

REGISTRO DE TRABAJO DE GRADO

FECHA

12

05

2023

DATOS DEL ESTUDIANTE (S)

NOMBRES: OMAIRA JISEL						APELLIDOS: OTAVO MADRIGAL	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO: 1022394477	
CORREO INSTITUCIONAL: ootavom@ucentral.edu.co						TELÉFONO: 3197117438	

NOMBRES: ANGELA MARCELA						APELLIDOS: VARGAS VARGAS	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO: 1030550600	
CORREO INSTITUCIONAL: avargasv@ucentral.edu.co						TELÉFONO: 3105771896	

NOMBRES: CESAR AUGUSTO						APELLIDOS: SUAREZ PEÑA	
TIPO IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO: 1032457689	
CORREO INSTITUCIONAL: csuarezp2@ucentral.edu.co						TELÉFONO: 3157871426	

MODALIDAD DE TRABAJO DE GRADO (Seleccione una opción)

I. Modalidad de investigación:	II. Modalidad de profundización: <input type="checkbox"/>
a. Proyecto final con resultados de nuevo conocimiento <input type="checkbox"/>	a. Trabajo monográfico <input type="checkbox"/>
	b. Pasantía nacional o internacional <input type="checkbox"/>

Línea de profundización

AVAL DEL DOCENTE DIRECTOR

NOMBRES: LUIS ANDRES CAMPOS MALDONADO	DEPARTAMENTO: INGENIERÍA Y CIENCIAS BÁSICAS
CORREO INSTITUCIONAL: lcampasm@ucentral.edu.co	TELÉFONO-EXT.: +573016627377

COMPONENTES

1. TÍTULO DEL TRABAJO DE GRADO O PASANTÍA

Entrenamiento de un modelo de procesamiento de lenguaje natural para búsquedas semánticas en el contexto jurídico de la Corte Constitucional.

2. INTRODUCCIÓN Y JUSTIFICACIÓN (máximo 1500 palabras)

“El fenómeno de la Legaltech, o tecnología aplicada a la comercialización y prestación de servicios legales, tiene ya un alcance mundial y casi imparable.” (Legaltechies, 2020), hay que reconocer que la analítica y la inteligencia artificial comienza a despertar gran interés en el ámbito jurídico, y es tal que los sistemas de administración de justicia en el mundo ven en estas herramientas una oportunidad para mejorar los procesos legales, así mismo las firmas de abogados las consideran un verdadero apoyo para apalancar las estrategias de defensa ante los juzgados y aumentar la probabilidad de ganancia de los casos y negocios.

El estado Colombiano a través del Ministerio de Tecnologías De La Información y Las Comunicaciones ha elaborado el documento Agenda Colombia Digital 2022-2026, la cual traza la ruta para articular la cultura digital con la transformación digital como motor de oportunidades e igualdad (Mintic, 2022),

La Corte Constitucional de Colombia es la entidad judicial encargada de velar por la integridad y la democracia de la Constitución, cimienta su misión en los estrictos términos enmarcados en los artículos 241 al 244 de la Constitución Política de Colombia y su visión en ser la entidad que mediante su jurisprudencia y ejercicio del poder público, sea reconocida nacional e internacionalmente. Su principal función es cuidar que se respete la constitución y los derechos humanos o fundamentales de las personas, y resolver los asuntos de tipo constitucionales, estudia las tutelas y que se puede convertir en una ley o por el contrario que no tiene los fundamentos para ser categorizado como tal.

Mientras el campo del derecho se basa en tanto en el lenguaje oral como en el escrito, y por otro lado, la ingeniería se basa en matemáticas, cada vez más se identifican necesidades y oportunidades para integrar algoritmos que permitan incorporar nuevos conceptos tanto jurídicos como tecnológicos que repercutan en el derecho y en la administración pública [1]. Surgen preguntas sobre el derecho, la justicia y la tecnología, tales como: ¿es la inteligencia artificial la oportunidad para construir un mundo mejor? ¿Puede la inteligencia artificial mejorar el derecho? ¿Es posible para un abogado tener un conocimiento profundo de las normas sin ser especializado? ¿Cuál es la importancia de la inteligencia artificial en el derecho? ¿Es labor del gobierno generar estrategias de inteligencia artificial? Preguntas que se generan debido al gran volumen de datos generado hoy en día y a medida

que el tamaño de los datos necesarios para ejecutar un proceso determinado empieza crecer y, además, a ser requerido por múltiples procesos a la vez.

Es así como en el presente proyecto se pretende lograr que los profesionales del derecho y el personal que labora en el sistema de administración de justicia, cuente con herramientas tecnológicas para la gestión de análisis y para desarrollar sus funciones con tiempos óptimos de búsqueda y respuesta, que eviten el vencimientos de términos en los procesos y beneficien a las partes que interviene en los mismos.

3. OBJETIVO GENERAL

Entrenar el modelo de lenguaje de procesamiento de lenguaje natural (NPL) llamado BERT en el contexto jurídico colombiano.

4. OBJETIVOS ESPECÍFICOS

Obtener las sentencias de la página de la Corte Constitucional por medio de una herramienta de web scraping desarrollada en Python.

Crear una opción que permita el análisis de tipologías o tópicos más frecuentes y relevantes en el tiempo en las sentencias emitidas por parte de la corte constitucional.

Desarrollar una interfaz de usuario de consulta y búsqueda para la jurisprudencia de la Corte Constitucional basado en el modelo entrenado BERT.

5. ANTECEDENTES (máximo 2000 palabras)

Con la creciente cantidad de información jurídica disponible, la necesidad de herramientas que puedan procesar y analizar grandes cantidades de texto se ha vuelto cada vez más importante. El procesamiento de lenguaje natural se ha convertido en una técnica popular para abordar este desafío, ya que permite buscar, clasificar y analizar grandes volúmenes de documentos legales en poco tiempo, ayudando a las áreas legales a automatizar tareas, identificar patrones, tendencias, mejorar la precisión y la eficiencia de los procesos. La adopción de la tecnología de PLN en el ámbito legal sigue aumentando, y se espera que continúe así a medida que se desarrollen nuevas aplicaciones y avances tecnológicos. En este sentido, nace la necesidad de aplicar este conocimiento en el contexto jurídico colombiano para empezar a apalancar las funciones de un sistema que por años se ha considerado lento dentro de la sociedad.

La revisión de antecedentes da inicio con un recorrido sobre los indicadores del sector jurisdiccional mostrando la realidad actualizada en relación con los procesos judiciales, además la Corporación Excelencia en la Justicia hace recomendaciones sobre los avances tecnológicos en los que se deberían apalancar las áreas, basado en estas sugerencias se consideran los avances del modelo BERT en el lenguaje español como una estrategia de apoyo a las áreas legales, teniendo en cuenta el alcance de este en relación con los análisis semánticos que está en

capacidad de realizar, también se evalúan publicaciones de carácter internacional que estudiaron a BERT como una estrategia para la solución de problemas de las áreas legales en relación con el procesamiento de lenguaje natural.

5.1 Índice de Congestión de la Rama Judicial en Colombia (Sector Jurisdiccional)

La Corporación excelencia en la Justicia, tenía como objetivo general de este proyecto estudiar modelos de Inteligencia Artificial basados en NLP para proponer estrategias de sentencia y herramientas de soporte a la actividad judicial en Colombia. La metodología se basó en incluir el uso de datos públicos recolectados en temas judiciales para entrenar modelos de aprendizaje automático y, posteriormente, evaluar su desempeño en la predicción de sentencias. La problemática identificada incluyó la congestión de procesos judiciales y la corrupción en la administración de justicia colombiana, lo que convierte al sistema judicial en una oportunidad de mejora urgente para el desarrollo del país.

En conclusión se pretendía dar a conocer estrategias para ayudar a mejorar la eficiencia y rapidez en la toma de decisiones judiciales, para así contribuir con la reducción de la congestión y la corrupción en el sistema judicial colombiano. Además, se destaca que la tecnología y la inteligencia artificial pueden ayudar en la interpretación de hechos y reglas aplicables, así como la identificación de posibles consecuencias. Esto puede ser una herramienta valiosa para los jueces en la estructuración de razonamientos para resolver casos.

5.2. Intersections between Law and Artificial Intelligence

La inteligencia artificial y la ley se cruzan en muchos niveles, porque esta influye no solo en la práctica legal al hacer que los abogados sean más eficientes en su trabajo o al automatizar algunos servicios legales, sino también en la ley misma. Se desafían los conceptos legales tradicionales, por lo que la ley debe adaptarse. Esta adaptación deberá continuar y corresponder a los nuevos desarrollos en la materia. Al mismo tiempo, la ley también dará forma a los desarrollos en IA al establecer nuevos estándares, pautas y limitaciones en los desarrollos en varios dominios de aplicaciones de IA. Será necesario realizar una extensa investigación legal para determinar las implicaciones sociales de implementar IA y robots en nuestra vida cotidiana. Por ejemplo, pronto se necesitará una evaluación de si la IA y los sistemas ciberfísicos deberían regularse de manera diferente, así como un análisis de cómo debería reaccionar la ley en caso de que la IA provoque un desempleo masivo.

5.3. La inteligencia artificial y su aplicación al campo del Derecho

La siguiente referencia, se extrajo de la revista Alegatos de la universidad autónoma metropolitana de México, cuyo objetivo es analizar el aporte y la función que trae la inteligencia artificial al derecho, así como explicar cuáles son los elementos necesarios para crear sistemas expertos jurídicos y la forma en la que procesan la información.

Por lo tanto es importante destacar que los sistemas expertos jurídicos tienen como principio emular el comportamiento del ser humano. Esto es posible gracias a diversas fuentes de entrada informativa al programa sobre cierta materia, algunos ejemplos son bases de datos que construyan la contextualización y la finalidad del sistema experto, así como, la información del usuario que es otorgada por medio de preguntas, lo que permite identificar un tópico o objeto y relacionarlo a las posibles respuestas.

Para lograr esto, se definió que la metodología usada para crear estos sistemas, estaba compuesta por 5 fases.

- Definición de un dominio o subdominio sobre el cual el versará

- Definición de los problemas que resolverá
- Identificación del grupo de expertos
- Adquisición del conocimiento
- Elaboración del prototipo

Siendo usada en cualquiera de las cuatro clases diferentes de sistemas expertos en el ámbito jurídico que menciona la autora : aquellos que reemplazan al experto, los que identifican el problema, los que ayudan a identificar el problema y los que asisten en la toma de decisiones. Este último fue el enfoque principal de su trabajo.

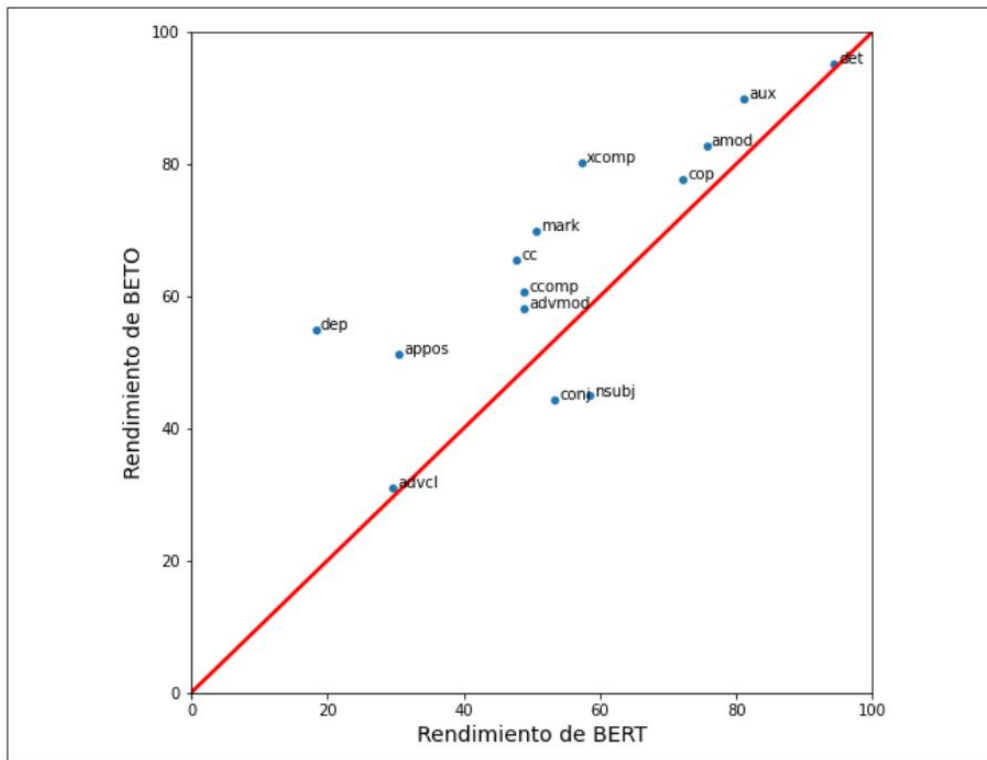
La autora concluye que la tecnología tiene un papel fundamental en el ejercicio del derecho, no solo en lo que respecta a la informática jurídica documental y la información jurídica de gestión. Las posibilidades de aplicar la informática metadocumentaria son amplias y pueden ayudar a los jueces a estructurar sus razonamientos y llegar a soluciones más efectivas.

5.4. Análisis de la sintaxis aprendida por BETO, un modelo de lenguaje en español basado en transformers.

Proyecto de investigación cuyo principal objetivo consistió en desarrollar un modelo de procesamiento de lenguaje natural más interpretable para el español (BETO), la investigación hace énfasis en el análisis específico con BERT y su relación con el idioma inglés, dando a conocer que los estudios sobre este quedan fuera de la zona hispano-hablante, y sus avances no son aplicables en esta, por lo tanto hace referencia a la necesidad de comprensión del comportamiento de BETO para desarrollar un modelo más cercano para el español.

La metodología implementada en el trabajo de referencia consistió primero en realizar un estudio de la sintaxis del modelo BETO, esto para comparar capacidades entre BETO y BERT y definir mapas de conocimiento. Segundo presentó un estudio sobre el comportamiento del modelo en base a mapas de conocimiento, resaltó patrones que presentan semejanzas entre mapas de relaciones similares y destacó los factores por los cuales ocurren estas similitudes. Tercero revisó factores que afectaron el rendimiento de reacciones sintácticas y evaluó cuales fueron causantes que la predicción fallara y como estos variaron según la relación de datos que fue analizada. Cuarto se enfocó en predecir cuando una relación no había sido reconocida correctamente, analizó si BETO contaba con tendencias a predecir lo mismo para distintas relaciones y estudio si el modelo poseía la capacidad de comprender una frase, incluso cuando no lograba descifrar sus dependencias sintácticas. Por último determinó si el modelo era capaz de combinar habilidades semánticas de un subconjunto de sus partes para generar un reconocimiento superior al utilizado, dando origen a conocimiento distribuido.

BERT vs BETO



Rendimiento de la mejor head en BETO y BERT para cada relación. Hay una alta correlación en el reconocimiento de las relaciones en ambos idiomas.

En las conclusiones de esta investigación se aborda la brecha que existe entre el idioma inglés y español para el entrenamiento de modelos de procesamiento de lenguaje natural, se concluye que BETO posee comprensión sintáctica de las frases que procesa. Además, posee un mejor rendimiento en coeficientes de atención y se observa que cuenta con patrones sintácticos. En cuanto a sus limitaciones se muestra que el principal factor para una predicción fallida es que la relación predicha se estructure de forma poco común. Para finalizar, la investigación espera que los resultados obtenidos sean útiles para darle un mayor alcance y comprensión a modelos de procesamiento de lenguaje natural como BETO, y que esto a su vez permitan el desarrollo de tecnologías de inteligencia artificial más avanzadas y seguras para el español. (Quiñonez, 2021)

5.5. Modelo de lenguaje español: RigoBERTa

Creado por el Instituto de Ingeniería del Conocimiento, entidad privada creada en 1989 por ADIC, asociación sin ánimo de lucro ubicada en la Universidad Autónoma de Madrid, dedicada a la investigación, desarrollo e innovación.

Este modelo de lenguaje en español está entrenado para la comprensión general del idioma hispano. Cuenta con la posibilidad de adaptarse a diferentes sectores, tales como: legal, salud, etc. para mejorar las aplicaciones del Procesamiento del Lenguaje Natural (PLN) en ámbitos específicos.

El principal objetivo de esta entidad fue desarrollar un modelo propio con más y mejores datos, mayor hardware de entrenamiento y mejor arquitectura de red neuronal que otros modelos del lenguaje español.

La metodología para desarrollar RigoBERTa surge del trabajo de un equipo multidisciplinar, formado principalmente por científicos de datos, expertos en el procesamiento de lenguaje natural (PLN) y lingüistas computacionales, quienes fueron los encargados del desarrollo y entrenamiento del modelo. Su arquitectura se basa en la arquitectura del modelo de lenguaje DeBERTa, este es un tipo de red neuronal de referencia que ya supera al rendimiento humano en tareas de procesamiento de lenguaje natural (PLN) en el idioma inglés. El entrenamiento del modelo se realizó con cuatro fuentes de datos en español que suponen más de 450.000 millones de palabras, esto para asegurar su calidad.

Se puede concluir que RigoBERTa se puede aplicar para resolver eficazmente distintas tareas de PLN, desde la clasificación de documentos hasta la detección de entidades, se adapta a distintos dominios de lenguaje, a través de un contexto determinado y representativo esto para reconocer la terminología específica de distintos ámbitos.

5.6. JurBERT: A Romanian BERT Model for Legal Judgement Prediction

Los autores proponen investigar la posibilidad de adaptar y aplicar modelos BERT para la predicción de sentencias judiciales en un conjunto de datos pequeño y ruidoso, en un idioma de pocos recursos (rumano). Usando datos con una representación realista, es decir, aquellos que son legibles por máquina y que están disponibles para los profesionales en este campo especializado. Los cuales fueron proporcionados por un banco rumano, y están compuestos por documentos judiciales originales que representan los tipos de casos más frecuentes relacionados con el dominio bancario.

En este trabajo, se emplea un estudio sobre la aplicabilidad de los métodos de PNL de última generación para la predicción de juicios legales en Rumania. Haciendo entrenamiento, ejecución y evaluación de los modelos con resultados prometedores en dos conjuntos de datos. El primero, contiene un resumen generado por humanos de argumentos clave, donde el modelo, jurBERT, supera las líneas base consideradas. Pasando al segundo conjunto de datos, que contiene todos los argumentos originales de las partes involucradas, en este caso jurBERT es solo un poco mejor que modelos mucho más simples, ya que tiene dificultades para manejar textos tan largos. Estas limitaciones de los modelos tipo BERT con respecto al tamaño máximo de entrada son un factor significativo que dificulta su desempeño.

5.7. El estado de la Legaltech en... Colombia

“Legaltechies es un proyecto del Legal LAB de Términos y Condiciones. Encabezado por Jorge Morell Ramos y junto al equipo de TyC, aborda el análisis en profundidad del universo Legaltech y el asesoramiento en la materia. Legaltechies ha sido responsable del primer mapa sobre la Legaltech española en 2016” (Legaltechies, 2020), en este mismo sentido, han abordado los avances tecnológicos aplicados al ámbito legal en otros países de habla hispana como Ecuador, Chile y Colombia, haciendo un recorrido histórico de la evolución en la materia y mapeando de manera general empresas privadas y públicas que han incursionado en lo que se conoce como legaltech.

Llegando a la conclusión de que “para un mercado joven como es el de la Legaltech colombiana, la mayoría de los principales nichos están ya presentes, existe una variedad de proyectos importante, se están desarrollando iniciativas públicas y privadas que potencian la innovación, y a todo ello se están sumando un nuevo tipo de estudio jurídico mucho más digitalizado y online (que sin duda merece su propio estudio). Por tanto, Colombia se posiciona sin duda como una de las referencias Legaltech en América Latina.” (Legaltechies, 2020) Se resaltan avances importantes en líneas de mercado como “plataformas de consultas jurídicas y reclamaciones online, los servicios para evidencias digitales, los relativos a la automatización de procesos legales y las plataformas para la creación de contratos online”, que usan modelos de procesamiento de lenguaje natural para apalancar y mejorar sus plataformas de servicios.

5.8. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

En esta publicación, se presenta “*Sentence-BERT*” (SBERT), una modificación de la red BERT que usa redes siamesas y tripletas que es capaz de derivar incrustaciones de oraciones semánticamente significativas. Esto permite que BERT se use para ciertas tareas que hasta ahora no eran aplicables. Las cuales incluyen la comparación de similitudes semánticas a gran escala, la agrupación y la recuperación de información a través de la búsqueda semántica. Esta modificación nace de la necesidad de mejorar el modelo BERT en términos de rendimiento computacional, ya que gestionar este tipo de soluciones demanda una cantidad importante de tiempo en procesamiento, el autor refiere que la búsqueda de oraciones semánticamente similares en un conjunto de datos de 10.000 podría tardar hasta 65 horas de entrenamiento, mientras que responder una sola consulta requeriría más de 50 horas.

La arquitectura de red siamesa permite derivar vectores de tamaño fijo para oraciones de entrada. Usando una medida de similitud como la similitud del coseno o la distancia Manhattan/Euclidiana, se pueden encontrar oraciones semánticamente similares. Estas medidas de similitud se pueden realizar de manera extremadamente eficiente en hardware moderno, lo que permite que SBERT se use para la búsqueda de similitud semántica, así como para la agrupación, reduciendo los tiempos antes mencionados de manera altamente significativa, donde el entrenamiento pasa a realizarse en 5 segundos y una consulta a unos milisegundos.

SBERT se puede adaptar a una tarea específica estableciendo un nuevo rendimiento de vanguardia en conjuntos de datos con similitud de argumentos o conjuntos de datos conformados por los tres elementos relevantes en la lingüística (el sujeto, el predicado y el objeto), para distinguir oraciones de diferentes secciones de un artículo de Wikipedia. (Reimers & Gurevych, 2019)

6. FUENTE DE LOS DATOS (Cómo se obtendrán los datos)

Se obtendrán los datos de la página de la Corte Constitucional utilizando una herramienta de web scraping desarrollada en Python, descargando documentos disponibles en la relatoría de la entidad con fecha de emisión de la sentencia hasta diciembre de 2022.

7. APLICACIÓN Y/O APOORTE ESPECÍFICO AL CAMPO

Aporta al área del derecho con la aplicación de metodologías de procesamiento de lenguaje natural en contextos específicos.

El trabajo se clasifica dentro del área del machine learning y el aprendizaje por refuerzo usando redes neuronales. Además, aporta al área de conocimiento del derecho.

8. METODOLOGÍA O ACTIVIDADES ESPECÍFICAS

1. Descarga por medio de webscraping de sentencias de la Corte Constitucional colombiana, disponibles de forma abierta en la relatoría de la página web de la entidad.
2. Limpieza de los textos para la generación del vocabulario que entrará a BERT. Esto eliminará palabras cortas ("stopwords"), signos de puntuación, correos electrónicos, caracteres especiales, entre otros. La generación de este vocabulario es indispensable para que la tokenización se haga de manera más eficiente, en el entendido que toda palabra que se encuentre en el vocabulario que se le entrega a BERT hará parte de la revisión de contexto, de lo contrario, el modelo partirá las palabras para poder hacer la calificación matemática en ambas direcciones.
3. Segunda limpieza de textos que eliminará correos electrónicos y caracteres especiales que no son relevantes para el entrenamiento de BERT. Esto generará una segunda base de datos de textos limpios con contexto jurídico. Se resalta el hecho de la necesidad de esta limpieza porque es importante mantener el contexto de los textos que entran al análisis, y por ejemplo, retirar las palabras cortas no respeta este principio.
4. Entrenamiento BERT: Con la limpieza del cuarto ítem y contando con el vocabulario actualizado del tercero, se empieza con el entrenamiento del modelo. El cual, hará la lectura de textos y podrá ser almacenado de manera física.
5. Utilización del modelo BERT entrenado en el campo jurídico colombiano para realizar a través de una interfaz de usuario búsquedas de documentos con relaciones semánticas.

9. RECURSOS

Recursos de personal:

Para la elaboración del proyecto se cuenta con el siguiente recurso humano:

- Angela Marcela Vargas Vargas - Economista.
- Cesar Augusto Suarez Peña - Ingeniero comercial.
- Omaira Jisel Otavo Madrigal - Administradora financiera.

- Luis Andres Campos Maldonado - Tutor de Maestría en Analítica de Datos.

Recursos hardware y software:

Hardware: Se utilizarán los equipos de cómputo de cada uno de los integrantes del presente proyecto de profundización.

Software: Se utilizará visual studio code para programar en python, miniconda como administrador de paquetes y se controlarán las versiones de los avances a través del repositorio de GitHub al que pertenecen todos los miembros del equipo de trabajo, en caso de ser necesario se utilizará un motor de base de datos (PostgreSQL).

10. CRONOGRAMA DE ACTIVIDADES

ACTIVIDADES A REALIZAR	Semanas de ejecución de cada actividad															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Actividad 1 (6 febrero - 18 marzo) - Presentación de propuestas de investigación por parte de los posibles tutores. - Consolidación del grupo de trabajo. - Establecimiento de la problemática a desarrollar en el anteproyecto. - Revisión semanal con tutor asignado						X										
Actividad 2 (21- marzo- 15 abril) -Reunión con el grupo de trabajo para asesoría y seguimiento. -Elaboración de cronograma de actividades. - Definición del título del trabajo de grado. -Elaboración de la introducción, justificación, objetivo general y objetivos específicos. - Revisión semanal con tutor asignado			X													
Actividad 3 (17-22 abril) -Revisión de Antecedentes y/o trabajos académicos relacionados con el problema de investigación. -Definición de la fuente de los datos - Revisión semanal con tutor asignado	X															
Actividad 4 (24-29 abril) -Verificación de la aplicación y/o aporte específico al campo relacionado al proyecto de investigación. - Revisión semanal con tutor asignado	X															
Actividad 5 (02-06 mayo) -Elaboración de metodología o actividades específicas - Revisión semanal con tutor asignado	X															
Actividad 6 (08-13 mayo) -Verificación de recursos Tecnológicos, humanos, entre otros. -Elaboración de fuentes de financiación. -Redacción de resultados esperados al llevar a cabo la propuesta. - Bibliografía - Revisión semanal con tutor asignado	X															

11. PRESUPUESTO (En caso de modalidad Investigación) **Y FUENTES DE FINANCIACIÓN** (En caso de modalidad Profundización)

Tipo	Categoría	Recurso	Descripción	Fuente de Financiación	Monto
Recursos Humanos	asesoría	Humano	Asesor externo especializado	Propios	0.00
Recursos disponibles	Infraestructura	Equipo	Laptop	Propios	0.00
		Equipo	Computador de escritorio	Propios	0.00
		Memoria USB	Memoria USB 16 GB	Propios	18,000.00
		Impresora	Impresora hp	Propios	0.00
Recursos necesarios	Materiales	Tóner	Pieza de tóner	Propios	45,000.00
		Papel	Hojas para la impresión de borradores	Propios	20,000.00
		Otros	Otros útiles (folder, lapiceros)	Propios	15,000.00
		Otros	Otros útiles (folder, lapiceros)	Propios	15,000.00
Servicios	Servicios	Internet y llamadas	Internet y llamadas telefónicas	Propios	200,000.00
	Servicios	Servicio de luz	Servicio de luz	Propios	308,000.00
	Servicios	Transporte	Transporte	Propios	200,000.00
Total fuentes de financiación					806,000.00

12. RESULTADOS ESPERADOS

A través de una interfaz de usuario realizar búsquedas de palabras específicas que den como salida una lista de sentencias de la corte que tengan relación con la búsqueda inicial, es decir, que este texto no necesariamente debe estar contenido en las sentencias encontradas porque el modelo debe estar en la capacidad de relacionar las similitudes semánticas dentro de los diferentes documentos. Y así, mejorar la precisión de las consultas de las búsquedas, lo que permite a los usuarios encontrar información relevante de manera más rápida y eficiente.

13. REFERENCIAS

Briggs, J. (n.d.). *Natural Language Processing (NLP) for Semantic Search*. Pinecone. Retrieved May 2, 2023, from <https://www.pinecone.io/learn/data-augmentation/>

Corporación Excelencia en la Justicia. (2023, March 22). *Índice de Congestión de la Rama Judicial en Colombia (Sector Jurisdiccional)*. Corporación Excelencia en la Justicia. Retrieved May 2, 2023, from <https://cej.org.co/indicadores-de-justicia/efectividad/indice-de-congestion-de-la-rama-judicial-en-colombia-sector-jurisdiccional/>

Instituto de Ingeniería del Conocimiento. (n.d.). *Modelo de lenguaje en español - RigoBERTa*. IIC. Retrieved May 2, 2023, from <https://www.iic.uam.es/inteligencia-artificial/procesamiento-del-lenguaje-natural/modelo-lenguaje-espanol-rigoberta/>

Legaltechies. (2020, September 2). *El estado de la Legaltech en... Colombia*. Legaltechies. Retrieved April 26, 2023, from <https://legaltechies.es/2020/09/02/el-estado-de-la-legaltech-en-colombia/>

Martínez Bahena, G. C. (2013, February 5). *La inteligencia artificial y su aplicación al campo del Derecho*. Corte Interamericana de Derechos Humanos. Retrieved May 2, 2023, from <https://www.corteidh.or.cr/tablas/r30570.pdf>

Masala, M., & Iacob, R. (2021, Noviembre 10). *JurBERT: A Romanian BERT Model for Legal Judgement Prediction* [JurBERT: un modelo BERT rumano para la predicción de juicios legales]. [aclanthology.org](https://aclanthology.org/2021.nllp-1.8.pdf).
<https://aclanthology.org/2021.nllp-1.8.pdf>

Quiñonez, A. (2021). *Análisis de la sintaxis aprendida por BETO, un modelo de lenguaje en español basado en transformers*. repositorio.uc.cl. <https://doi.org/10.7764/tesisUC/ING/60582>

Reimers, N., & Gurevych, I. (2019, Agosto 27). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arxiv.org. Retrieved May 2, 2023, from <https://arxiv.org/pdf/1908.10084.pdf>

14. DATOS DE LA ORGANIZACIÓN (En caso de pasantía)

NOMBRE DE LA ORGANIZACIÓN:



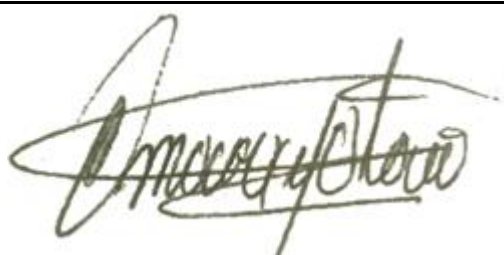
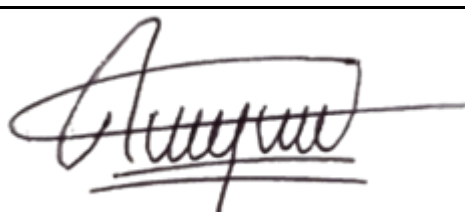
RESPONSABLE DE LA ORGANIZACIÓN:

CORREO:

TELÉFONO:

DURACIÓN VINCULACIÓN (horas):

15. FIRMAS

FIRMA DEL ESTUDIANTE:	FIRMA DEL DOCENTE DIRECTOR O TUTOR
 <p>Angela Marcela Varga Vargas</p>	 <p>Luis Andrés Campos Maldonado</p>
 <p>Omaira Jisel Otravo Madrigal</p>	
 <p>Cesar Augusto Suarez Peña</p>	

16. DATOS DE TRÁMITE COMITÉ DE INVESTIGACIÓN DEL PROGRAMA (Espacio para diligenciar por el Comité del Programa)

No. CONSECUTIVO	
No. ACTA	
FECHA	