



UNIVERSIDAD CATÓLICA DEL NORTE
ESCUELA DE INGENIERÍA

LEGALSENT NEUTRALIZER.

Proyecto de identificación de sesgos emocionales en documentos legales
mediante Análisis de Sentimientos

Exequiel Vera G.

Natalie Valdés M.

Profesor guía: Dr. Juan Bekios Calfa

Coquimbo, Chile

2024

RESUMEN

El presente informe describe el proyecto Legalsent Neutralizer, el cual busca abordar la problemática de los sesgos emocionales presentes en documentos legales mediante el uso de análisis de sentimientos y procesamiento de lenguaje natural (NLP). La necesidad de este proyecto radica en la importancia de asegurar la imparcialidad en el lenguaje legal, garantizando que los documentos reflejen un tono neutral y equilibrado. Para lograr este objetivo, se propone adaptar un modelo de machine learning, específicamente variantes del modelo BERT, que permita evaluar y clasificar la carga emocional en textos jurídicos.

Se han definido varias fases para la implementación del proyecto, incluyendo la planificación, investigación de modelos y bases de datos, adaptación del modelo, evaluación y validación, despliegue y documentación. Los resultados esperados incluyen la creación de un modelo de machine learning capaz de identificar sentimientos en textos legales, una interfaz web para el análisis de documentos, y un informe detallado que documente todo el proceso y los hallazgos obtenidos. Este proyecto tiene el potencial de contribuir significativamente a la promoción de la justicia y equidad en el sistema judicial mediante la identificación y corrección de sesgos emocionales en los documentos legales.

ABSTRACT

The "Legalsent Neutralizer" project aims to address the issue of emotional biases in legal documents through the use of sentiment analysis and natural language processing (NLP). Ensuring impartiality in legal language is crucial, and this project focuses on adapting a machine learning model, specifically BERT variants, to evaluate and classify the emotional content of legal texts.

The project includes phases such as planning, researching models and databases, adapting the model, evaluation and validation, deployment, and documentation. Expected results are the development of a machine learning model capable of identifying sentiments in legal texts, an interactive web interface for document analysis, and a detailed report documenting the process and findings. This project aims to significantly contribute to promoting justice and equity in the judicial system by identifying and correcting emotional biases in legal documents.

NOMENCLATURA

BERT (*Bidirectional Encoder Representations from Transformers*)

CSV (*Comma-Separated Values*)

GPU (*Graphics Processing Unit*)

NLP (*Natural Language Processing*)

TRL (*Technology Readiness Level*)

UCN (*Universidad Católica del Norte*)

GLOSARIO

Backlog Registro donde se almacenan todas las historias de usuario y tareas pendientes, dichas tareas pueden ir aumentando a medida que avance el proyecto a raíz de reuniones de equipo o *stakeholders*.

Framework Toda aquella abstracción de *software* (formada por la combinación de múltiples librerías) que entrega las herramientas necesarias para crear y desarrollar aplicaciones (y soluciones) de forma estandarizada y reusable.

Historias de usuario Narración desde la perspectiva del usuario objetivo que busca plantear la necesidad de cierta funcionalidad de *software*. Es también el mecanismo o soporte utilizado en el Análisis de Requisitos que busca sintetizar requerimientos y objetivos de un sistema de *software*. ?

Librería Conjunto de archivos que aportan funcionalidad a un programa al entregar una solución a un problema recurrente

Módulo Se refiere a las distintas secciones en que se subdivide el proyecto las cuales se centran en los diferentes problemas para la conformación del mismo.

Scrum Master Es el encargado de que se lleven a cabo de manera exitosa los valores que son descritos por intermedio de la metodología ágil en el modelo scrum.

Sprint Es un período de tiempo limitado donde el equipo se asigna un cierto número de tareas a realizar.

Stakeholders Actores que tienen interés o algún tipo de relación con una empresa o proyecto y a los cuales puede afectarles, ya sea de manera directa u indirecta la toma de decisiones.

User Story Mapping Es un diagrama que permite a los gerentes de producto, desarrolladores y diseñadores poner un método detrás de la continua complejidad del diseño del producto y crear una experiencia de usuario (UX) increíble, ?

TABLA DE CONTENIDOS

1	INTRODUCCIÓN	1
1.1	Descripción de la institución	1
1.2	Descripción del problema	1
1.3	Justificación del problema	2
1.4	Objetivos	3
1.4.1	Objetivo General	3
1.4.2	Objetivos Específicos	3
1.5	Descripción detallada del trabajo a realizar	3
1.6	Resultados esperados	6
1.7	Metodología de trabajo	7
1.8	Recursos, limitaciones y supuestos	8
1.9	Plan de trabajo	9
1.9.1	Fase 1: Planificación del Proyecto	9
1.9.2	Fase 2: Investigación de Modelos y Bases de Datos	9
1.9.3	Fase 3: Adaptación del Modelo	10
1.9.4	Fase 4: Evaluación y Validación	11
1.9.5	Fase 5: Despliegue	11
1.9.6	Fase 6: Documentación	12
1.10	Nivel de madurez tecnológica	12
2	MARCO TEÓRICO	13
2.1	Procesamiento de Lenguaje Natural (NLP)	13

TABLA DE CONTENIDOS

2.2	Modelos de Lenguaje	14
2.2.1	Modelos BERT	14
2.2.2	Modelos BERTA	14
2.2.3	Transformers	14
2.3	Técnicas de Aprendizaje Automático	15
2.3.1	Redes Neuronales	15
2.3.2	Fine-Tuning	15
2.4	Técnicas de NLP Específicas	15
2.4.1	Tokenización	15
2.4.2	Embedding	16
2.4.3	Masked Language Modeling	16
2.4.4	Next Sentence	16
2.5	Metodologías Ágiles	16
2.5.1	Scrum	16
2.5.2	Backlog	17
3		18
3.1	Requisitos del problema	18
3.1.1	Requisitos Funcionales	18
3.1.2	Requisitos No Funcionales	19
3.1.3	Limitaciones y Supuestos	19
3.1.4	Necesidades del Usuario	20
4	RESULTADOS	21
4.0.1	BERT	21
4.0.2	BERTA	22
4.0.3	BERT VS BERTA	22
5	CONCLUSIÓN	24

TABLA DE CONTENIDOS

ANEXOS	27
A Resultados de BERT	27
B Resultados de BERTA	28

ÍNDICE DE FIGURAS

Figura A.1	Clasificación de sentimiento de los textos jurídicos con BERT . . .	27
Figura B.1	Clasificación de sentimiento de los textos jurídicos con BERT . . .	28

Capítulo 1

INTRODUCCIÓN

1.1. Descripción de la institución

La Universidad Católica del Norte en Coquimbo es una institución de educación superior ubicada en la cuarta región de Chile. Fundada en 1956, la UCN es una de las universidades más importantes y reconocidas en el norte del país, ofreciendo una amplia gama de programas académicos en diversas áreas del conocimiento, incluyendo ciencias naturales, ciencias sociales, humanidades, ingeniería, negocios y más. La universidad se destaca por su compromiso con la excelencia académica, la investigación y el servicio a la comunidad. Además de sus programas académicos, la UCN en Coquimbo ofrece una variedad de actividades extracurriculares y servicios para estudiantes, incluyendo clubes estudiantiles, deportes, actividades culturales y más, contribuyendo así, al enriquecimiento de la experiencia universitaria.

1.2. Descripción del problema

Legalsent Neutralizer es un proyecto que consiste en la aplicación de técnicas de análisis de sentimientos en documentos legales para evaluar

si el tono del texto refleja una imparcialidad adecuada bajo un contexto legal, mediante el uso de procesamiento de lenguaje natural y análisis de sentimientos, que permiten identificar cualquier expresión o connotación emocional que pueda sesgar el contenido del documento. La necesidad de abordar este problema radica en promover la equidad y la justicia en el sistema legal al detectar y corregir posibles sesgos emocionales puedan tener los jueces que dictan las sentencias. Al garantizar la neutralidad se contribuye a un sistema legal más justo y transparente, donde los documentos reflejen un enfoque imparcial y equilibrado.

1.3. Justificación del problema

Dentro del sistema legal, la imparcialidad es esencial para garantizar la justicia y la equidad; sin embargo, el lenguaje en los documentos legales puede estar cargado de emociones que introducen sesgos inadvertidos, afectando la percepción y la interpretación de los hechos. En Chile, aunque el proceso judicial sigue un procedimiento riguroso, asegurar que los documentos reflejen verdadera neutralidad podría presentarse como un desafío. Legalsent Neutralizer surge para abordar esta necesidad mediante la integración de procesamiento de lenguaje natural y análisis de sentimientos, con el objetivo de identificar y corregir sesgos emocionales en los textos legales. Al hacerlo, se promueve un sistema judicial más justo y equitativo, se fortalece la confianza pública en las instituciones judiciales, y se demuestra cómo las tecnologías avanzadas pueden mejorar la objetividad y transparencia en la redacción de documentos legales, asegurando que las decisiones se basen únicamente en hechos y leyes, libres de influencias emocionales.

1.4. Objetivos

1.4.1. Objetivo General

- Adaptar un modelo de machine learning que determine la emocionalidad de un texto jurídico en un plazo de 3 meses.

1.4.2. Objetivos Específicos

1. Planificar las actividades del proyecto para gestionar la correcta ejecución de todas las tareas.
2. Seleccionar técnicas de análisis de sentimientos y modelos pre-entrenados de NLP.
3. Implementar modelos de redes neuronales utilizando técnicas de NLP.
4. Validar con el cliente el modelo final.
5. Desplegar el modelo final.

1.5. Descripción detallada del trabajo a realizar

1. Investigación de la problemática.
 - Entender el problema y su relevancia, como este trabajo es crucial para un sistema legal más justo.
2. Investigación del marco teórico.
 - Buscar los conceptos claves para entender el problema.
 - Investigar cual es el tipo de estructura mas adecuada para el problema que se quiere resolver.

3. Adquisición del Modelo BERTA.

- Obtener un modelo preentrenado de BERT desde Hugging Face, especializado en análisis de sentimientos.

4. Preparación de Datos.

- Recolectar y preparar un conjunto de datos de textos jurídicos chilenos.
- Dividir los textos en oraciones utilizando una librería adecuada, como la librería de tokens de transformers.

5. Predicción y Etiquetación.

- Utilizar el modelo entrenado para predecir los sentimientos de las oraciones en los textos jurídicos.
- Etiquetar cada oración con su sentimiento correspondiente (positivo, negativo, neutral).

6. Cálculo de Métricas de Neutralidad.

- Analizar las predicciones del modelo para calcular métricas de neutralidad en los textos.
- Evaluar la proporción de oraciones etiquetadas como neutrales y su distribución en los textos.

7. Desplegar el modelo.

- Crear una página web para poder clasificar textos en formato PDF.
- Subir la página a la red.

8. Informe de Resultados.

- Documentar los resultados obtenidos como el análisis de neutralidad.
- Preparar un informe detallado que describa el proceso, los resultados y las conclusiones del proyecto.

1.6. Resultados esperados

Legalsent Neutralizer tiene como objetivo principal desarrollar y adaptar un modelo de machine learning para evaluar la emocionalidad de textos jurídicos, asegurando que los documentos legales reflejen un tono neutral y adecuado. Los resultados esperados de este proyecto incluyen:

1. **Modelo de Machine Learning Adaptado:** Se espera la adaptación de un modelo basado en técnicas avanzadas de procesamiento de lenguaje natural (NLP) y análisis de sentimientos que pueda identificar y evaluar la carga emocional en textos legales. Este modelo debe ser capaz de distinguir entre sentimientos positivos, negativos y neutrales, proporcionando una evaluación precisa de la emocionalidad del documento.
2. **Análisis Detallado de Sentimientos en Documentos Legales:** El proyecto resultará en un sistema capaz de analizar extensos corpus de textos jurídicos chilenos, etiquetando oraciones individuales según su carga emocional. Este análisis detallado permitirá identificar patrones y tendencias en la redacción de documentos legales que puedan introducir sesgos emocionales inadvertidos.
3. **Interfaz Web para Clasificación de Textos:** Desarrollo de una plataforma web interactiva que permita a los usuarios subir documentos legales en formato PDF y recibir un análisis detallado de la emocionalidad de los textos. Esta interfaz facilitará el acceso a la tecnología desarrollada, permitiendo a los profesionales legales y otras partes interesadas evaluar la neutralidad de los documentos de manera eficiente.
4. **Informe Exhaustivo de Resultados:** Al finalizar el proyecto, se generará

un informe detallado que documente todo el proceso de desarrollo, desde la investigación inicial y la preparación de datos hasta la implementación del modelo y la validación de resultados. Este informe incluirá análisis cuantitativos y cualitativos de la efectividad del modelo, así como recomendaciones para futuras mejoras y aplicaciones potenciales en otros contextos legales.

1.7. Metodología de trabajo

Para la fase de gestión del proyecto se utilizaron diferentes herramientas para trazar el camino a seguir, con el objetivo de lograr los resultados esperados. Entre las metodologías utilizadas se encuentran:

Agile inception

Para comenzar con el entendimiento de la problemática, es clave ver a grandes rasgos lo necesario para resolver el problema. El punto de comienzo y el objetivo a alcanzar son información vital para trazar un camino. Por esto, se utilizaron técnicas como el Roadmap, con el fin de construir los objetivos específicos y el objetivo general. Además, un User History Map es relevante para, en un futuro, tener el product backlog y poder definir los sprints semanales.

Entendimiento del Negocio (Fase de CRISP-DM)

En esta etapa, se llevó a cabo un análisis exhaustivo de los requisitos del negocio y de las necesidades específicas del proyecto. Se identificaron los objetivos principales, los objetivos específicos, los stakeholders involucrados y el product owner. Se definieron las métricas clave de rendimiento para evaluar el éxito del modelo.

Desarrollo Ágil (Scrum)

Se adoptó el marco de trabajo Scrum para organizar el desarrollo del proyecto en sprints de dos semanas. Se formó un equipo compuesto por un líder, un Data Scientist y un scrum master. Se programaron revisiones formales del progreso del proyecto cada dos días, donde se evaluaron los avances realizados y se tomaron decisiones basadas en los resultados obtenidos hasta el momento. Estas revisiones permitieron realizar ajustes rápidos en la dirección del proyecto y garantizar que se cumplieran los objetivos establecidos.

Al integrar la metodología CRISP-DM con prácticas ágiles como Scrum, se logró un enfoque estructurado y flexible para la correcta planificación del proyecto, permitiendo una adaptación continua a medida que se descubrían nuevas necesidades o desafíos en el camino.

1.8. Recursos, limitaciones y supuestos

Dentro de los recursos disponibles para la realización del proyecto, se cuenta con el apoyo de la fiscalía, la cual facilitará los textos jurídicos con los cuales se determinará la emocionalidad de los mismos.

Se trabajará con Hugging Face para encontrar modelos que se adapten a la situación, donde hay una amplia variedad de modelos entrenados de uso libre.

Sin embargo, se enfrenta una gran limitación respecto a los datos para el entrenamiento del modelo, ya que no existen bases de datos etiquetadas con textos jurídicos legales. Además, en general, los modelos ya entrenados están en inglés. Al no tener datos etiquetados, no se puede hacer transfer learning o fine-tuning a los modelos, por lo cual existe el riesgo de que el modelo no generalice bien y no pueda identificar emocionalidades en los textos.

Sumado a lo anterior, no existen métricas para evaluar si el funcionamiento del modelo es correcto y predice de buena manera los textos jurídicos. El único método válido

es que un jurado con estudios al respecto lea las oraciones y su predicción, y nos dé su veredicto sobre si están bien clasificadas. Esta sería la única manera de comprobar la efectividad del modelo.

1.9. Plan de trabajo

El plan de trabajo para el proyecto se organiza en varias fases, cada una con tareas específicas y una duración estimada. A continuación se presenta una descripción detallada de cada fase y sus respectivas tareas:

1.9.1. Fase 1: Planificación del Proyecto

En esta fase inicial, se establecen las bases del proyecto mediante la definición clara de objetivos, metodologías y el alcance del proyecto.

Duración: 2 semanas

Tareas:

- Presentación de la planificación
- Construir un Roadmap
- User history map
- Investigar metodología Scrum
- Redactar objetivos general y específicos
- Project Charter

1.9.2. Fase 2: Investigación de Modelos y Bases de Datos

Durante esta fase, se realiza una exhaustiva investigación sobre los modelos de análisis de sentimientos más adecuados para el proyecto.

Duración: 3 semanas

Tareas:

- Definir modelos a utilizar
- Investigar NLP
- Investigar embeddings
- Buscar bases de datos etiquetadas
- Investigar a fondo BERT
- Buscar bases de datos adicionales
- Buscar códigos de BERT
- Buscar modelos de BERT

1.9.3. Fase 3: Adaptación del Modelo

En esta fase, se ajustan y refinan los modelos seleccionados para que se adapten específicamente al análisis de textos jurídicos en el contexto chileno.

Duración: 4 semanas

Tareas:

- Hacer split de los documentos
- Modificar objetivos
- Hacer presentación de avances
- Realizar primeras pruebas de predicción

1.9.4. Fase 4: Evaluación y Validación

Esta fase se centra en la validación del modelo adaptado.

Duración: 4 semanas

Tareas:

- Desplegar el modelo en Google Colab
- Validar el modelo con el Product Owner
- Desplegar el modelo
- Cargar modelo en la GPU
- Guardar predicciones en CSV
- Crear gráficos analíticos de las predicciones
- Obtener porcentaje de los sentimientos presentes en los textos

1.9.5. Fase 5: Despliegue

En la fase de despliegue, se desarrolla una plataforma web interactiva que permite a los usuarios subir documentos legales y recibir análisis de sentimientos.

Duración: 3 semanas

Tareas:

- Desarrollar una página web interactiva para la clasificación de textos en formato PDF.
- Integrar el modelo de análisis de sentimientos en la plataforma web y realizar pruebas de usabilidad.
- Subir la página web a la red y asegurar su accesibilidad para los usuarios finales.

1.9.6. Fase 6: Documentación

La fase final del proyecto implica la documentación detallada de todo el proceso, desde la investigación inicial hasta el despliegue y validación

Duración: 2 semanas

Tareas:

- Documentar todos los resultados obtenidos, incluyendo el análisis de neutralidad y la efectividad del modelo.
- Preparar un informe detallado que describa el proceso, los resultados y las conclusiones del proyecto.

1.10. Nivel de madurez tecnológica

El proyecto se encuentra en un nivel de madurez tecnológica (TRL) 3. Este nivel indica que la tecnología ha sido probada de manera experimental y que existen pruebas de concepto que validan la viabilidad de los principios fundamentales.

En esta fase, los principios analíticos y experimentales críticos han sido validados. Se han desarrollado y probado dos modelos de análisis de sentimientos, BERT y BERTA, que entregan la clasificación de los textos, que si bien han mostrado resultados prometedores en pruebas de laboratorio, la validación externa con la fiscalía y la implementación en un entorno real aún están pendientes. El proyecto está en una fase donde los componentes básicos funcionan en un entorno controlado y simulado, pero requiere más pruebas y validación en situaciones del mundo real para avanzar a niveles superiores de madurez tecnológica.

Capítulo 2

MARCO TEÓRICO

En este capítulo se presenta el marco teórico que sustenta a Legalsent Neutralizer, proporcionando una base conceptual sólida para el desarrollo y la implementación del sistema. Se abordan las principales metodologías y tecnologías utilizadas en el análisis de sentimientos aplicado a documentos legales, con énfasis en el procesamiento de lenguaje natural (NLP) y el aprendizaje automático.

2.1. Procesamiento de Lenguaje Natural (NLP)

El procesamiento de lenguaje natural (NLP) es una rama de la inteligencia artificial que se centra en la interacción entre computadoras y el lenguaje humano. Su objetivo es permitir que las computadoras comprendan, interpreten y generen lenguaje humano de manera efectiva. En el contexto del proyecto, el NLP se utiliza para analizar y evaluar la carga emocional de los textos jurídicos.

2.2. Modelos de Lenguaje

2.2.1. Modelos BERT

BERT (Bidirectional Encoder Representations from Transformers) es un modelo de lenguaje basado en la arquitectura Transformers desarrollado por Google. BERT es conocido por su capacidad para comprender el contexto bidireccional en el que aparecen las palabras en una oración, lo que lo hace muy efectivo en tareas de procesamiento de lenguaje natural, como la clasificación de texto, la traducción automática y la generación de texto. En el proyecto, se utilizarán variantes de BERT para el análisis de sentimientos.

2.2.2. Modelos BERTA

BERTA es una variante de BERT optimizada para el análisis de sentimientos. Los modelos BERTA son entrenados específicamente para tareas de clasificación emocional, mejorando la precisión y eficiencia en la identificación de sesgos emocionales en textos legales. Al igual que BERT, BERTA utiliza la arquitectura Transformers, pero se enfoca más en capturar matices emocionales, lo que lo hace particularmente útil para el objetivo de este proyecto.

2.2.3. Transformers

Los Transformers son una arquitectura de red neuronal diseñada para modelar relaciones a largo plazo en conjuntos de datos secuenciales, como texto o audio. Se destacan por su capacidad para capturar dependencias de largo alcance y han demostrado ser altamente efectivos en una variedad de tareas de aprendizaje automático, incluido el procesamiento de lenguaje

natural.

2.3. Técnicas de Aprendizaje Automático

2.3.1. Redes Neuronales

Las redes neuronales son modelos computacionales inspirados en el funcionamiento del cerebro humano. Están compuestas por capas de nodos interconectados que procesan datos y aprenden patrones para realizar tareas específicas, como reconocimiento de imágenes y procesamiento de lenguaje natural. En el proyecto, se utilizarán redes neuronales para implementar modelos de análisis de sentimientos.

2.3.2. Fine-Tuning

El Fine-Tuning es una técnica en el aprendizaje automático donde un modelo pre-entrenado se ajusta o adapta a una tarea específica mediante el ajuste de sus pesos y parámetros utilizando datos etiquetados adicionales. Esto permite que el modelo pre-entrenado se especialice en una tarea particular sin necesidad de entrenarlo desde cero.

2.4. Técnicas de NLP Específicas

2.4.1. Tokenización

La tokenización es el proceso de dividir un texto en unidades más pequeñas, llamadas tokens. Estos tokens pueden ser palabras individuales, subpalabras o caracteres, y se utilizan como entrada para modelos de procesamiento de lenguaje natural. La tokenización es un paso crucial en el preprocesamiento de datos para el análisis de sentimientos.

2.4.2. Embedding

El embedding es una técnica utilizada en el procesamiento de lenguaje natural para representar palabras o tokens como vectores de números reales en un espacio dimensional. Estos vectores capturan el significado semántico de las palabras y se utilizan como entrada para modelos de aprendizaje automático.

2.4.3. Masked Language Modeling

El Masked Language Modeling es una técnica de pre-entrenamiento utilizada en modelos de lenguaje como BERT. Consiste en ocultar aleatoriamente ciertas palabras o tokens en una oración de entrada y entrenar al modelo para predecir las palabras ocultas en función del contexto circundante.

2.4.4. Next Sentence

La predicción de la próxima oración (Next Sentence Prediction) es una tarea utilizada en el pre-entrenamiento de modelos de lenguaje como BERT. El modelo se entrena para determinar si una oración dada es la continuación lógica de la oración anterior en un corpus de texto, lo que ayuda al modelo a comprender mejor la relación entre las oraciones en un texto continuo.

2.5. Metodologías Ágiles

2.5.1. Scrum

Scrum es una metodología ágil que se utiliza para gestionar el desarrollo de proyectos de manera eficiente y flexible. En Scrum, el trabajo se

organiza en períodos de tiempo limitados llamados sprints, donde el equipo se asigna un cierto número de tareas a realizar. Un Scrum Master es el encargado de asegurar que los valores y prácticas de Scrum se lleven a cabo de manera exitosa.

2.5.2. Backlog

El backlog es un registro donde se almacenan todas las historias de usuario y tareas pendientes. Estas tareas pueden ir aumentando a medida que avance el proyecto a raíz de reuniones de equipo o stakeholders. El backlog es una herramienta esencial en la gestión de proyectos ágiles, proporcionando un listado priorizado de las tareas que deben completarse.

Capítulo 3

3.1. Requisitos del problema

Legalsent Neutralizer busca abordar la problemática de sesgos emocionales presentes en documentos legales. En el contexto judicial, la imparcialidad es esencial para asegurar la justicia y equidad, pero a menudo el lenguaje utilizado en dichos documentos puede estar cargado de emociones que introducen sesgos inadvertidos. Estos sesgos pueden afectar la percepción e interpretación de los hechos, poniendo en riesgo la objetividad de las decisiones legales. El objetivo del proyecto es analizar los textos proporcionados a través de análisis estadísticos y determinar si contienen sesgos o son neutrales.

3.1.1. Requisitos Funcionales

- **Análisis de Sentimientos:** El sistema debe ser capaz de analizar el contenido de los documentos legales y clasificar cada oración según su carga emocional (positiva, negativa, neutral).
- **Detección de Sesgos:** El sistema debe identificar y señalar las oraciones que contienen un sesgo emocional, proporcionando un reporte detallado.

- Interfaz de Usuario: Debe haber una interfaz web donde los usuarios puedan subir documentos en formato PDF para su análisis.
- Generación de Reportes: El sistema debe generar reportes de análisis que incluyan métricas de neutralidad y sugerencias para mejorar la imparcialidad del texto.
- Base de Datos: El sistema debe almacenar los documentos analizados y los resultados de los análisis para futuras referencias y comparaciones.

3.1.2. Requisitos No Funcionales

- Rendimiento: El sistema debe ser capaz de procesar y analizar documentos en un tiempo considerablemente bajo.
- Seguridad: Debe garantizarse la confidencialidad y seguridad de los documentos analizados, evitando cualquier acceso no autorizado.
- Usabilidad: La interfaz de usuario debe ser intuitiva y fácil de usar para que cualquier persona con conocimientos básicos de informática pueda utilizarla sin problemas.
- Escalabilidad: El sistema debe ser escalable para manejar un creciente número de usuarios y documentos sin pérdida de rendimiento.

3.1.3. Limitaciones y Supuestos

- Disponibilidad limitada de conjuntos de datos etiquetados con emociones específicas en textos legales en español.
- Modelos pre-entrenados disponibles principalmente en inglés, lo que podría requerir una adaptación significativa.

- Evaluación de la efectividad del sistema dependiente de la validación por expertos legales.
- Se asume que los documentos proporcionados para análisis están en formato digital legible (PDF).
- Se espera contar con la colaboración de instituciones legales para la validación de los resultados.
- Se supone que el modelo BERT o variantes adaptadas serán suficientes para capturar las emociones en los textos legales.

3.1.4. Necesidades del Usuario

Los usuarios principales del sistema serán profesionales legales, jueces, abogados, y otros stakeholders en el sistema judicial. Ellos necesitan una herramienta que les permita asegurar que los documentos legales que redactan y revisan sean imparciales y objetivos. Las necesidades específicas incluyen:

- Facilidad de uso: Una plataforma intuitiva donde puedan cargar documentos y obtener resultados rápidamente.
- Precisión en el análisis: Alta precisión en la detección de sesgos emocionales para confiar en los resultados proporcionados.
- Confidencialidad: Garantía de que los documentos analizados no serán divulgados a terceros.

Capítulo 4

RESULTADOS

En cuanto al análisis de resultados se utilizaron dos modelos para poder compararlos y ver cual es el que lograba un trabajo más efectivo en cuanto a predecir las emocionalidades. Estos dos modelos son BERT y BERTA, en el cual el segundo es una variación del primero, que a priori debería entregar mejores resultados.

4.0.1. BERT

Respecto a BERT, los resultados obtenidos discrepan bastante de la realidad. Se obtienen, en promedio, los siguientes porcentajes de cada emocionalidad (Ver Anexo A):

- Positivo: 13.18 %
- Negativo: 85.47 %
- Neutral: 1.35 %

Se observa una clara predominancia de la emocionalidad negativa dentro de los textos. Sería arriesgado decir si esto concuerda con la realidad de los textos, ya que debe ser evaluado por un experto en la materia.

Sin embargo, a simple lectura, los textos no parecen tener emocionalidad, por lo cual se podría indicar que el modelo no está generalizando bien. Además, en promedio, los textos solo presentan un 1.35 % de neutralidad.

4.0.2. BERTA

Ahora bien, respecto a BERTA, ocurre una situación particular, y es que la mayoría de los casos son neutros. A priori, se puede interpretar que los textos jurídicos son completamente neutros. Sin embargo, al hacer pruebas exhaustivas con textos con una clara tendencia a alguna emocionalidad, también los clasificaba como neutros.

Esto se debe principalmente a que este modelo de BERTA está entrenado con datos en inglés, por lo cual no generaliza bien al español. Esto puede sesgar la interpretación de los textos jurídicos, haciendo pensar que todos son neutros cuando, en realidad, el modelo no puede generalizar correctamente y clasifica la mayoría de los textos en español como neutros.

El porcentaje promedio de las emocionalidades es (Ver Anexo B):

- Positivo: 0,00 %
- Negativo: 0,17 %
- Neutral: 99,98 %

4.0.3. BERT VS BERTA

Al comparar los dos modelos, no sería apropiado decir que alguno clasifica de manera óptima las emocionalidades dentro de los textos jurídicos. Desde un punto de vista lógico, BERTA hace un mejor trabajo, ya que, a

priori, los textos jurídicos en su mayoría deberían ser presuntamente neutros. Sin embargo, como se mencionó anteriormente, esto puede estar sesgado por las limitaciones del modelo. Para tener un mejor criterio sobre la clasificación del trabajo realizado, debería ser chequeado por un especialista en la materia.

Para obtener mejores resultados, se debería hacer un proceso de transfer learning del modelo de BERTA, entrenándolo con palabras en español y luego realizar un proceso de fine-tuning para lograr que generalice con textos jurídicos. Debido a que el léxico de estos textos es distinto, sería idóneo que el modelo se entrenara con documentos de este estilo. Sin embargo, para ello es necesario contar con una base de datos etiquetada.

Capítulo 5

CONCLUSIÓN

Legalsent Neutralizer ha sido diseñado para abordar la problemática de los sesgos emocionales en documentos legales mediante el uso de técnicas avanzadas de procesamiento de lenguaje natural (NLP) y análisis de sentimientos. Este proyecto tenía como objetivo principal desarrollar un sistema capaz de identificar y corregir sesgos emocionales, asegurando así la imparcialidad y neutralidad en los textos legales.

A lo largo del desarrollo del proyecto, se implementaron y compararon dos modelos de machine learning, BERT y su variante BERTA, para evaluar su eficacia en la detección de emociones en textos jurídicos. Los resultados mostraron que ambos modelos presentaron limitaciones en su capacidad para generalizar adecuadamente en textos en español. BERT tendió a clasificar las oraciones con una carga emocional negativa predominante, mientras que BERTA clasificó la mayoría de los textos como neutrales, independientemente de su contenido emocional real.

Estos hallazgos destacan la necesidad de realizar un proceso de transferencia de aprendizaje (transfer learning) y ajuste fino (fine-tuning) utilizando datos en español, especialmente textos jurídicos, para mejorar la precisión y efectividad de los modelos en el contexto legal chileno. La

implementación de una interfaz web interactiva permitirá a los usuarios subir y analizar documentos legales de manera eficiente, facilitando el acceso a esta tecnología innovadora.

Las conclusiones del proyecto indican que, aunque se han realizado avances significativos hacia la identificación y corrección de sesgos emocionales en documentos legales, es necesario continuar con el desarrollo y la validación del modelo para asegurar su efectividad en entornos del mundo real. Este proyecto contribuye de manera significativa a la promoción de la justicia y la equidad en el sistema judicial, demostrando cómo las tecnologías avanzadas pueden mejorar la objetividad y transparencia en la redacción y revisión de documentos legales.

Legalsent Neutralizer no solo ha confirmado la viabilidad del uso de técnicas de NLP y análisis de sentimientos en el ámbito legal, sino que también ha establecido una base sólida para futuras investigaciones y desarrollos en este campo, subrayando la importancia de adaptar los modelos a contextos específicos y la necesidad de una validación continua con expertos en la materia.

REFERENCIAS

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. Recuperado de <https://arxiv.org/abs/1810.04805>

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, 10(1), 1-309. doi:10.2200/S00762ED1V01Y201703HLT037

Pennington, J., Socher, R., Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. doi:10.3115/v1/D14-1162

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. Recuperado de <https://arxiv.org/abs/1706.03762>

ANEXOS A

Resultados de BERT

Archivo (RIT)	positivo	neutral	negativo	Total etiquetas	% positivo	% neutral	% negativo
128-2023	59	3	222	284	0,21	1,06	78,17
230-2023	10	0	102	112	8,93	0,00	91,07
418-2021	401	23	1206	1630	24,60	1,41	73,99
429-2022	26	2	153	181	14,36	1,10	84,53
585-23	22	0	201	223	9,87	0,00	90,13
603-2023	31	4	85	120	25,83	3,33	70,83
677-2023	38	1	80	119	31,93	0,84	67,23
780-2023	81	5	334	420	19,29	1,19	79,52
788-2023	23	5	89	117	19,66	4,27	76,07
301-2023	25	1	80	106	23,58	0,94	75,47
720-2023	22	1	84	107	20,56	0,93	78,50
230-2022	338	10	1248	1596	21,18	0,63	78,20
114-2023	88	4	202	294	29,93	1,36	68,71
117-2023	21	2	104	127	16,54	1,57	81,89
233-2020	167	9	1283	1459	11,45	0,62	87,94
252-2023	15	1	84	100	15,00	1,00	84,00
259-2023	30	0	98	128	23,44	0,00	76,56
273-2023	69	3	225	297	23,23	1,01	75,76
276-2022	75	2	294	371	20,22	0,54	79,25
279-2021	3	0	85	88	3,41	0,00	96,59
291-2022	25	2	112	139	17,99	1,44	80,58
323-2021	19	0	96	115	16,52	0,00	83,48
334-2023	20	1	114	135	14,81	0,74	84,44
365-2022	26	0	126	152	17,11	0,00	82,89
38-2022	54	5	203	262	20,61	1,91	77,48
383-2023	21	5	91	117	17,95	4,27	77,78
432-2022	28	1	162	191	14,66	0,52	84,82
432-2022	72	4	511	587	12,27	0,68	87,05
592-2023	11	2	129	142	7,75	1,41	90,85
91-2023	4	0	70	74	5,41	0,00	94,59
93-2022	45	1	183	229	19,65	0,44	79,91
2100331958-7	39	4	253	296	13,18	1,35	85,47

Figura A.1: Clasificación de sentimiento de los textos jurídicos con BERT

ANEXOS B

Resultados de BERTA

Archivo (RIT)	positivo	neutral	negativo	Total etiquetas	% positivo	% neutral	% negativo
128-2023	59	3	222	284	0,21	1,06	78,17
230-2023	10	0	102	112	8,93	0,00	91,07
418-2021	401	23	1206	1630	24,60	1,41	73,99
429-2022	26	2	153	181	14,36	1,10	84,53
585-23	22	0	201	223	9,87	0,00	90,13
603-2023	31	4	85	120	25,83	3,33	70,83
677-2023	38	1	80	119	31,93	0,84	67,23
780-2023	81	5	334	420	19,29	1,19	79,52
788-2023	23	5	89	117	19,66	4,27	76,07
301-2023	25	1	80	106	23,58	0,94	75,47
720-2023	22	1	84	107	20,56	0,93	78,50
230-2022	338	10	1248	1596	21,18	0,63	78,20
114-2023	88	4	202	294	29,93	1,36	68,71
117-2023	21	2	104	127	16,54	1,57	81,89
233-2020	167	9	1283	1459	11,45	0,62	87,94
252-2023	15	1	84	100	15,00	1,00	84,00
259-2023	30	0	98	128	23,44	0,00	76,56
273-2023	69	3	225	297	23,23	1,01	75,76
276-2022	75	2	294	371	20,22	0,54	79,25
279-2021	3	0	85	88	3,41	0,00	96,59
291-2022	25	2	112	139	17,99	1,44	80,58
323-2021	19	0	96	115	16,52	0,00	83,48
334-2023	20	1	114	135	14,81	0,74	84,44
365-2022	26	0	126	152	17,11	0,00	82,89
38-2022	54	5	203	262	20,61	1,91	77,48
383-2023	21	5	91	117	17,95	4,27	77,78
432-2022	28	1	162	191	14,66	0,52	84,82
432-2022	72	4	511	587	12,27	0,68	87,05
592-2023	11	2	129	142	7,75	1,41	90,85
91-2023	4	0	70	74	5,41	0,00	94,59
93-2022	45	1	183	229	19,65	0,44	79,91
2100331958-7	39	4	253	296	13,18	1,35	85,47

Figura B.1: Clasificación de sentimiento de los textos jurídicos con BERT