

Facultad de Energía, las Industrias y los Recursos Naturales no Renovables

Carrera de Computación

Prototipo de Software antiplagio para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.

Anti-plagiarism control software prototype for the Computer Engineering career at the National University of Loja, using neural networks.

Linea de investigación: Sistemas Inteligentes

PROYECTO DE INVESTIGACIÓN DE INTEGRACIÓN CURRICULAR.

Autor:

- ♦ ORCID, Jaime O. Paqui-Medina, jaime.paqui@unl.edu.ec
- ♦ ORCID, Santiago Alexander Roman Silva, santiago.roman@unl.edu.ec

Tutor:

• Pablo F. Ordoñez-Ordoñez, Mg.Sc.

Cotutor especialista:

• Oscar Cumbicus



Certificación de Tutoría

En calidad de Tutor y Cotutor del Proyecto de Trabajo de Titulación PTT, certificamos la tutela a Jaime O. Paqui-Medina, con el tema - **Prototipo de Software antiplagio** para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.

Anti-plagiarism control software prototype for the Computer Engineering career at the National University of Loja, using neural networks., quien ha cumplido con todas las observaciones requeridas. Es todo cuanto puedo decir en honor a la verdad, facultando al interesado hacer uso de la presente, así como el trámite de pertinencia del presente proyecto.

Loja, 24 de agosto de 2022

Atentamente, Pablo F. Ordoñez-Ordoñez, Mg.Sc.

TUTOR

OSCAR digitalmente por OSCAR MIGUEL CUMBICUS CUMBICUS Fecha: S PINEDA 2022.08.24 10:06:44-05'00' OSCAR Cumbicus

COTUTOR

Certificación de Autoría del Proyecto

Yo Jaime O. Paqui-Medina, y Santiago Alexander Roman Silva, estudiantes de la Universidad Nacional de Loja, declaramos en forma libre y voluntaria que el presente Proyecto de Trabajo de Titulación que versa sobre - **Prototipo de Software antiplagio para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.**

Anti-plagiarism control software prototype for the Computer Engineering career at the National University of Loja, using neural networks., así como la expresiones vertidas en la misma son autoría del compareciente, quien ha realizado en base a recopilación bibliográfica primaria y secundaria. En consecuencia asumo la responsabilidad de la originalidad de la misma y el cuidado al remitirse a las fuentes bibliográficas respectivas para fundamentar el contenido expuesto.

Atentamente,

Jaime O. Paqui-Medina

Atentamente,

San Pull

Santiago Alexander Roman Silva

Índice general

1.	Problemática Probl	1
	1.1. Situación Problemática	1
	1.2. Problema de Investigación	3
2.	<u>Justificación</u>	4
3.	Objetivos	6
J.	3.1. General	6
	3.2. Específicos	6
	5.2. Especificos	U
4.	Marco Teórico	7
	4.1. Plagio	7
	4.2. Formulario de Google	7
	4.3. Diagrama BPMN	7
	4.3.1. Características de BPMN	7
	4.3.2. Modelación de procesos	8
	4.3.3. Elementos de los diagramas	8
	4.4. Python	8
	4.5. Django	9
	4.5.1. El modelo	9
	4.5.2. La vista en Django (Template)	9
	4.5.3. La plantilla	10
	4.6. Modelo 4+1	10
	4.7. Redes neuronales	10
	4.8. Visual Studio Code	11
	4.9. scikit-learn	11
	4.10. Trabajos Relacionados	12
	4.10.1. Creación de un prototipo de sistema de detección de plagio con	- 1 0
	NLTK.	12
	4.10.2. Diseño e implementación de un sistema prototipo para detección	
	de plagio en imágenes y un módulo de análisis de similitud de	10
	documentos académicos	12
	4.10.3. Software para Detección automática de similitud en Documentos de Investigación en la Universidad Nacional del Altiplano Puno	
	2016	13
	<u> </u>	то

iv Índice general

4.10.4. Diseño e implementación de una plataforma genérica para desarrolla	ır
y probar nuevas técnicas de detección de plagio en textos	13
4.10.5. Diseño e Implementación de una herramienta en Studium para	
evitar los plagios en asignaturas de Ingeniería	13
5. Metodología	14
6. Cronograma	15
7. Presupuesto	16
7.1. Análisis de Puntos de Función	16
7.2. Cálculo de Puntos de Función	17
7.3. Cálculo del Esfuerzo	17
7.4. Cálculo del Presupuesto	18
Bibliografía	20
Lista de Acrónimos y Abreviaturas	22
A. Anexo I	23
B. Anexo II	27
C. Anexo III	33
D. Anexo IV	35

Índice de figuras

A.1. Pregunta 1.	23
A.2. Pregunta 2	24
A.3. Pregunta 3	24
A.4. Pregunta 4	25
A.5. Pregunta 5	25
A.6. Pregunta 6	26
B.1. Correo de los encuestados.	27
B.2. Lista de nombres de los encuestados.	28
B.3. Resultado acerca del conocimiento de plagio y herramienta que más	
conocen los encuestados.	28
B.4. Resultado sobre el beneficio de detectar plagio.	29
B.5. Resultado sobre la importancia de la utilidad del software anti-plagio.	30
B.6. Resultado sobre la importancia de utilizar software anti-plagio propio	
de la carrera	31
B.7. Resultado del Por qué es importante utilizar software anti-plagio propio	
de la carrera.	32
C.1. Tabla de las encuestas realizadas y sus respuestas	33
C.2. Tabla de las encuestas realizadas y sus respuestas	34
	25
D.1. Malla Curricular de la carrea Ingeniería en Ciencias de la Computación.	35

Índice de tablas

5.1.	Metodología del PTT	14
7.1.	Posibles Requisitos del Sistema	16
7.2.	Número de puntos de función sin ajustar	17
7.3.	Numero de puntos de función con ajuste	18

Índice de tablas vii

Prototipo de Software antiplagio para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.

Anti-plagiarism control software prototype for the Computer Engineering career at the National University of Loja, using neural networks.

Linea de investigación: Sistemas Inteligentes

1. Problemática

1.1. Situación Problemática

Entre las herramientas que pueden ayudar a mejorar la probidad académica de los estudiantes, aún desde los años pre-universitarios, se encuentran los software o herramientas "antiplagio". No obstante, las principales herramientas antiplagio son comerciales y necesitan de un pago anticipado, estas son: (Turnitin Urkund Gramman iThenticate entre otras como ejemplos), reduciendo así a las personas que tienen acceso a estas herramientas para detectar y corregir el problema del plagio académico.

En el Ecuador se realizó un análisis para evaluar el desempeño de algunas herramientas de acceso gratuito a través de internet. Se diseñó una búsqueda simple en Google con las palabras "herramientas antiplagio gratis" y "free antiplagiarism tools". El análisis mostró tres herramientas gratuitas con un desempeño aceptable para detectar copias de textos procedentes de internet: Small Seo Tools, DupliChecker.com y Search Engine Reports, siendo la primera la de mejor desempeño en condiciones reales.

Juan Fernández, director de la carrera de Informática de la Universidad Estatal de Guayaquil de la Región Costa, dice que estos programas son necesarios. Indica que ellos trabajan en un 90 por ciento con la herramienta Urkund, plataforma pagada, para analizar los trabajos de titulación: Donde permiten hasta un 10 por ciento de coincidencias, debido a errores al momento de citar algo, pero si es mayor se le pide al estudiante que realice las respectivas correcciones [2]. La Universidad Nacional de Loja no es la excepción a estos casos pues muchos estudiantes y docentes no siempre tienen acceso a estas herramientas por lo que deciden no aplicarlas e ignorar el posible plagio que pueda existir en los documentos académicos, generando así que los estudiantes se lleguen a acostumbrar a aplicar estas prácticas no éticas que a la larga significa una deficiencia en el aprendizaje del estudiante en cuestión. B

Seguidamente, en la UNL mediante la encuesta a 10 docentes de la carrera de computación, se obtuvo resultado de 8 individuos. B.2 Se indica que el 100 por ciento de

¹https://www.ayudauniversitaria.com/turnitin-gratis/

²https://www.urkund.com/es/el-sistema-urkund/

³https://www.grammarly.com/

⁴https://www.ithenticate.com/

2 Problemática

los docentes encuestados manifiestan que sería muy beneficioso que se cree un software anti-plagio de la carrera el cual sea gratuito, por las siguientes razones: B.4

- Garantiza que se respete la propiedad intelectual del autor. Además facilita la revisión de tareas y trabajos de titulación, ya que garantiza que son de elaboración propia de los estudiantes.
- Evitar el exceso de plagio en los trabajos de titulación.
- Mejorar la calidad de los trabajos de docentes y alumnos tanto en la parte de investigación como en la protección de derechos de autor, estas herramientas también permitirían evitar problemas legales dentro de este contexto. B.6
- Los software que detectan el plagio con resultados aceptables en el análisis son pagados.
- La mayoría de sistemas anti-plagio tienen un costo elevado, lo que dificulta la adquisición de los mismos.
- Es necesario, ya que no se dispone de este tipo de herramienta en la carrera.

Posteriormente, se pregunto acerca de las herramientas conocidas por parte de los docentes encuestados, donde se obtuvo los siguientes resultados: B.3

- Urkund: Conocida por cuatro docentes, dando un porcentaje del 50 por ciento de los encuestados.
- PlagScan: Conocida por cinco docentes, dando un porcentaje del 62,5 por ciento de los encuestados.
- DOCODE: Conocida por ningún docente, dando un porcentaje del 0 por ciento de los encuestados.
- Compilato: Conocida por ningún docente, dando un porcentaje del 0 por ciento de los encuestados.
- Turnitin: Conocida por dos docentes, dando un porcentaje del 25 por ciento de los encuestados.
- Strikeplagiarism: Conocida por ningún docentes, dando un porcentaje del 0 por ciento de los encuestados.
- Ninguno: Conocida por un docente, dando un porcentaje del 12,5 por ciento de los encuestados.

Concluyendo con los datos recolectados podemos entender que la herramienta más conocida por parte de los docentes es PlagScan y que además no todos los docentes utilizan las mismas herramientas para controlar el plagio, entonces teniendo en cuenta las respuestas obtenidas acerca de la importancia de utilizar software anti-plagio propio de la carrera, ellos manifiestan lo siguiente: B.7

- Contribución a la sociedad.
- Tener una propia herramienta, que sea fácil de usar y funcione correctamente.
- Que la herramienta sea accesible para la carrera, considerando de manera general todas las áreas de estudio.
- Tener un sistema personalizado con las necesidades que tiene la carrera, ya que siempre es una ventaja disponer de herramientas de este tipo.

Por otro lado, se indica que es importante que se utilice un software anti-plagio en la carrera, por calidad y ética, para verificar el plagio en tareas y/o actividades, porque esto garantiza que se respete la propiedad intelectual del autor, para mejorar la calidad de los trabajos de docentes y alumnos tanto en la parte de investigación como en la protección de derechos de autor, estas herramientas también permitirían evitar problemas legales dentro de este contexto. B.5

Aunque las herramientas anti-plagio son muy variadas y no todas estas herramientas son comerciales, las herramientas que están disponibles de forma gratuita no generan una correcta revisión del plagio dando como resultado una deficiencia que explica por qué no se utilizan estas herramientas normalmente. Una solución a este problema sería desarrollar un software que pueda ser utilizado por la institución que esté dirigida a la comunidad específica de estudiantes de la UNL haciendo que esto sea más eficiente. C

1.2. Problema de Investigación

¿Como se lograra controlar el plagio mediante el desarrollo de un software de control antiplagio para la carrera de ingeniería en Computación en la Universidad Nacional de Loja mediante el uso de redes neuronales?

2. Justificación

El poder implementar un sistema que permita detectar plagio, es con la finalidad de evitar que los estudiantes de la carrera de Ingeniería en Computación puedan tomar ideas, textos u otra información de otras personas e introducir en su trabajo como propio sin reconocer la autoría en ninguna de las formas existentes para hacerlo, es decir, sin usar los mecanismos de referencia.

De esta manera los docentes tendrían más control de toda la parte documentaría de los trabajos PIC, mediante el uso de esta herramienta que se pretende desarrollar, y detectar el típico copia y pega que realizan a menudo los estudiantes en sus trabajos, entonces con la implementación del software, se estaría ayudando a los estudiantes a mejorar su redacción, ya que esta constituye uno de los pilares principales para el desarrollo y evaluación de las habilidades de investigación y comunicación, las mismas que se consideran de vital importancia para promover el aprendizaje en los sistemas educativos actuales que se enfocan a una enseñanza más autodidacta y personalizada. [4]

Dentro de la Universidad Nacional de Loja, es necesario un software para controlar el plagio, que este sea gratuito y accesible para toda la carrera de Computación, ya que de esta forma se lograra garantizar que se respete la propiedad intelectual del autor, además facilitara la revisión de tareas y trabajos de titulación, ya que garantizara que son de elaboración propia de los estudiantes, y de esta manera mejorar la calidad de los trabajos de docentes y alumnos tanto en la parte de investigación como en la protección de derechos de autor, estas herramientas también permitirían evitar problemas legales dentro de este contexto.

Teniendo en cuenta los conocimientos adquiridos en las materias tales como:

' Programación Orientada a Objetos, Base de Datos, Diseño de Software, Desarrollo Basado en Plataformas, Teoría de Autómatas y Computabilidad Avanzada, Sistemas Distribuidos, Software Engineering Models, Software Engineering Management, Software Quality, Software Security, las mismas que fueron recibidas en todo el transcurso de la carrera de Ingeniería en Ciencias de la Computación, de la Facultad de energía, las industrias y los recursos naturales no renovables, pertenecientes a la UNL, consideramos que estamos en la capacidad de crear este software propuesto para la Carrera. D.1

Por lo tanto la implementación de este software apoyara a una mejora para la educación en general, contribución con la sociedad académica, ya que garantizara una

educación equitativa y de calidad y por ende promoverá oportunidades de aprendizaje ayudando a los estudiantes a mejorar su redacción en sus trabajos PIC. [5]

3. Objetivos

3.1. General

Desarrollar un prototipo de software de control antiplagio para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.

3.2. Específicos

- OB1 Diseñar el modelo de control antiplagio, mediante redes neuronales con Scikit-Learn.
- OB2 Diseñar el software antiplagio utilizando Django y aplicando del modelo de software 4+1.
- OB3 Evaluar la ejecución del software antiplagio, dentro de la carrera de Ingeniería en Computación, en los proyectos PIC.

4. Marco Teórico

4.1. Plagio

El plagio es una infracción de derechos de autor de cualquier trabajo producido al copiar ese trabajo sin el permiso de la persona que creó el trabajo o que posee o posee los derechos sobre el mismo. dicho trabajo y presentar ese trabajo como trabajo original, y esto incluye usar todo o parte del trabajo de otros.

4.2. Formulario de Google

Formularios de Google es un software de administración de encuestas que se incluye como parte del conjunto gratuito Google Docs Editors basado en la web que ofrece Google. Esta útil herramienta también nos permite planificar eventos, hacer preguntas a nuestros compañeros o amigos y así recopilar otro tipo de información de forma fácil y sencilla para lograr analizarla de forma sencilla.

4.3. Diagrama BPMN

Es el nuevo estándar para el modelado de procesos de negocio y servicios web. También es una notación a través de la cual se expresan los procesos de negocio en un diagrama de procesos de negocio (BPD). Este estándar agrupa la planificación y gestión del flujo de trabajo, así como el modelado y la arquitectura.

4.3.1. Características de BPMN

- Proporciona un lenguaje gráfico común, con el fin de facilitar su comprensión a los usuarios de negocios.
- Integra las funciones empresariales.
- Utiliza una Arquitectura Orientada por Servicios (SOA), con el objetivo de adaptarse rápidamente a los cambios y oportunidades del negocio.
- Combina las capacidades del software y la experiencia de negocio para optimizar los procesos y facilitar la innovación del negocio.

8 Marco Teórico

4.3.2. Modelación de procesos

Es la captura de una secuencia de actividades de negocio, y de la información de soporte.Los procesos de negocio describen la manera cómo una empresa alcanza sus objetivos. A continuación se tiene los siguientes niveles de proceso de modelado.

- Mapas de proceso: Son diagramas de flujo simple de las actividades.
- Descripciones de proceso: Conforman una extensión del anterior, y manejan información adicional pero no suficiente para definir completamente el ejecución actual.
- Modelos de proceso: Son diagramas de flujo extendido con suficiente información para que el proceso pueda ser analizado, simulado, y/o ejecutado.

4.3.3. Elementos de los diagramas

La función del BPMN es crear un mecanismo simple para realizar modelos de procesos de negocio, con todos sus elementos gráficos, y que al mismo tiempo sea posible gestionar la complejidad.Las cuatro categorías básicas de elementos son:

- Objetos de flujo: Un diagrama de procesos de negocio está compuesto por tres elementos básicos, que son los objetos de flujo. De esta forma, los modeladores no tienen que reconocer un gran número de formas diferentes. Los tres objetos de flujo son: Eventos, Actividad, Gateway.
- Objetos conectores: Conectan los objetos de flujo de un proceso, y definen el orden de ejecución de las actividades. Los tipos de conectores son: Secuencia, Mensaje y Asociación.
- Swimlanes (canales):Son un mecanismo empleado para organizar actividades en categorías separadas visualmente, con el fin de ilustrar diferentes capacidades funcionales o responsabilidades.
- Artefactos: Son objetos gráficos que proveen información adicional de los elementos dentro de un proceso, sin afectar el flujo del proceso. Se tiene actualmente dos tipo de artefactos predefinidos: Grupos y Anotaciones.

4.4. Python

Python es un lenguaje de programación de alto nivel, orientado a objetos, con una semántica dinámica integrada, principalmente para el desarrollo web y de aplicaciones informáticas. Es muy atractivo en el campo del Desarrollo Rápido de Aplicaciones (RAD) porque ofrece tipificación dinámica y opciones de encuadernación dinámicas,

4.5. Django

uno de los beneficios más importantes de Python es que tanto la librería estándar como el intérprete están disponibles gratuitamente, tanto en forma binaria como en forma de fuente. A diferencia de otros lenguajes como Java o .NET, se trata de un lenguaje interpretado, es decir, que no es necesario compilarlo para ejecutar las aplicaciones escritas en Python, sino que se ejecutan directamente por el ordenador utilizando un programa denominado interpretador, por lo que no es necesario "traducirlo" a lenguaje máquina.

Python cuenta con una amplia bibliotecas de librerías que permiten obtener diversos recursos de código abierto aplicables para inteligencia artificial. Python posee una amplia cantidad de recursos que permiten el aprendizaje de una inteligencia artificial.

4.5. Django

Django nace como un proyecto para publicación de noticias de Lawrence Journal-World(Revista de publicación de noticias), lo interesante de Django es que desde un principio fue construido como una herramienta para resolver problemas reales en un entorno empresarial, fue diseñado para optimizar el tiempo de desarrollo y los requerimientos exigentes de los desarrolladores web. Django se suele llamar un framework MVC, y con razón. Es muy fuerte mente influenciado por MVC y es incluso posible argumentar que la terminología MVC es el único lugar patrón de cambios en Django. Sin embargo más allá de las clasificaciones que podrían existir, está el entender cómo funciona realmente, al entenderlo se puede llegar a dominarlo. Esto se debe a que los desarrolladores no tuvieron la intención de seguir algún patrón de desarrollo, sino hacer el framework lo más funcional posible. En Django, las tres capas básicas son el modelo, la vista, y la plantilla. [10]

4.5.1. El modelo

El modelo define los datos almacenados, se encuentra en forma de clases de Python, cada tipo de dato que debe ser almacenado se encuentra en una variable con ciertos parámetros, que posee también métodos.

4.5.2. La vista en Django (Template)

La vista se presenta en forma de funciones en Python, su propósito es determinar qué datos serán visualizados, entre otras cosas más. El ORM de Django permite escribir código Python en lugar de SQL para hacer las consultas que necesita la vista. La vista también se encarga de tareas conocidas como el envío de correo electrónico, la autenticación con servicios externos y la validación de datos a través de formularios. [10]

10 Marco Teórico

4.5.3. La plantilla

La plantilla es básicamente una página HTML con algunas etiquetas extras propias de Django, en sí no solamente crea contenido en HTML (también XML, CSS, Javascript, CSV, etc).

Python - Django tiene unas propiedades únicas, como ser el desarrollo ágil, ya que con esta herramienta podemos construir un sistema web en cuestión de días, es una herramienta de tercera generación, trabaja en base a un modelo de desarrollo MTV, es una herramienta para profesionales perfeccionistas que buscan la funcionalidad en un 100 por ciento.

4.6. Modelo 4+1

Que se utiliza para describir la arquitectura de un sistema software intensivo basado en el uso de múltiples puntos de vista. Las vistas suelen describir el sistema desde el punto de vista de diferentes interesados, tales como usuarios finales, desarrolladores o directores de proyecto. Las cuatro vistas del modelo son: vista lógica, vista de desarrollo, vista de proceso y vista física. Además, una selección de casos de uso o escenarios suele utilizarse para ilustrar la arquitectura sirviendo como una vista más.

- Vista lógica: La vista lógica está enfocada en describir la estructura y funcionalidad del sistema.
- Vista de desarrollo: La vista de desarrollo ilustra el sistema desde la perspectiva del programador y está enfocado en la administración de los artefactos de software.
- Vista de proceso: La vista de proceso trata los aspectos dinámicos del sistema, explica los procesos de sistema y cómo se comunican. Se enfoca en el comportamiento del sistema en tiempo de ejecución.
- Vista física: La vista física describe el sistema desde el punto de vista de un ingeniero de sistemas. Está relacionada con la topología de componentes de software en la capa física, así como las conexiones físicas entre estos componentes.
- Escenarios: La descripción de la arquitectura se ilustra utilizando un conjunto de casos de uso, o escenarios lo que genera una quinta vista.

4.7. Redes neuronales

Una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la información: Funciona simultaneando un número elevado de unidades de procesamiento interconectadas que parecen versiones abstractas de neuronas.

Las unidades de procesamiento se organizan en capas. Hay tres partes normalmente en una red neuronal : una capa de entrada, con unidades que representan los campos de entrada; una o varias capas ocultas; y una capa de salida, con una unidad o unidades que representa el campo o los campos de destino. Las unidades se conectan con ponderaciones.

Al principio, todas las ponderaciones son aleatorias y las respuestas que resultan de la red son, posiblemente, disparatadas. La red aprende a través del entrenamiento. Continuamente se presentan a la red ejemplos para los que se conoce el resultado, y las respuestas que proporciona se comparan con los resultados conocidos. La información procedente de esta comparación se pasa hacia atrás a través de la red, cambiando las ponderaciones gradualmente.

4.8. Visual Studio Code

Visual Studio Code ofrece a los desarrolladores una nueva opción de herramienta de desarrollo que combina la simplicidad y experiencia optimizada de un editor de código con lo mejor que necesitan los desarrolladores para su ciclo central de código-construcción-depuración. Visual Studio Code es el primer editor de código y el primero multiplataforma, herramienta de desarrollo, compatible con OS X, Linux y Windows, en la familia Visual Studio. En esencia, Visual Studio Code (VS Code) cuenta con un potente y rápido editor de código ideal para el día a día. La versión Beta de Code ya tiene muchas de las funciones que los desarrolladores necesitan en un código y texto editor, incluida la navegación, soporte de teclado con enlaces personalizables, resaltado de sintaxis, coincidencia de corchetes, sangría automática y fragmentos, con soporte para docenas de idiomas. [13]

4.9. scikit-learn

La librería scikit-learn, también llamada sklearn, es un conjunto de rutinas escritas en Python para hacer análisis predictivo, que incluyen clasificadores, algoritmos de clusterización, etc. Está basada en NumPy, SciPy y matplotlib, de forma que es fácil aprovechar el código que usan estas librerías. [14]

incorpora varias funciones para preprocesar los datos:

- Normalización: ajustar las variables numéricas para que tengan media 0 y varianza 1, o bien que estén en un rango
- Transformaciones no lineales: basadas en cuartiles y exponentes, para transformar variables con distribuciones muy sesgadas

12 Marco Teórico

• Discretización: se trata de convertir una variable numérica en un conjunto de valores posibles, según algún criterio.

- Valores perdidos: cuando por algunos registros faltan datos de alguna variable, es posible imputar un valor en función de algún criterio automatizable, por ejemplo, sustituirlo por la mediana.
- Creación de interacciones entre variables mediante el uso de polinomios.

4.10. Trabajos Relacionados

4.10.1. Creación de un prototipo de sistema de detección de plagio con NLTK.

El objetivo de este trabajo es fundamentalmente realizar un prototipo de sistema de detección de plagio en base a los conocimientos adquiridos durante el grado y durante la etapa de documentación. Partimos de la hipótesis de que con los conocimientos del grado más la investigación propia de un trabajo de estas características es posible realizar un prototipo que permita detectar un determinado tipo de plagio y aportar conocimiento a lo que se ha investigado previamente para esta tarea. Como puntos concretos a trabajar para alcanzar este objetivo general, hemos pensado en investigar lo que se ha dicho y hecho para el fin de la detección de plagio, así como también probar las diferentes tecnologías que hay disponibles, desarrollar capacidades de programación, y pensar en un algoritmo de detección propio.

4.10.2. Diseño e implementación de un sistema prototipo para detección de plagio en imágenes y un módulo de análisis de similitud de documentos académicos.

El propósito de este tema es el poder tratar y buscar una solución a las implicaciones éticas y educativas al momento de realizar o manejar información para las distintas presentaciones académicas ya sea por parte del docente o estudiante. Para este tema bastante amplio y complejo podemos partir como una de las partes fundamentales como es la cultura de la ciudadanía, esto lleva que tengamos esta costumbre desde niños como es en la enseñanza de escuelas y posteriormente en colegio y universidades haciendo que esta acción sea habitual sin tener en cuenta el daña causado.

4.10.3. Software para Detección automática de similitud en Documentos de Investigación en la Universidad Nacional del Altiplano Puno 2016.

El trabajo de investigación es una propuesta de un Software para la verificación y detección de manera automática de la similitud en los trabajos de investigación en la Universidad Nacional del Altiplano de Puno. Los procesos para la detección de la similitud se divide en tres partes, en primer lugar se realiza las tareas de pre-análisis, que consiste en la conversión de los documentos en los formatos PDF, DOC a un archivo de texto plano. La segunda parte consiste en un análisis a priori, en la que utilizamos algoritmos no tan costosos para determinar si el documento en cuestión es posiblemente plagiado. Finalmente se tiene el análisis riguroso que tienen como entrada todos los documentos que pasaron la primera fase y a través de estos procesos se va a determinar cuál es el nivel de similitud del documento solicitado para el análisis.

4.10.4. Diseño e implementación de una plataforma genérica para desarrollar y probar nuevas técnicas de detección de plagio en textos.

En este trabajo se presenta un modelo de plataforma de software para desarrollar y evaluar los algoritmos de detección de plagio. La plataforma se basa en un diseño modular escalable, que implementa un conjunto de servicios que posibiliten realizar automáticamente tareas como: análisis sintáctico y semántico a través de WordNet y Freeling, extracción automática de texto de múltiples formatos de archivos (PDF, Word y texto), extracción de contenido de páginas web (el uso de algunos motores de búsqueda como Google, Yandex, Yahoo, Bing), el almacenamiento, la carga y el uso de algoritmos de detección de plagio. 16

4.10.5. Diseño e Implementación de una herramienta en Studium para evitar los plagios en asignaturas de Ingeniería

El proyecto de innovación docente que se pretende justificar consiste en implementar una solución válida para todos los lenguajes de programación de las asignaturas de grado, con el objetivo de que se integre y se desarrolle una solución basada en Studium. El profesor al crear una tarea o entrega activará el sistema anticopia, de una forma sencilla. La solución estará basada en software libre con el objetivo de minimizar los costes. Los alumnos siempre y cuando quiera el profesor podrán obtener un informe sobre el análisis del sistema anticopia, de forma que podrían visualizar la similitud de su ejercicio respecto a sus compañeros. Esta funcionalidad será opcional y únicamente estaría disponible después del envío de la tarea.

5. Metodología

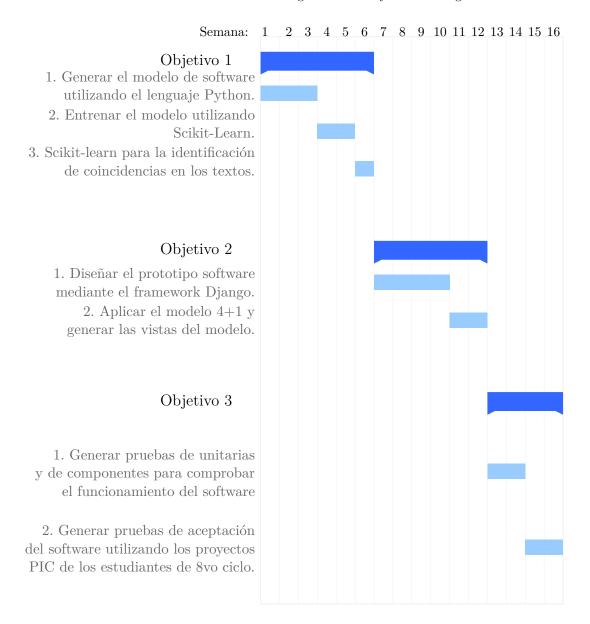
Tabla 5.1: Metodología del PTT

OBJ	OBJETIVOS		ALCANCE	PRODUCTO	MÉTODOS	MATER./TECN.	LUGAR	RESPONSABLE
OB1	3.2		Diseñar el modelo mediante el lenguaje Python que marcara el marco de trabajo para el software anti plagio. Utilizar redes neuronales con Scikit-Learn para identificar coincidencias dentro los textos.	Modelo.	Metodología ICONIX	Scikit-Learn Redes Neuronales	Lab de software	Santiago R. Jaime P.
ОВ2	3.2		Generar el prototipo necesario para el sistema web utilizando el framework Django Generar: Vista Lógica, Vista de Despliegue, Vista de Procesos y Vista Física.	Prototipo de software	Modelo 4+1. Lenguaje de Programación Python.	Visual Studio Code Python	UNL(Carrera de computación)	Santiago R. Jaime P.
ОВ3	3.2		Generar pruebas unitarias, de componentes y aceptación para asegurarse que el software funcione correctamente. Probar el software anti plagio para los Proyectos de Integración Curricular de octavo ciclo de la carrera de Ingeniería en computación.	Pruebas funcionales	Pruebas unitarias. Pruebas de componentes. Prueba de aceptación.	Proyectos PIC	UNL(Carrera de computación)	Santiago R. Jaime P.

6. Cronograma

Cronograma del Proyecto de Integración Curricular

Mes 1



7. Presupuesto

7.1. Análisis de Puntos de Función

Para el presente proyecto se utilizará el análisis de puntos de función para estimar el presupuesto, el cual en si es su unidad de medida. Por lo tanto, con este se pretende representar cantidad; en este caso, la medición es independiente de la tecnología utilizada para desarrollar el software.

Para el proceso para medir o evaluar el desempeño funcional con los puntos de función se hace uso de los requisitos funcionales que el usuario solicitó, como se describe en los artefactos de desarrollo.

En la figura [7.1], se muestra los posibles requisitos funcionales que se pueden presentar en el sistema para poder hacer el cálculo estimado del presupuesto del proyecto, ya que aún no se hace un levantamiento real de los requisitos.

A partir de los requisitos funcionales se procede a hacer el cálculo de los puntos de función si ajustar como se muestra en la figura 7.2, y los puntos de función con ajuste como se muestra en la figura 7.3.

Tabla 7.1: Posibles Requisitos del Sistema

Referencia	Nombre del Requisito	Tipo Operación	Peso
RF01	Registrarse y loguearse a la aplicación	EI	4
RF02	Subir documentos a la aplicación(Create, Update)	EI	4
RF03	Obtener la base de datos con los documentos	EI	4
RF04	Buscar documentos	EQ	4
RF05	Generar el porcentaje total de plagio	EI	4
RF06	Eliminar documentos(Delete)	EI	4
RF07	Listado de documentos(Read)	EO	5
RF08	Reporte de los documentos registrados	EO	5
RF09	1 Reporte de plagio	EO	5
	4 Tablas BD	ILF	40

Tipo de función de usuario	Nivel de complejidad	N^{o}	*	Peso	=	Total
Entradas	Baja			3		
EI	Media	5		4		20
151	Alta			6		
Salidas	Baja			4		
EO	Media	3		5		15
EO	Alta			7		
Consultas	Baja			3		
EQ	Media	1		4		4
EQ	Alta			6		
Archivos	Baja			7		
ILF	Media	4		10		40
ILF	Alta			15		
Trak and a a a	Baja			5		
$egin{array}{c} ext{Interfaces} \ ext{EIF} \end{array}$	Media			7		
EIF	Alta			10		
Número de Pu	ntos Función sin ajusta	r:				79

Tabla 7.2: Número de puntos de función sin ajustar

7.2. Cálculo de Puntos de Función

Una vez obtenido los puntos de función sin ajustar y los puntos de función con ajuste, se procede hacer el cálculo total de lo puntos de función, con la siguiente fórmula:

$$PF = PUNTOSFUNCINSINAJUSTAR * (0.65 + 0.01 * TOTALGI)$$

$$= 79 * (0.65 + 0.01 * 37)$$

$$= 80.58$$
(7.1)

7.3. Cálculo del Esfuerzo

Para hacer el cálculo del esfuerzo o el cálculo de trabajo de un desarrollador se necesita los puntos de función y las horas de trabajo dependiendo de las lineas de código del lenguaje de programación.

$$H/Ht = PF * HORASPFPROMEDIO$$

$$= 80.58 * 8$$

$$= 645$$
(7.2)

18 Presupuesto

Grado Factores de Influencia en la Dificultad del Sistema (0-5)1. Comunicaciones de datos 4 2. Procesamiento distribuido 4 2 3. Objetivos de rendimiento 4. Configuración de uso intensivo 1 5. Tasas de transacción rápidas 1 6. Entrada de datos en línea 5 7. Amigabilidad en el diseño 3 3 8. Actualización de datos en línea 9. Procesamiento complejo 1 10. Reusabilidad 3 11. Facilidad de instalación 0 12. Facilidad operacional 3 2 13. Adaptabilidad (inst mult) 14. Versatilidad (cambios) 5 37 Total GI

Tabla 7.3: Numero de puntos de función con ajuste

$$DASDETRABAJO(DT) = H/5$$

$$= 645/5$$

$$= 129$$
(7.3)

$$MESESDETRABAJO(MT) = DT/20$$

$$= 129/20$$

$$= 6.45$$

$$(7.4)$$

7.4. Cálculo del Presupuesto

Una vez obtenido el cálculo de los puntos de función y el cálculo del esfuerzo, se procede a calcular la estimación del presupuesto. Para ello primeramente se debe definir los siguiente:

SUELDO MENSUAL PROGRAMADOR (JR.)=600 OTROS GASTO(SERVICIOS + MATERIALES + IMPREVISTOS) = 2400

Y por último se procede a realizar el cálculo del presupuesto:

$$PRESUPUESTO = (PROGRAMADORES * DM * SUELDO) + OTROSGASTOS$$

$$= (2*6.45*600) + 2400$$

$$= 10,140$$

$$(7.5)$$

Bibliografía

- [1] D. Díaz Arce, "Herramientas "antiplagio": ¿son confiables? Estudio de casos.," Edutec. Revista Electrónica de Tecnología Educativa, no. 61, p. a365, 2017.
- [2] N. E. Universo, "Universidades de ecuador utilizan programas informáticos para detectar el plagio intelectual," sep 2019.
- [3] O. Augusto and R. Ortega, "Creación de un prototipo de sistema de detección de plagio con NLTK," 2020.
- [4] J. P. Lozano Quito and J. E. Zhunio Zhunio, "Diseño e implementación de un sistema prototipo para detección de plagio en imágenes y un módulo de análisis de similitud de documentos académicos," Universidad Politécnica Salesiana, Facultad de Ingenierías, pp. 1–121, 2014.
- [5] M. Rieckmann, Educación para los Objetivos de Desarrollo Sostenible: objetivos de aprendizaje. UNESCO Publishing, 2017.
- [6] A. S. Zalazar, S. Gonnet, and H. Leone, "Un Modelo para Contratos de Cloud Computing," pp. 303–317, 2013.
- [7] I. Nizhenkivska, O. Kuznetsova, and V. Naroha, "Uso De Formularios De Google Para Comprobar Los Conocimientos De Los Estudiantes De Farmacia En Biología Y Química," Fundamentos De La Investigación Científica Moderna, pp. 1–4, 2021.
- [8] S. Analitica, "Manual de diagramación de procesos bajo estándar bpmn," Colombia: Analítica SAS, 2016.
- [9] I. Challenger Pérez, Y. Díaz Ricardo, and R. Becerra García, "El lenguaje de programación Python/The programming language Python," Revista Ciencias Holguín, vol. 20, pp. 1–13, 2014.
- [10] J. L. Condori Ayala, "Phython-djangoframework de desarrollo web para perfeccionistasbasado en el modelo mtv," Revista de información, tecnología y sociedad, p. 36, 2012.
- [11] P. Kruchten, "Planos arquitectónicos: El modelo de 4+ 1 vistas de la arquitectura del software," IEEE software, vol. 12, no. 6, pp. 42–50, 1995.

Bibliografía 21

[12] D. J. Matich, "Redes neuronales: Conceptos básicos y aplicaciones," Universidad Tecnológica Nacional, México, vol. 41, pp. 12–16, 2001.

- [13] T. Kahlert and K. Giza, "Visual Studio Code Code Editing. Redefined," Microsoft, vol. 1, no. March, pp. 1–26, 2016.
- [14] O. Kramer, "Scikit-learn," in Machine learning for evolution strategies, pp. 45–53, Springer, 2016.
- [15] E. A. Aliaga Payehuanca, "Software para detección automática de similitud en documentos de investigación en la universidad nacional del altiplano puno 2016," 2017.
- [16] M. F. Barrera Maura and N. H. Fajardo Heras, "Diseño e implementación de una plataforma genérica para desarrollar y probar nuevas técnicas de detección de plagio en textos," B.S. thesis, 2014.
- [17] G. Villarrubia González, J. F. De Paz, A. González Briones, V. F. López Batista, A. López Barriuso, A. B. Gil González, J. M. Corchado Rodríguez, M. Navarro Cáceres, P. I. Vega Cruz, Á. L. Sánchez Lázaro, et al., "Diseño e implementación de una herramienta en studium para evitar los plagios en asignaturas de ingeniería," 2018.

Lista de Acrónimos y Abreviaturas

TT Trabajo de Titulación.

A. Anexo I

Formulario con las respectivas preguntas, para realizar la encuesta correspondiente.

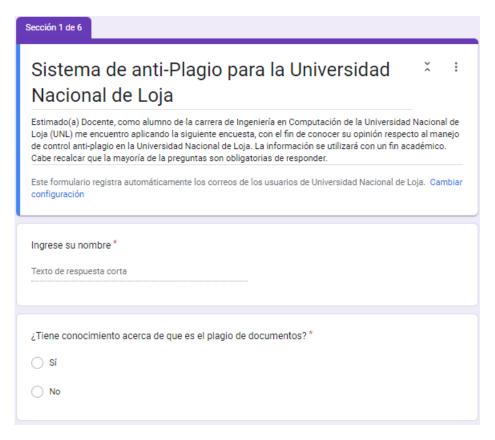


Figura A.1: Pregunta 1.

24 Anexo I

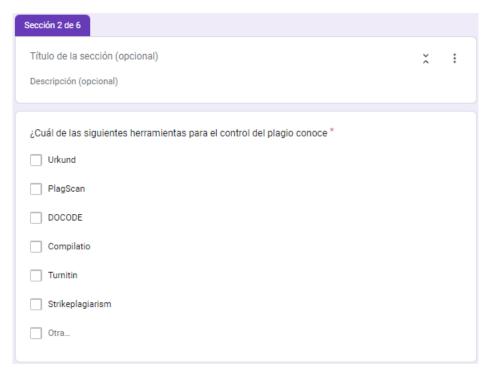


Figura A.2: Pregunta 2.



Figura A.3: Pregunta 3.

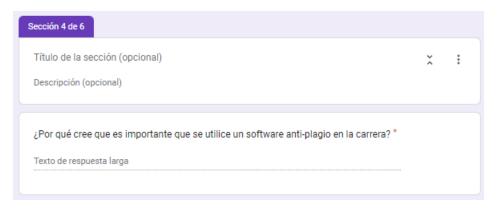


Figura A.4: Pregunta 4.

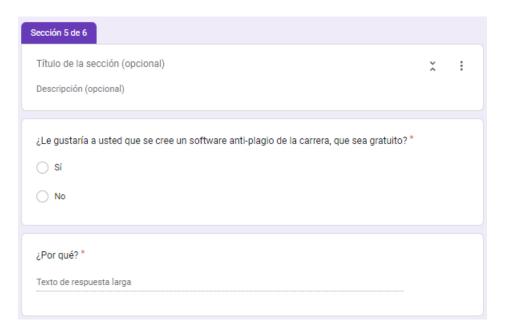


Figura A.5: Pregunta 5.

26 Anexo I

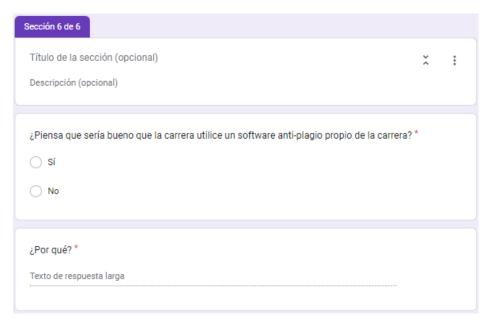


Figura A.6: Pregunta 6.

B. Anexo II

Respuestas obtenidas de los encuestados mediante el formulario aplicado del Anexo I.

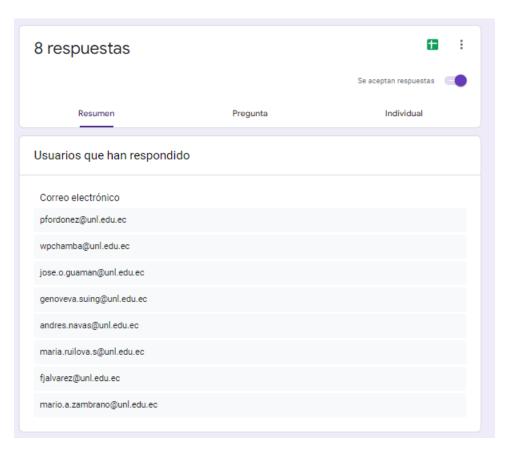


Figura B.1: Correo de los encuestados.

28 Anexo II



Figura B.2: Lista de nombres de los encuestados.

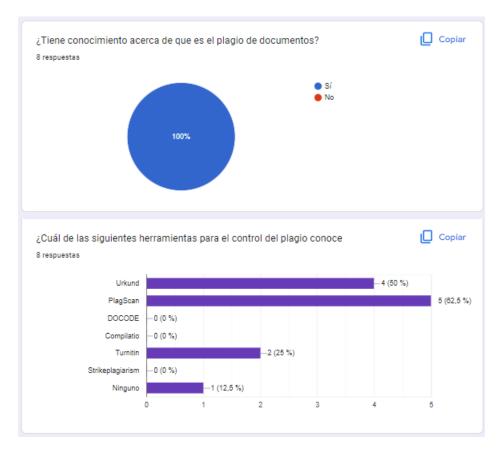


Figura B.3: Resultado acerca del conocimiento de plagio y herramienta que más conocen los encuestados.

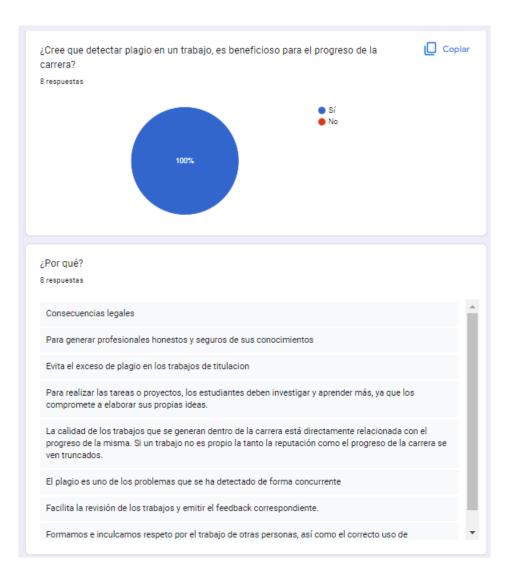


Figura B.4: Resultado sobre el beneficio de detectar plagio.

30 Anexo II

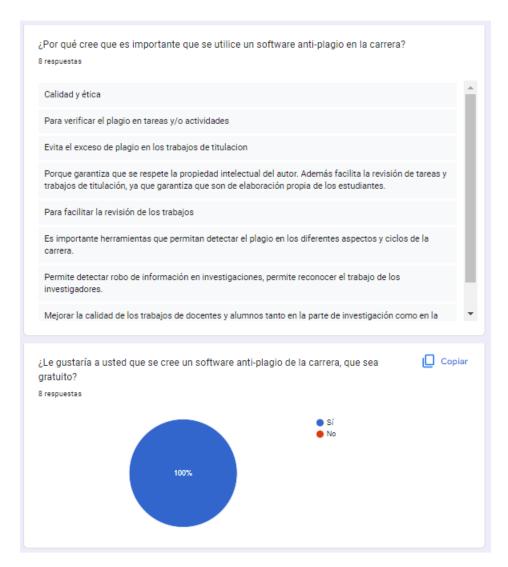


Figura B.5: Resultado sobre la importancia de la utilidad del software anti-plagio.

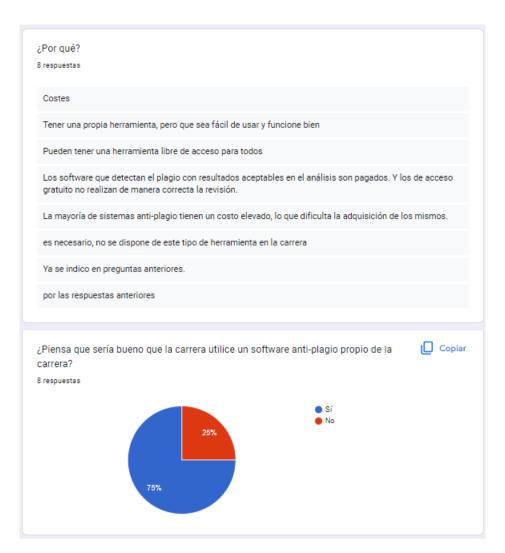


Figura B.6: Resultado sobre la importancia de utilizar software anti-plagio propio de la carrera

32 Anexo II



Figura B.7: Resultado del Por qué es importante utilizar software anti-plagio propio de la carrera.

C. Anexo III

Respuestas obtenidas de los encuestados mediante el formulario aplicado del Anexo I, reflejados en una tabla.

Dirección de correo electrónico	Ingrese su nombre	¿Tiene conocimiento acerca de que es el plagio de documentos?	¿Cuál de las siguientes herramientas para el control del plagio conoce		plagio en un trabajo, es progreso de la carrera?	¿Por qué cree que es importante que se utilice un software anti-plagio en la carrera?	¿Le gustaria a usted que se cree un software anti-plagio de la carrera, que sea gratuito?			
					¿Por qué?			¿Por qué?		¿Por qué?
pfordonez@unl.edu.ec	Pablo Ordoñez	Si	Urkund, PlagScan	Si	Consecuencias legales	Calidad y ética	Si	Costes	Si	Contribución a la sociedad académica
wpchamba@unl.edu.ec	Wilman Chamba	Si	Ninguno	Si	Para generar profesionales honestos y seguros de sus conocimientos	Para verificar el plagio en tareas y/o actividades	Si	Tener una propia herramienta, pero que sea fácil de usar y funcione bien	Si	Tener una propia herramienta, pero que sea fácil de usar y funcione bien
jose.o.guaman@unl.edu.	Jose Guaman	Si	Urkund	Sí	Evita el exceso de plagio en los trabajos de titulacion	Evita el exceso de plagio en los trabajos de titulacion	Sí	Pueden tener una herramienta libre de acceso para todos	No	Pueden tener una herramienta libre de acceso para todos
genoveva.suing@unl.edu	.Genoveva Suing Albito	sSi	PlagScan, Turnitin	Sí	Para realizar las tareas o proyectos, los estudiantes deben investigar y aprender más, ya que los compromete a elaborar sus propias ideas.	Porque garantiza que se respete la propiedad intelectual del autor. Además facilita la revisión de tareas y trabajos de litulación, ya que garantiza que son de elaboración propia de los estudiantes.	Sí	Los software que detectan el plagio con resultados aceptables en el análisis son pagados. Y los de aceptados gratuito no realizan de manera correcta la revisión.		Se podría considera de manera general todas la áreas de estudio.

Figura C.1: Tabla de las encuestas realizadas y sus respuestas.

34 Anexo III

					·	i e				a contract of the contract of
andres.navas@unl.edu.ed	Andrés Roberto Navas	s Sí	PlagScan	Sí	La calidad de los trabajos que se generan dentro de la carrera está directamente relacionada con el progreso de la misma. Si un trabajo no es propio la tanto la reputación como el progreso de la carrera se ven truncados.	Para facilitar la revisión de los trabajos		La mayoría de sistemas anti-plagio tienen un costo elevado, lo que dificulta la adquisición de los mismos.		Tendríamos un sistema personalizado con las necesidades de la carrera
maria.ruilova.s@unl.edu.e	María del Cisne Ruilov	v St	Urkund, PlagScan, Tu		problemas que se ha detectado de forma	Es importante herramientas que permitan detectar el plagio en los diferentes aspectos y ciclos de la carrera.	Sí	es necesario, no se dispone de este tipo de herramienta en la carrera	Sí	siempre es una ventaja disponer de herramientas de este tipo
pfordonez@unl.edu.ec	Pablo Ordoñez	Sí	Urkund, PlagScan	Sí	Consecuencias legales	Calidad y ética	Sí	Costes	Sí	Contribución a la sociedad académica
wpchamba@unl.edu.ec	Wilman Chamba	Sí	Ninguno		Para generar profesionales honestos y seguros de sus conocimientos	Para verificar el plagio en tareas y/o actividades		Tener una propia herramienta, pero que sea fácil de usar y funcione bien		Tener una propia herramienta, pero que sea fácil de usar y funcione bien

Figura C.2: Tabla de las encuestas realizadas y sus respuestas.

D. Anexo IV

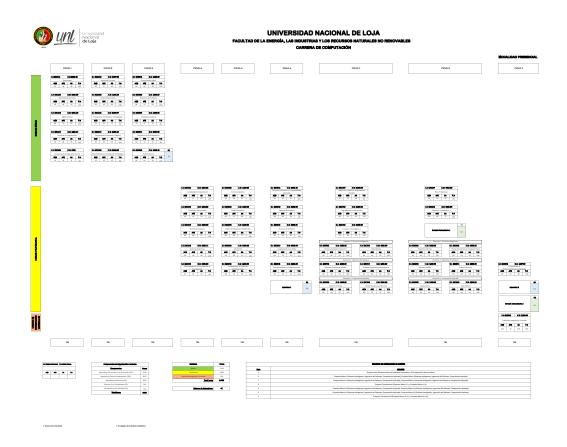


Figura D.1: Malla Curricular de la carrea Ingeniería en Ciencias de la Computación.

E. Anexo V. URL de Proyecto de Integración Curricular

https://es.overleaf.com/read/ntthcjjspjpt

F. Anexo VI

Cuadro tutorías



UNIVERSIDAD NACIONAL DE LOJA COMISIÓN DE ARTICULACIÓN DE LAS FUNCIONES SUSTANTIVAS FORMATO PARA EL REGISTRO DE LAS ACTIVIDADES DE TUTORÍA PARA LA ELABORACIÓN DEL TRABAJO DE INTEGRACIÓN CURRICULAR/TITULACIÓN

Facultad:	Facultad de la Energía las Industrias y los Recursos no Renovables
Carrera/Programa:	Carrera Computación
Nombre y apellidos del docente asesor/director:	Oscar Cumbicus
Tema del trabajo de integración curricular/litulación:	Prototipo de Software antiplagio para la carrera de Ingeniería en Computación de la Universidad Nacional de Loja, utilizando redes neuronales.
Nombres y apellidos del/los aspirante/s:	Jaime Paqui y Santiago Roman
Período académico ordinario:	Abril 2022 - Septiembre 2022

Nro.	Fecha	Tiempo empleado en la	Tema tratado en la tutoría	Recomendaciones del Asesor/Director	Modal	idad	Firma del estudiante / Registro Virtual	
NITO.	recna	tutoría (horas)	rema tratado en la tutoria	Recomendaciones dei Asesor/Director	Presencial	Virtual	rema dei estudiar	ite / Kegistro Virtual
1	28/6/2022	1	Reunión inicial para poner en contexto la propuesta	Generar encuestas Explicaciones adicionales	x		Sand State	30
2	1/7/2022	1	Revison de la encuesta	Corregir encuesta Explicaciones adicionales	x		Salar	
3	29/7/2022	1	Revisón de PIC	Corrección del Tema del PIC Corrección en las Objetivas del PIC Corrección de la Problemática		х	Sun Park	
4	9/8/2022	1	Revisón de PIC	Corrección de la Problemática (citor las entrevistas). Corrección en la Justificación del PIC (justificar académicamente, técnicamente y social). Agregar más anexos		х	San	
5	11/8/2022	1	Revisón de PIC	Corrección en los Objetivos del PIC Mejorar Metodología Corrección en el Cronograma del PIC Corrección de las faltas de Ortografía		х	Sussia	W.
6	8/16/2022	1	Revisón de PIC	Corrección del Presupuesto del PIC Corrección en Marco Teórico Corrección en la Bibliografía		х	Salar -	
7	8/22/2022	1	Revisión del PIC	Carrección en los Objetivos del PIC Carrección en el Alcance del PIC Revisión general de todo el PIC		x	Salara -	M
9	23/8/2022	1	Rivisión Final PIC	Corrección en el Presupuesto del PIC Corrección en la plantilla del PIC	x		Salar	W

				•			ALC:	34/4- F
3	29/7/2022	1	Revisón de PIC	Carrección del Tema del PIC Carrección en las Objetivos del PIC Carrección de la Problemática		x	Salar	W.
4	9/8/2022	1	Revisón de PIC	Corrección de la Problemática (citar las entrevistas) Corrección en la Justificación del PIC (justificar académicamente, técnicamente y social) Agregar más anexos		х	Sparker	
5	11/8/2022	1	Revisón de PIC	Corrección en los Objetivos del PIC Mejorar Metodología Corrección en el Cronograma del PIC Corrección de las faltas de Ortografía		х	5	
6	8/16/2022	1	Revisón de PIC	Corrección del Presupuesto del PIC Corrección en Marco Teórico Corrección en la Bibliografía		x	Sala -	
7	8/22/2022	1	Revisión del PIC	Carrección en las Objetivos del PIC Carrección en el Alcance del PIC Revisión general de todo el PIC		x	Salar -	W.
9	23/8/2022	1	Rivisión Final PIC	Carrección en el Presupuesto del PIC Carrección en la plantilla del PIC	x		Salar	AN.
Fecha de presentación: 8/24/2022								

Firma del docente

Firma del Director/a y/o Encargado de la Gestión Académica de la Carrera