la aplicación.

Fuente: Elaboración propia.

Tabla cruzada APROX 5 seg*APROX 10 seg				
Recuento		APROX 10 seg		
		0	1	Total
APROX 5 seg	0	0	14	14
	1	16	0	16
Total		16	14	30

Figura 4.6: Trabajadores usando la aplicación.

Fuente: Elaboración propia.

Estos resultados nos muestran que se llegó al objetivo, cumpliendo con optimizar uno de los principales recursos, el tiempo. Con esto concluimos logrando todos los objetivos planteados para nuestra tesis.

Tabla 4.3: Resultados con tiempos, sin usar la aplicación y usando la aplicación

Sin aplicación	Con aplicación
760.2 segundos	7.33 segundos

Fuente: Elaboración propia.

Logrando reducir el tiempo de 760.2 segundos a 7.33 segundos, esto significa un ahorro de tiempo en más de 752 segundos en la búsqueda por documento.

# Capítulo 5

## Consideraciones finales

### 5.1. Conclusiones

Al finalizar la investigación se pudo cumplir con los objetivos específicos propuestos, los cuales se detallan a continuación:

1. Durante esta investigación se pudo explicar la problemática de la demora en la búsqueda de documentos así como la solución aplicando un software la cual contiene como principal mecanismo el uso de los sistemas inteligentes, en este caso se determinó usar redes neuronales, perceptrón y backpropagation, la cual nos permitió establecer un lapso de tiempo muy reducido al momento de la búsqueda de los documentos.
2. Se cumplió con la implementación de un prototipo para la clasificación de documentos, la cual es muy sencilla y útil para el usuario de cualquier oficina.
3. La investigación se logró una sensibilidad del 94.6 %, precisión del 100 % y exactitud del 94.7 % en el criterio de subir con éxito y encontró el documento.
4. La investigación se logró una sensibilidad del 92.42 %, precisión del 99.18 % y exactitud del 92.42 % en el criterio de subir con éxito y encontró el documento.

titud del 91.7 % en el criterio de documento sin error ortográfico y encontró el documento.

5. Se optimizó reducir el tiempo de 760.2 segundos a 7.33 segundos, esto significa un ahorro de tiempo en más de 752 segundos en la búsqueda por documento.

## 5.2. Trabajos futuros

Si bien esta investigación cumple con la hipótesis y objetivos específicos planteados, durante la presente tesis, se presentaron algunas ideas que podrían mejorar la aplicación, por ejemplo:

- Diseñar más a detalle la jerarquía de las oficinas de la Universidad Nacional de Trujillo, lo cual se podría llegar hasta cada escuela, como sus jerarquías internas dentro de ellas, creando una clasificación mas minuciosa.
- Proponer diseñar una red local en todas las oficinas de la Universidad Nacional de Trujillo para su clasificación de los documentos de manera que permita mejorar aun mas los tiempos de emisión y recepción de documentos a diferentes oficinas.
- Crear un app móvil y de esta forma tener un alcance a más usuarios.

# Referencias bibliográficas

- Cruz, A. (2012). Definiciones de sistemas inteligentes y 5 ejemplos de sistemas inteligentes. url=<http://www.academia.edu/10376257/Definiciones-de-sistemas-inteligentes-y-5-ejemplos-de-sistemas-inteligentes>. [Online; accedido Octubre 05, 2019].
- Erlinb (2017). Redes neuronales: Conceptos básicos y aplicaciones. url=[www.perfectcodex.com/redneuronal/2017/12/27/conceptos-basicos/](http://www.perfectcodex.com/redneuronal/2017/12/27/conceptos-basicos/). [Online; accedido Octubre 01, 2019].
- Espitia Corredor, Daniel Eduardo (2017). Entrenamiento del perceptron. url=[www.platzi.com/tutoriales/1352-ia/2619-entrenamiento-del-perceptron/](http://www.platzi.com/tutoriales/1352-ia/2619-entrenamiento-del-perceptron/). [Online; accedido Octubre 06, 2019].
- Faloutsos, C. and Oard, D. W. (1998). A survey of information retrieval and filtering methods. Technical report, Technical Report CS-TR3514.
- Freeman, J. A. and Skapura, D. M. (1991). *Neural networks: algorithms, applications, and programming techniques*. Addison Wesley Longman Publishing Co., Inc.
- Java (2019). ¿qué es la tecnología java y para qué la necesito? url=[www.java.com/es/download/faq/whatisjava.xml](http://www.java.com/es/download/faq/whatisjava.xml). [Online; accedido Octubre 07, 2019].

Lagos, P. S. (2002). Commonkads y el lenguaje de modelado unificado oé uml. *Ingeniería informática*, page 4.

Marrone, P. (2007). Joone: Java object oriented neural engine-the complete guide. *URL* <http://www.joone.org>.

Matich, D. J. (2001). Redes neuronales: Conceptos básicos y aplicaciones. *Universidad Tecnológica Nacional, México*.

McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In *Philosophical logic and artificial intelligence*, pages 161–190. Springer.

Michael A. Nielsen (2015). Redes neuronales y aprendizaje profundo: Cómo funciona el algoritmo de backpropagation. url=www.neuralnetworksanddeeplearning.com/chap2.html. [Online; accedido Octubre 01, 2019].

Munt, A. M. (2018). Introducción a los modelos de redes neuronales artificiales el perceptrón simple y multicapa. *Universidad de Zaragoza, España*.

MySQL Workbench (2019). Mysql workbench. url=www.mysql.com/products/workbench/. [Online; accedido Octubre 07, 2019].

Netbeans (2019). Apache netbeans ide. [Online; accedido Octubre 07, 2019].

Otzen, T. and Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology*, 35(1):227–232.

Prince, B. D. (2009). Inteligencia artificial. ¿qué es? url=http://inteligencia-artificial-delabarrita.blogspot.pe/2009/10/en-que-consiste-y-en-que-se-fundamenta.html. [Online; accedido Octubre 04, 2019].

Steinbach, M., Karypis, G., Kumar, V., et al. (2000). A comparison of document clustering techniques. In *KDD workshop on text mining*, pages 525–526. Boston.

BIBLIOTECA  
Y  
DE CIENCIAS  
Y  
MATEMÁTICAS  
FÍSICAS

# Apéndice A

## Código de la aplicación

Tabla A.1: Tabla con los documentos de nuestra población

N	Subir con éxito	Documento sin error ortográfico	Encontró el documento
1	1	1	0
2	1	0	0
3	1	1	1
4	1	1	1
5	1	1	1
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	0	0	0
11	1	1	1
12	1	1	1
13	1	1	1
14	1	0	0
15	1	0	0
16	0	0	0
17	0	0	0
18	1	1	1
19	1	1	1
20	1	1	1
21	1	1	1
22	1	1	1

23	1	1	1
24	1	1	1
25	1	1	1
26	1	1	1
27	1	1	1
28	1	1	1
29	1	1	1
30	1	1	1
31	1	1	1
32	1	1	1
33	1	1	1
34	1	1	1
35	1	1	1
36	1	1	1
37	1	1	1
38	1	1	1
39	1	1	1
40	1	1	1
41	1	1	1
42	1	1	1
43	1	1	1
44	1	1	1
45	1	1	1
46	1	1	1
47	1	1	1
48	1	1	1
49	1	1	1
50	1	1	1
51	1	1	1
52	1	1	1
53	0	0	0
54	1	1	1
55	1	1	1
56	1	1	1
57	1	1	1
58	1	1	1
59	1	1	1
60	1	1	1
61	1	1	1
62	1	1	1
63	1	1	1
64	1	1	1
65	1	1	1
66	1	1	1

67	1	1	1
68	0	0	0
69	1	1	1
70	1	1	1
71	1	1	1
72	1	1	1
73	1	1	1
74	1	1	1
75	1	1	1
76	1	1	1
77	1	1	1
78	1	1	1
79	1	1	1
80	1	1	1
81	1	1	1
82	1	1	1
83	1	1	1
84	1	1	1
85	1	1	1
86	1	1	1
87	1	1	1
88	1	1	1
89	1	1	1
90	1	1	1
91	1	1	1
92	1	1	1
93	1	1	1
94	1	1	1
95	1	1	1
96	1	1	1
97	1	1	1
98	1	1	1
99	1	1	1
100	1	1	1
101	1	1	1
102	1	1	1
103	0	0	0
104	1	1	1
105	1	1	1
106	1	1	1
107	1	1	1
108	1	1	1
109	1	1	1
110	1	1	1

111	1	0	1
112	1	1	1
113	1	1	1
114	1	1	1
115	1	1	1
116	1	1	1
117	1	1	1
118	1	1	1
119	1	1	1
120	1	1	1
121	1	1	1
122	1	1	1
123	1	1	1
124	1	1	1
125	1	1	1
126	1	1	1
127	1	1	1
128	0	0	0
129	1	1	1
130	1	1	1
131	1	1	1
132	1	1	1
133	1	1	1
134	1	1	1

Fuente: Elaboración propia.

Tabla A.2: Encuesta a trabajadores sin la aplicación.

N de trabajadores	Aprox. 10 minutos	Aprox. 20 minutos
1	1	0
2	1	0
3	1	0
4	0	1
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	0	1
11	1	0
12	0	1
13	1	0
14	1	0
15	0	1
16	0	1
17	0	1
18	1	0
19	1	0
20	1	0
21	0	1
22	1	0
23	1	0
24	1	0
25	1	0
26	1	0
27	1	0
28	1	0
29	1	0
30	0	1

Fuente: Elaboración propia.

Tabla A.3: Encuesta a trabajadores con la aplicación.

N de trabajadores	Aprox. 5 segundos	Aprox. 10 segundos
1	1	0
2	1	0
3	0	1
4	0	1
5	1	0
6	1	0
7	1	0
8	0	1
9	1	0
10	0	1
11	1	0
12	0	1
13	0	1
14	1	0
15	0	1
16	0	1
17	0	1
18	1	0
19	0	1
20	1	0
21	0	1
22	1	0
23	0	1
24	1	0
25	1	0
26	1	0
27	1	0
28	0	1
29	1	0
30	0	1

Fuente: Elaboración propia.

Tabla A.4: Destino de los documentos.

<b>Tipo de documento</b>	
Facultades	Ciencias Sociales
	Ingeniería
	Medicina
	Enfermería
	Ciencias Económicas
	Ciencias Agropecuarias
	Ingeniería Química
	Ed. y cc. de la comunicación
	Farmacia y Bioquímica
	Ciencias Físicas y Matemáticas
Area administrativa	Derecho y Ciencias Políticas
	Ciencias biológicas
	Estomatología
	Unidad de presupuesto
	Org. control institucional
Rector	Unidad de recursos humanos
	Unidad de contabilidad
	Unidad de tesorería
	Dirección de calidad universitaria
	Unidad de logística
	Vice rectorado
	Dirección de relaciones internacionales
Exteriores	Dirección de planificación y desarrollo
	Dirección general de administración
Exteriores	Dirección de investigación y desarrollo
	Rectorado (viáticos)
	Secretaría general
Universidades	
Empresas	

Fuente: Elaboración propia.



Figura A.1: Inicio de la interfaz de la aplicación.

Fuente: Elaboración propia.

The screenshot shows a table titled "ADMINISTRACION DE CUENTAS" with four columns: DNI, Nombre, Apellido Paterno, and Apellido Materno. The data is as follows:

DNI	Nombre	Apellido Paterno	Apellido Materno
333333	Administrador		
46795930	PAUL	SANCHEZ	RODRIGUEZ
70481511	Luis	Contreras	Pumamango

At the bottom left is a "Atras" button, and at the bottom right is an "Añadir Usuario" button.

Figura A.2: Interfaz de administración de los usuarios.

Fuente: Elaboración propia.

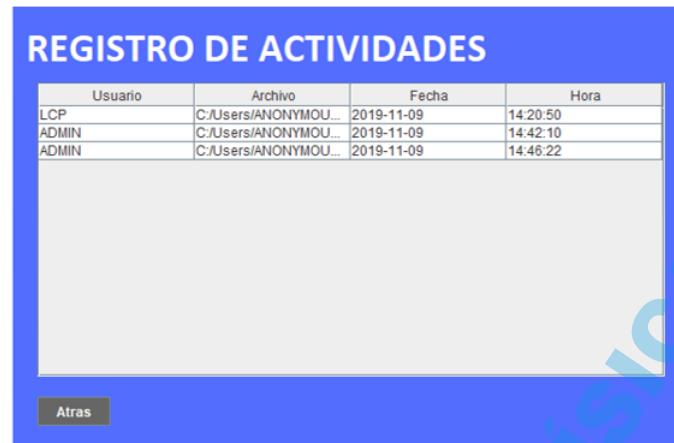


Figura A.3: Interfaz de registro de actividades.

Fuente: Elaboración propia.



Figura A.4: Destino de los documentos.

Fuente: Elaboración propia.

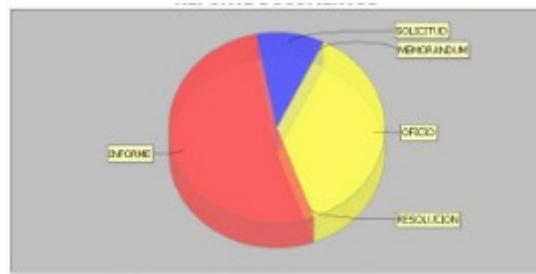


Figura A.5: Tipos de documentos.

Fuente: Elaboración propia.

```
JFileChooser chooser = new JFileChooser();
chooser.setMultiSelectionEnabled(true);
Component frame = null;
chooser.showOpenDialog(frame);
File[] files = chooser.getSelectedFiles();
for(int i = 0 ; i<files.length;i++){
    try {
        procesandoDocumento(files[i]);
    } catch (SQLException ex) {
        Logger.getLogger(CargarDocumentos.class.getName()).log(Level.SEVERE, null, ex);
    }
}
/*
```

Figura A.6: Función para cargar muchos documentos a la vez.

Fuente: Elaboración propia.

```
Documento doc = new Documento(ruta,idUser);
Pattern pattern = Pattern.compile(".*doc");
Matcher matcher = pattern.matcher(doc.getRuta());
boolean identificado = false;
if (matcher.matches()){
    identificado = true;
    doc.readDocFile();
}
pattern = Pattern.compile(".*docx");
matcher = pattern.matcher(doc.getRuta());
if (matcher.matches()){
    identificado = true;
    doc.readDocxFile();
}
pattern = Pattern.compile(".*pdf");
matcher = pattern.matcher(doc.getRuta());
if (matcher.matches()){
    identificado = true;
}
if(!identificado){
    System.out.println("INVALIDO");
}
```

Figura A.7: La librería POI permite ingresar los tipos de documentos: .doc, .docx y .pdf.

Fuente: Elaboración propia.

```
/**Get the total number of paragraphs*/
String[] paragraphs = we.getParagraphText();
List<String> parrafos = new ArrayList<String>();
for (int i = 0; i < paragraphs.length; i++) {
    if(paragraphs[i].toString().length()>4){
        parrafos.add(paragraphs[i].toString().toUpperCase());
    }
}
String texto = "";
int lineasMinimas = 10;
if (parrafos.size()<10){
    lineasMinimas = parrafos.size();
}
for (int i = 0; i< lineasMinimas; i++){
    String cadena = parrafos.get(i).replaceAll("\n\r", "");
    cadena = cadena.replaceAll("\t", " ");
    cadena = cadena.replaceAll("\n", " ");
    cadena = cadena.replaceAll("\r", " ");
    cadena = cadena.replaceAll("      ", " ");
    texto += cadena+" ";
}
procesarTexto(texto,rutaarchivo);
}
```

Figura A.8: Evaluación del arreglo.

Fuente: Elaboración propia.

```
package clasificadordedocumentos.Utils;  
import javax.swing.JOptionPane;  
  
public class RedNeuronal {  
  
    public static int analizarTipo(int[] patron){  
        int filas_entrada = 1;  
        int col_entrada = 7;  
        int neuronas_entrada = 7;  
        int neuronas_oculta = 6;  
        int neuronas_salida = 5;  
        int i, j, k;  
  
        double[][] pesos_entrada_oculta = {{-7.444237021930663,-0.37779879764022867,-1.4090979425538694,4.8393627918  
{0.26435945646199077,6.1783006354098,4.147205266095795,-6.008878470974494,2.5589069738195236,-7.530549393450167},  
{6.91647722081312,-5.720326746168036,-0.7583900767178557,7.8460792472112875,-3.54771553412731,-2.101905197211127},  
{-0.7151708207310346,0.44021000187774884,0.27921740154286595,0.35391748113143595,-0.11319220477188462,-0.29684517531  
{-3.071086112498734,2.779491338230045,-0.9364620510641689,-4.208826804763281,-8.143339426841514,7.0688761447254285},  
{9.70034200031392,-9.072453235667107,-5.332641429696277,-6.706186653064318,0.5767424715924138,4.585572412697804},  
{-0.9838189459272035,0.5156963822861802,0.2908343170394419,0.7649285188140318,-0.02239220620860778,-0.12351362220695  
  
        double[][] pesos_oculta_salida = {{-10.835681064068183,-4.861042123983531,5.484828617313627,-4.9395642763345  
{-2.05423308258619,9.27511976951813,-7.383329675485055,2.853373503345491,-5.6112188369101474},
```

Figura A.9: Se muestra los valores de la red neuronal.

Fuente: Elaboración propia.

```
public class Usuario {  
    private int dni;  
    private String nombres;  
    private String apellidopaterno;  
    private String apellidomaterno;  
    private String contrasena;  
  
    public Usuario(int dni, String nombres, String apellidopaterno, String apellidomaterno,  
        this.dni = dni;  
        this.nombres = nombres;  
        this.apellidopaterno = apellidopaterno;  
        this.apellidomaterno = apellidomaterno;  
        this.contrasena = contrasena;  
    }
```

Figura A.10: Se muestra los nombres de las variables del usuario.

Fuente: Elaboración propia.



## UNIVERSIDAD NACIONAL DE TRUJILLO

### DECLARACION JURADA RR-384-2018/UNT

Los autores suscritos en el presente documento DECLARAMOS BAJO JURAMENTO que somos los responsables legales de la calidad y originalidad del contenido del Proyecto de Investigación Científica, así como, del Informe de Investigación Científica realizado.

Titulado: Aplicación de sistemas inteligentes para la clasificación automática de documentos

#### PROYECTO DE INVESTIGACIÓN CIENTÍFICA

PROY DE TRABAJO DE INVESTIGACIÓN (PREGRADO)   
PROYECTO DE TESIS PREGRADO   
PROYECTO DE TESIS MAESTRÍA   
PROYECTO DE TESIS DOCTORADO

#### INFORME FINAL DE INVESTIGACIÓN CIENTÍFICA

TRABAJO DE INVESTIGACIÓN (PREGRADO)  (X)  
TESIS PREGRADO   
TESIS MAESTRÍA   
TESIS DOCTORADO

El equipo investigador Integrado por:

Nº	APELLIDOS Y NOMBRES	FACULTAD/DEPARTAMENTO	CATEGORIA DOCENTE ASESOR	CÓDIGO Docente Número Matrícula del estudiante	Autor Coautor Asesor
01	Contreras Pumamungo Luis Anthony	FCFYM		052701510	Autor
02	Sánchez Rodríguez Paul Denis	FCFYM		012701110	Autor

Trujillo, 20 de noviembre de 2019

FIRMA

DNI

70493427

FIRMA

DNI

46795930

FIRMA

DNI



UNIVERSIDAD NACIONAL DE TRUJILLO

CARTA DE AUTORIZACIÓN DE PUBLICACIÓN DE TRABAJO DE INVESTIGACIÓN EN  
REPOSITORIO DIGITAL RENATI-SUNEDU  
RR-384-2018/UNT

Trujillo, 20 de noviembre de 2019

Los autores suscritos del INFORME FINAL DE INVESTIGACIÓN CIENTIFICA

Titulado: Aplicación de sistemas inteligentes para la clasificación automática de documentos

AUTORIZAMOS SU PUBLICACIÓN EN EL REPOSITORIO DIGITAL INSTITUCIONAL, REPOSITORIO RENATI-SUNEDU, ALICIA-CONCYTEC CON EL SIGUIENTE TIPO DE ACCESO:

- A. Acceso Abierto:
- B. Acceso Restringido  (datos del autor y resumen del trabajo)
- C. No autorizo su Publicación

Si eligió la opción restringido o NO autoriza su publicación sírvase justificar:

---

---

ESTUDIANTES DE PREGRADO: TRABAJO DE INVESTIGACIÓN

TESIS  TESIS DOCTORADO

ESTUDIANTES DE POSGRADO : TESIS MAESTRÍA

DOCENTES: INFORME DE INVESTIGACIÓN

El equipo investigador Integrado por:

Nº	APELLIDOS Y NOMBRES	FACULTAD	CONDICIÓN DOCENTE (NOMBRADO, CONTRATADO, EMÉRITO, estudiante, OTROS)	CÓDIGO Docente Numero Matrícula del estudiante	Autor Coautor Asesor
01	Contreras Pumamango Luis Anthony	FCFYM		052701510	Autor
02	Sánchez Rodríguez Paul Denis	FCFYM		012701110	Autor

FIRMA

70493427

DNI

FIRMA

46795930

DNI

FIRMA

DNI