

**PROTOTIPO DE INTERFAZ INTERACTIVA DE ANÁLISIS DE EMOCIONES EN  
TEXTOS**

**Juan Sebastián Olivares Cortés**

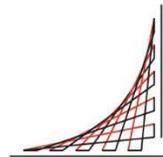
**Trabajo Dirigido**

**Tutor**

**Juan Manuel López López PhD  
Mónica Patricia Pineda Vargas**



**Universidad del  
Rosario**



**ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO**

**UNIVERSIDAD DEL ROSARIO  
ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ D.C  
2021**

## AGRADECIMIENTOS

En este espacio quiero agradecer inmensamente a Dios por haberme dado unos padres tan maravillosos los cuales me han apoyado a lo largo de mi vida en todas las formas posibles en las que se puede apoyar a un hijo, por brindarme siempre su cariño y amor incondicional y siempre estar ahí aun en medio de todas las dificultades y adversidades.

Gracias a Dios por mi papá Juan Francisco por recordarme cada día que en la vida todo se puede lograr con esfuerzo, dedicación y ganas de salir adelante aun así cuando parece que nada tiene sentido ni forma, a mi mamá Luz Marina por nunca dejarme ir a la universidad sin su bendición y sin mi coquita de almuerzo diario, por sostener absolutamente nuestro hogar y núcleo familiar mediante su orden y carisma particular.

A mi hermana Diana, por enseñarme que el valor de la vida siempre estará en una sonrisa y que la risa es la mejor anestesia ante cualquier dolor.

Gracias a Dios por poner en mi camino a María Fernanda Garzón, una mujer guerrera, humilde y echada para adelante que me enseñó a valorar cada esfuerzo que hacen todos los días mis padres, por forjar el valor del amor, respeto y unión en nuestra relación y por ser luz en mis días difíciles con tu cariño y una constante guía en mi diario vivir.

Infinitas gracias a Juan Manuel López por abrirme las puertas al semillero PROMISE desde que entré a la universidad, por depositar su confianza en mí y en mis capacidades como estudiante, porque más que un profesor ha sido un pilar fundamental en mi formación como persona y profesional en mi carrera universitaria.

A mis amigos, en especial: Checho, Mateito, Juanmi, Lau y Aleja con los cuales compartí momentos únicos, clases y proyectos. Por su carisma y forma de ser, por apoyarme y abrirme las puertas en sus vidas. Les agradezco por su amistad y les deseo lo mejor en sus carreras profesionales y su futuro.

A GrupoDot y en especial Mónica Pineda por darme la oportunidad de empezar mi carrera profesional, por su orientación y ayuda fundamental en el desarrollo de este proyecto dirigido junto con Juan Manuel López y por priorizar las ganas de aprender y la formación de perfiles profesionales antes que la experiencia.

A todos los miembros del semillero PROMISE, por su interés y participación en mejorar la calidad de educación en la universidad, por divulgar conocimiento y aprendizaje sin nada a cambio.

A la Escuela Colombiana de Ingeniería Julio Garavito y la Universidad del Rosario por hacer posible el programa de Ingeniería Biomédica y formar profesionales íntegros y competentes, capaces de afrontar las problemáticas y realidades a nivel nacional e internacional.

A todos ustedes, gracias por su confianza y apoyo incondicional en mi carrera profesional, que Dios se los retribuya en muchas bendiciones, éxitos y metas cumplidas para ustedes y sus familias.

## TABLA DE CONTENIDO

1. <b>INTRODUCCIÓN</b> .....	7
1.1 Modelos inteligencia artificial para procesamiento de lenguaje natural.....	13
2. <b>OBJETIVOS</b> .....	20
2.1 General.....	20
2.2 Específicos.....	20
3. <b>METODOLOGÍA</b> .....	21
3.1 Bases de datos.....	21
3.2 Consolidación de múltiples bases de datos.....	23
3.3 Procesamiento.....	26
3.4 Modelos de clasificación y regresión.....	27
3.4.1 Modelo de regresión.....	27
3.4.2 Modelo de clasificación.....	27
3.5 Integración del modelo generado con interfaz gráfica de usuario web.....	28
4. <b>RESULTADOS</b> .....	31
4.1 Modelo de regresión.....	31
4.2 Modelo de clasificación.....	32
4.3 Interfaz gráfica de usuario.....	35
5. <b>DISCUSIÓN</b> .....	40
6. <b>RECOMENDACIONES Y TRABAJOS FUTUROS</b> .....	42
7. <b>CONCLUSIONES</b> .....	42
8. <b>BIBLIOGRAFÍA</b> .....	44

## LISTA DE TABLAS

**Tabla 1.** DESCRIPCIÓN BASE DE DATOS ENCONTRADAS

**Tabla 2.** DESCRIPCIÓN BASE DE DATOS UNIFICADA

**Tabla 3.** DESEMPEÑO MEJORES MODELOS DE CLASIFICACIÓN DE EMOCIONES

## LISTA DE FIGURAS

**Figura 1.** Pronóstico de ingresos producto de soluciones en inteligencia artificial a nivel mundial desde 2018 a 2025 por regiones (en millones de dólares estadounidenses)

**Figura 2.** Ubicación del área de procesamiento de lenguaje natural dentro de la inteligencia artificial.

**Figura 3.** Pronóstico de ingresos relacionados con el área de procesamiento de lenguaje natural por categoría.

**Figura 4.** Modelo dimensional de emociones planteado por Russell.

**Figura 5.** Ejemplos de tokenización.

**Figura 6.** Ejemplo de generación de word embeddings usando la técnica bag of words.

**Figura 7.** Ejemplo de formación de distintos n-gramas a partir de un texto

**Figura 8.** Ejemplo de formación de incrustación ponderada de la palabra “stick” para dos capas LSTM.

**Figura 9.** Ejemplo de estructuras de incrustaciones del modelo GTP para distintas tareas de procesamiento de lenguaje natural.

**Figura 10.** Formación de la incrustación de entrada a partir de tokens individuales para el modelo BERT con dos frases A (my dog is cute) y B (he likes playing).

**Figura 11.** Comparación de modelos de arquitecturas pre-entradas de BERT, GPT y ELMo (de izquierda a derecha).

**Figura 12.** Comparación de desempeño de modelos para la prueba de rendimiento GLUE. ELMo (recuadro rojo), GPT (recuadro naranja) y BERT (recuadro verde).

**Figura 13.** Representación gráfica del número de parámetros en relación con algoritmos y modelos de procesamiento de lenguaje natural.

**Figura 14.** Transformación de datos numéricos a categóricos de la base de datos EMOBANK.

**Figura 15.** Esquema de funcionamiento interfaz gráfica de usuario web.

**Figura 16.** Diagrama de Gantt del proyecto.

**Figura 17.** Valores verdaderos y predichos de valencia y activación juntos para el conjunto de datos de prueba EMOBANK.

**Figura 18.** Valor F1 ponderado para cada una de las bases de datos utilizadas con cada método de preprocesamiento.

**Figura 19.** Valor F1 ponderado para la base de datos unificada con cada método de preprocesamiento.

**Figura 20.** Matriz de confusión para el conjunto de prueba del modelo entrenado con la técnica de remoción de palabras vacías.

**Figura 21.** Vista principal de usuario de la interfaz gráfica de usuario.

**Figura 22.** Sección “¡Clasifica tu emoción!” de la vista principal de la interfaz gráfica de usuario web.

**Figura 23.** Sección “Cómo se clasifican las emociones” de la vista principal de la interfaz gráfica de usuario web.

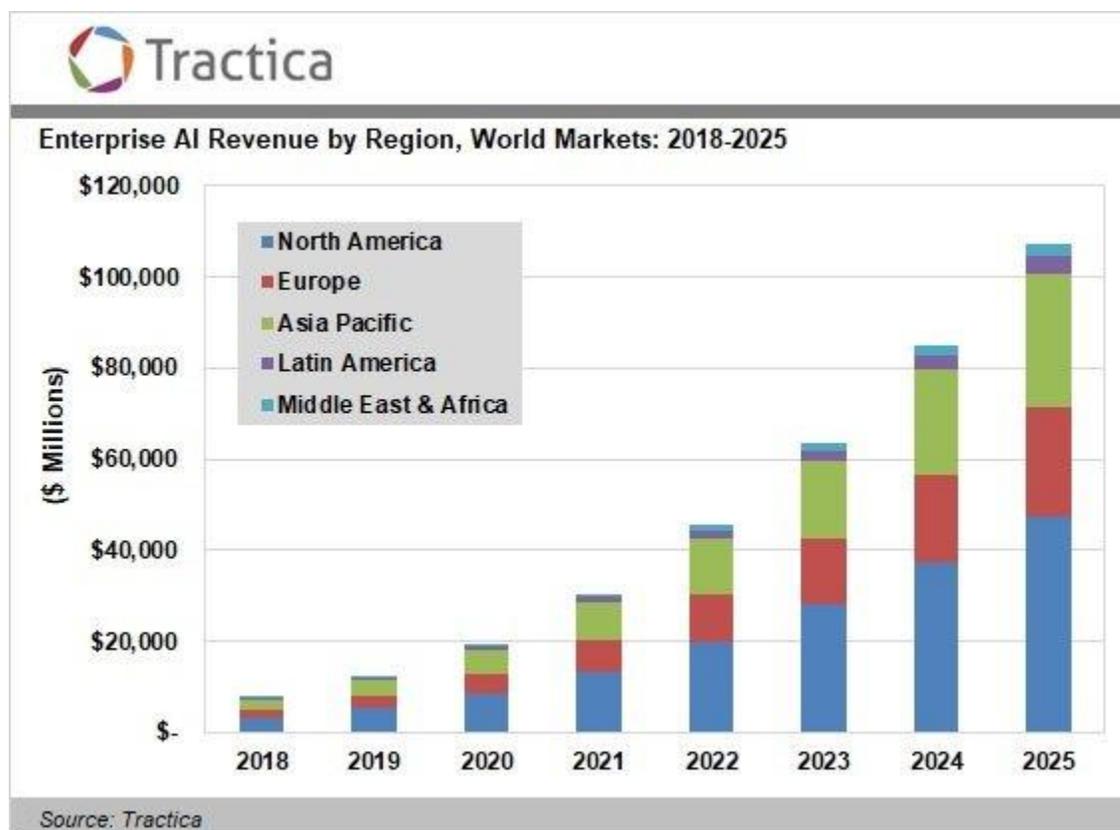
**Figura 24.** Sección “Por qué es importante analizar emociones” de la vista principal de la interfaz gráfica de usuario web.

**Figura 25.** Sección “¡Analiza y entiende tu resultado!” de la vista de análisis de la interfaz gráfica de usuario web.

**Figura 26.** Sección “Análisis por frase y entidades” de la vista de análisis de la interfaz, donde se puede hacer un análisis de sentimiento (valencia) del texto por frase.

## 1. INTRODUCCIÓN

El continuo avance tecnológico y el desarrollo de nuevas técnicas que permiten aplicar la inteligencia artificial cada vez en más campos, sin duda se ha convertido en base fundamental de las empresas para obtener mayores ingresos. Según Tractica<sup>1</sup>, firma de inteligencia de mercado enfocada en tecnologías emergentes, se pronostica que los ingresos a nivel mundial relacionados con inteligencia artificial y software se incrementarán desde 2018 a 2025 como se puede observar en la figura 1.



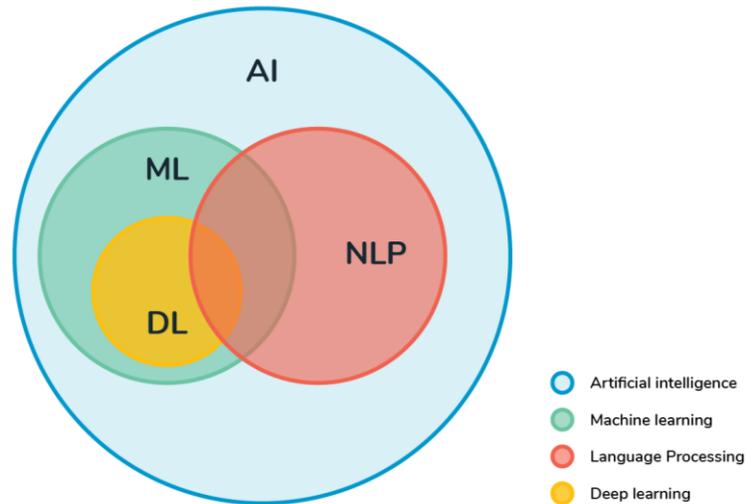
**Figura 1.** Pronóstico de ingresos producto de soluciones en inteligencia artificial a nivel mundial desde 2018 a 2025 por regiones (en millones de dólares estadounidenses). Tomada de [1].

Se estima que Estados Unidos será el mercado regional que experimentará el mayor crecimiento en el mercado alcanzando más de 50 billones de dólares en 2025, seguido de Europa con un crecimiento estimado de 26.54 billones de dólares, Asia con 32.89 millones de dólares, Latinoamérica con 4.57 billones de dólares y por último Medio Oriente y África con 3.03 billones de dólares [1].

Si bien, estas estimaciones están basadas a nivel general en inteligencia artificial, es decir incluyendo soluciones hechas con aprendizaje de máquina (Machine Learning), aprendizaje

<sup>1</sup> <https://www.crunchbase.com/organization/tractica>

profundo (Deep Learning) y sus categorías internas. Existe un campo en particular donde el desarrollo se ha incrementado sustancialmente en los últimos 5 años, este es el caso del área de procesamiento de lenguaje natural **PLN**, conocido también por sus siglas en inglés **NLP** (Natural Language Processing) [2].



**Figura 2.** Ubicación del área de procesamiento de lenguaje natural dentro de la inteligencia artificial. Tomado de [3].

De forma simple, el PLN pretende facilitar la interacción entre máquinas y humanos. Esta rama de la inteligencia artificial, que a su vez es un subcampo de la lingüística, trata de abstraer información del lenguaje, a partir del reconocimiento de patrones, brindando a los computadores las herramientas para entender, interpretar y manipular el lenguaje humano, desde su aplicación en tareas simples como el autocompletado de frases cuando escribimos un documento o llenamos un formulario, hasta la traducción de textos a otro lenguaje con reglas semánticas y sintácticas diferentes, extracción de perspectivas y análisis de inmensas cantidades de información de datos no estructurados como: correos electrónicos, documentos escaneados, publicaciones en redes sociales, minería de datos interfaces cerebro-computadora y psicología [4][5][8].

Si bien, la mayoría de los sectores económicos y de investigación se vieron afectados por la pandemia del COVID-19, la demanda de soluciones y servicios basadas en PLN se espera que crezca debido al aumento en la demanda de la mejora en la experiencia del servicio al cliente junto con la necesidad de soluciones que perfeccionen las operaciones internas de las empresas y disminuyan la tasa de abandono de clientes [6]. Además, se pronostica que el aumento en los ingresos relacionados con productos y servicios basados en PLN seguirán creciendo tanto a nivel de software como de hardware, pero teniendo su mayor impacto en la aplicación de ambos con la construcción de soluciones inteligentes y servicios para la industria como se puede ver en la figura 3.



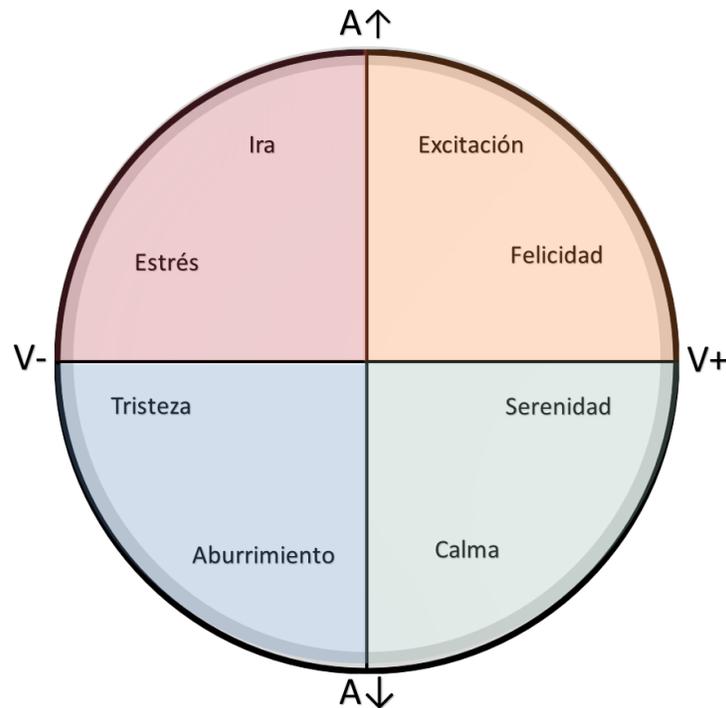
**Figura 3.** Pronóstico de ingresos relacionados con el área de procesamiento de lenguaje natural por categoría. Tomado de [7].

Además, dentro de las múltiples tareas que se pueden resolver utilizando PLN, está el campo de reconocimiento de emociones, el cual está ligado directamente a un problema actual y presente en la literatura cuya aplicación brinda soluciones en múltiples áreas de la industria y educación como lo son la minería de datos, interfaces cerebro-computadora y psicología [8].

Con el fin de explicar y clasificar emociones, desde siglos atrás se ha trabajado en modelos de emociones desde distintas perspectivas como lo son la filosofía con tratados filosóficos como es el caso de la obra “Pasiones del alma” en la cual se proponen y se investigan las pasiones primarias: maravilla, amor, odio, deseo, alegría, tristeza [9]. También, ha existido un trabajo muy profundo y constante desde la psicología con modelos principalmente basados en dimensiones que han sido tratados desde la década de los 80 como lo fue Wilhelm Wundtm, el cual postuló que las emociones pueden ser descritas bajo tres dimensiones, con el trabajo de Wundtm y la contribución de múltiples teorías construccionistas se generaron modelos posteriores que permitieron la consolidación de la definición de modelos dimensionales de emociones [10].

La base teórica sobre la cual los modelos dimensionales de emociones están contruidos parten del principio que todos los estados afectivos relacionados a las emociones están relacionados directamente con un sistema neurofisiológico común e interconectado [11]. De hecho, el modelo de emociones de Russell plantea que existen dos dimensiones bajo las cuales se pueden ubicar las emociones en relación con el nivel de placer y alerta asociado con la respuesta del sistema neurofisiológico y que la combinación lineal de estos niveles permite la clasificación de emociones, estos niveles se pueden ubicar de forma gráfica en un modelo circunplejo compuesto de dos ejes principales sobre los cuales se puede ubicar qué tan agradable o desagradable es la respuesta de los sistemas neurofisiológicos ante

un estímulo, lo cual se conoce como **valencia** y el incremento o la disminución de la respuesta de estos sistemas conocido como **activación** [12]. La valencia nos indica qué tan positivo o negativo es la sensación que estamos sintiendo y su va desde una valencia negativa hacia valencia positiva. Por otra parte, la activación está relacionada con el nivel de calma o alerta que tengo frente a un estímulo y su escala va desde baja activación hasta alta activación. El modelo de emociones planteado por Russell se puede observar en la figura 4.



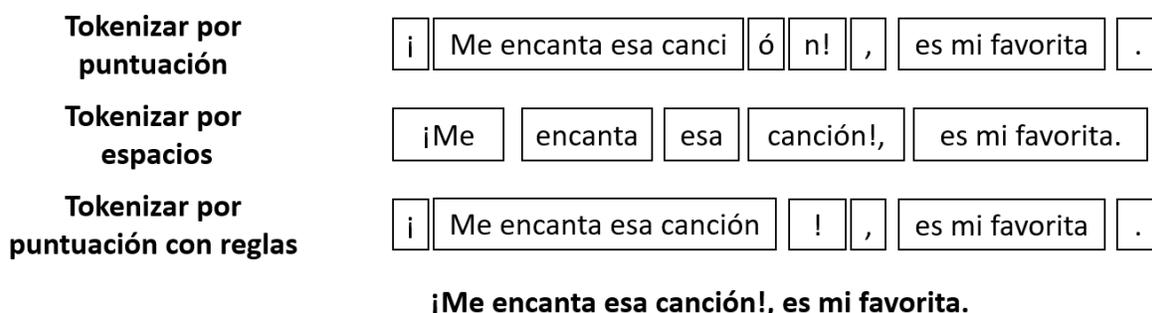
**Figura 4.** Modelo dimensional de emociones planteado por Russell. El eje horizontal corresponde a la **valencia** y el eje vertical representa la **activación**. Adaptado de [12].

Para los seres humanos, entender las emociones no es una tarea fácil y a menudo es muy complejo describir, interpretar y evaluar las emociones que estamos sintiendo. Reconocer las emociones es de vital importancia debido a que este proceso está relacionado con la toma de decisiones lo cual continuamente nos lleva a evaluar situaciones e identificar potenciales fuentes de conflicto y tener control al iniciar la respuesta a dichas acciones con el fin de resolver el conflicto y generar una respuesta adecuada frente al estímulo [13][14]. Por esta razón, generar un sistema que facilite la clasificación y reconocimiento de emociones en texto basado en un modelo dimensional principalmente fundamentado en un sistema neurofisiológico que regula estas emociones, no solo permite identificar la relación de la emoción general del texto con la respuesta neurofisiológica de nuestro cuerpo, sino que favorece a la mejora en la toma de decisiones e inclusive a la supervivencia del ser humano [11][14].

En este punto, ya sabemos que el uso y la importancia del PLN tiene un gran impacto a nivel mundial y que se puede aplicar en gran diversidad de tareas y de industrias como lo es el área de reconocimiento de emociones la cual es de vital importancia para el ser

humano, no solo con el fin de mejorar procesos internos de las empresas sino para la toma de decisiones. Para que las máquinas aprendan a tomar frases y partes de textos con el fin de capturar el sentido y significado de la oración y además tener la capacidad de hallar relaciones subyacentes entre estas frases y palabras es necesario realizar una transformación, específicamente una transformación numérica en la que las frases tengan una representación vectorial del texto crudo entendibles para una máquina. A este proceso se le llama tokenización, (*tokenization* en inglés) y es la base fundamental para que los sistemas de procesamiento de lenguaje natural puedan entender, abstraer y reconocer patrones en los textos a nivel de lenguaje de máquina.

Este proceso consiste en separar partes de un texto en partes más pequeñas llamadas *tokens*, los cuales pueden ser palabras, caracteres o partes de palabras [15]. Sin embargo, esta separación puede hacerse teniendo como base distintas reglas para dividir cada palabra como podemos ver en la figura 5.



**Figura 5.** Ejemplos de tokenización. Adaptado de [15].

Dependiendo de la regla que se escoja para dividir el texto se tendrán mayor o menor cantidad de *tokens* y como se puede intuir, existen infinidad de reglas que se pueden crear para separar estos *tokens*. Entonces, ¿qué reglas se deben escoger para tokenizar los textos? La respuesta es: depende del tipo de modelo y arquitectura de procesamiento de lenguaje natural y de la tarea que se quiera resolver (clasificación de texto, reconocimiento de entidades en un texto, generación de texto, resumen etc.).

Sin embargo, al obtener *tokens* aún se tienen palabras o partes del texto representadas por letras o caracteres, entonces, ¿cómo generar una representación vectorial que pueda ser entendida por una máquina? Así como existen distintas formas de tokenizar el texto, también existen varias técnicas para representar estos tokens de forma vectorial a lo cual se le conoce como *word embeddings*. Con el fin de comprender la importancia y relevancia de los *word embeddings* en el entrenamiento y desempeño de tareas relacionadas con el procesamiento de lenguaje natural, a continuación, hablaremos acerca de algunas técnicas de generación de *word embeddings*.

Una de las técnicas más simples para representar de forma vectorial un texto se llama bolsa de palabras, también conocida como *bag of words* en inglés. Esta técnica consiste en codificar el texto de entrada basado en la frecuencia con la cual las palabras o tokens aparecen dentro del texto generando un vector compuesto de números binarios en donde

el valor de 1 se asigna si el token está presente en el texto y 0 si no está presente. A continuación, podemos ver una representación gráfica de esta técnica.

Texto de entrada	Me	gusta	comer	hamburguesa	pizza	la	y
Me gusta comer hamburguesa	1	1	1	1	0	0	0
Me gusta comer pizza	1	1	1	0	1	0	0
Me gusta la hamburguesa y la pizza	1	1	0	1	1	2	1

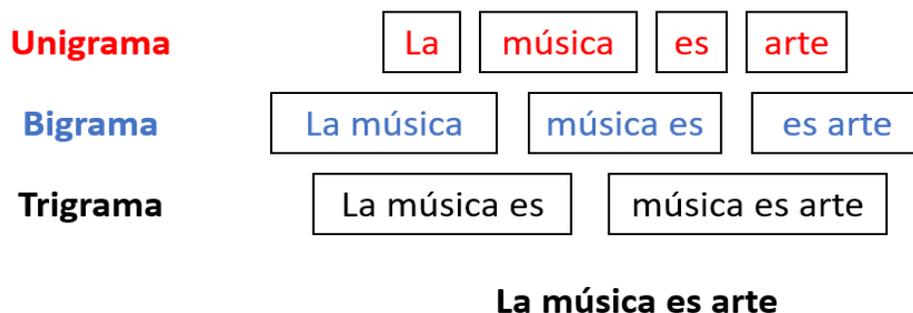
**Figura 6.** Ejemplo de generación de *word embeddings* usando la técnica *bag of words*.

Como se puede observar, con esta técnica se puede generar una representación vectorial bastante simple de un texto de entrada. Sin embargo, a medida que nuestro texto tiene mayor cantidad de tokens no repetidos, la dimensionalidad de los vectores aumenta. Además, palabras que se repiten dentro del texto tienden a ser palabras comunes como conectores, preposiciones y conjunciones lo que hace que de forma indirecta estas palabras tengan una mayor relevancia e importancia dentro del texto [16].

Con el fin de mejorar este problema, existen técnicas como TF-IDF conocido como *term frequency-inverse document frequency*. De forma sencilla, esta técnica se basa en determinar la frecuencia relativa de palabras en un documento y realizar una comparación con la proporción inversa de esa palabra sobre todo el texto. De tal manera que asigna un peso mayor a palabras relevantes dentro del texto que no están repetidas y les asigna un peso menor a palabras comunes en el texto como artículos y preposiciones. Esta técnica es de gran utilidad para sistemas de recuperación de información debido a que tiene en cuenta de forma directa las palabras clave dentro del texto y refleja que tan importantes son estas palabras en un documento o corpus. [17].

Es importante mencionar que, las técnicas mencionadas anteriormente no tienen en cuenta el contexto de un token o palabra dentro de la oración, es decir que son *word embeddings* sin contexto. Por ejemplo, en la oración: “**Sobre** la mesa está un libro que habla **sobre** animales, dentro del cual hay un **sobre** con tarjetas”, se puede ver que la palabra “sobre” está presente 3 veces con un significado distinto, “Sobre la mesa” como preposición, “sobre animales” como locución preposicional y “sobre con tarjetas” como sustantivo.

Para mejorar un poco este problema existen también técnicas como n-gramas o *n-grams* en inglés, las cuales de forma simple agrupan una secuencia n de palabras y asigna una probabilidad a la ocurrencia de un n-grama o la probabilidad de que una palabra aparezca justo después de una secuencia de palabras. Esta técnica permite un mejor desempeño en tareas como autocompletar texto y revisión de ortografía al tener en cuenta la representación de conjuntos de palabras lo cual brinda un mejor contexto [18]. Una representación gráfica de cómo se pueden obtener distintos n-gramas a partir de un texto se puede observar en la figura 7.



**Figura 7.** Ejemplo de formación de distintos n-gramas a partir de un texto.

Además, existen modelos más complejos de generación de *word embeddings* basados en modelos pre entrenados en una gran cantidad de datos (corpus) cuya ventaja principal es poder ser utilizados para distintas tareas de procesamiento de lenguaje natural. Actualmente, existen varios tipos de arquitecturas y modelos que se pueden utilizar para realizar tareas de procesamiento de lenguaje natural, cada arquitectura tiene un paradigma distinto y trabaja internamente de forma independiente. Para entender un poco más los fundamentos y las bases sobre las cuales están creados estos modelos, explicaremos un poco acerca de los siguientes algoritmos/arquitecturas: ELMo, GPT y BERT.

## 1.1 Modelos inteligencia artificial para procesamiento de lenguaje natural

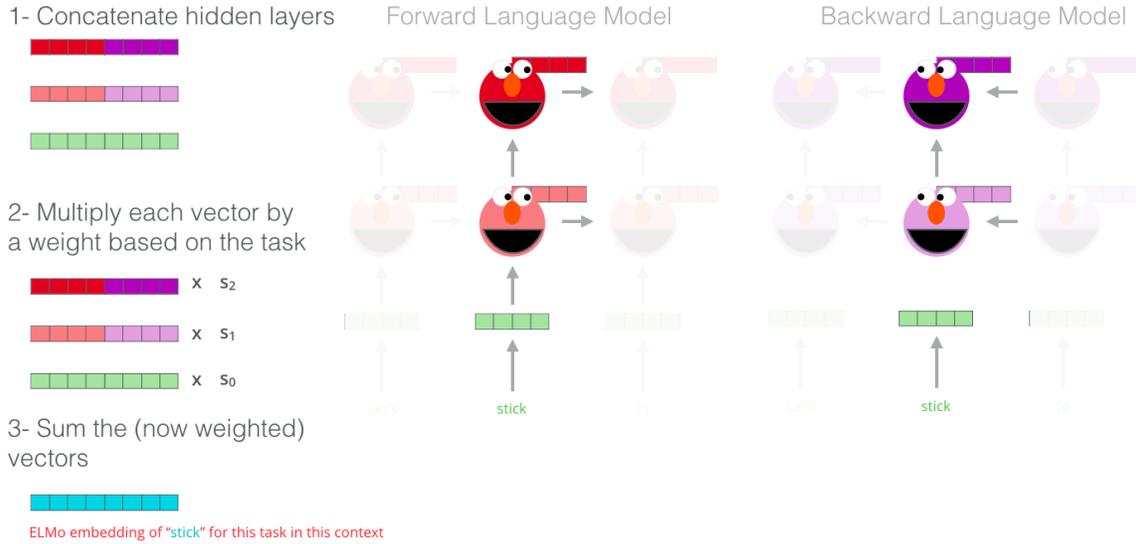
### ELMo

Este algoritmo fue uno de los primeros en adoptar el paradigma de tokenización y usar modelos pre entrenados de incrustación de palabras (*word embeddings*), cuya meta es inicializar la primera o primeras capas de un modelo de procesamiento de lenguaje natural para entrenar el modelo entero con representaciones jerárquicas de los tokens de entrada, lo cual permite a la máquina aprender una gran variedad de características desde las bases hasta conceptos de alto nivel semántico [19].

Su funcionamiento se basa en asignar una representación que es función de toda la frase de entrada a cada token. Para lograr esto, utilizan vectores pre-entrenados de una red bidireccional *Long-Short Term Memory* (LSTM) entrenada en un gran corpus lingüístico con el fin de resolver tareas de modelamiento de lenguaje, es decir de predecir la probabilidad de una secuencia de palabras. De ahí viene su nombre, *Embeddings from Language Models* (ELMo), incrustaciones de modelos de lenguaje [20].

Este modelo genera representaciones que son función de todas las capas internas de la red bidireccional LSTM y aprende combinaciones lineales de los vectores que son concatenados a medida que se avanza en las capas de la red LSTM y luego se genera una incrustación contextualizada al unir la incrustación inicial con la concatenación ponderada de la representación de las siguientes capas [19].

## Embedding of "stick" in "Let's stick to" - Step #2



**Figura 8.** Ejemplo de formación de incrustación ponderada de la palabra "stick" para dos capas LSTM. Tomado de [21]

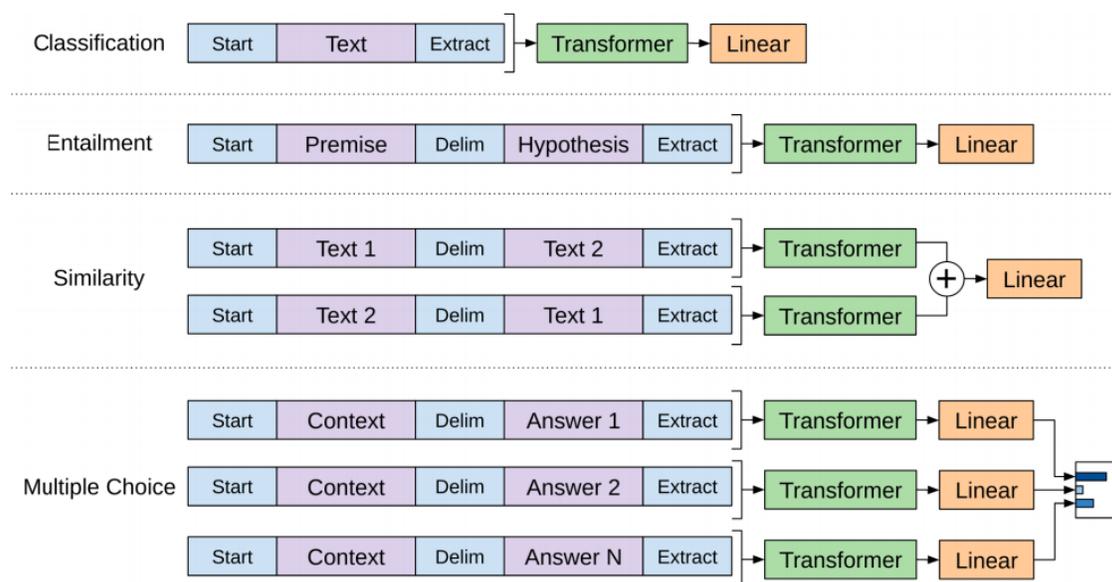
## GPT

Este modelo, creado por Open AI<sup>2</sup>, consiste en dos fases. Una fase de pre-entrenamiento no supervisado en la cual se busca modelar el lenguaje entrenando incrustaciones de palabras en un gran corpus lingüístico (tal como lo hace ELMo) y una fase de ajuste fino (*fine-tuning*) supervisado en donde se le enseña al modelo tareas discriminativas con datos anotados [22].

La diferencia con ELMo es que en vez de utilizar LSTM de dos capas, utiliza transformadores, los cuales son estructuras Encoder-Decoder (Codificador-Decodificador) que permiten un mejor manejo de las relaciones entre tokens a largo plazo mejor que los bloques LSTM, específicamente haciendo uso únicamente de los decodificadores con una capa de autoatención, conocida como *self-attention* en inglés. De ahí su abreviatura GPT (*Generative pre-training*) [21].

Teniendo en cuenta esta base, luego proponen hacer ajuste fino con la base ya pre-entrenada de modelamiento de lenguaje con decodificadores variando la estructura de los modelos con el fin de modificar la salida del modelo para dar solución a distintas tareas como lo es la clasificación de un texto en un tema específico, determinar qué tan similar es un texto en relación con otro y demás tareas como se puede ver en la figura 9.

<sup>2</sup> <https://openai.com/about/>



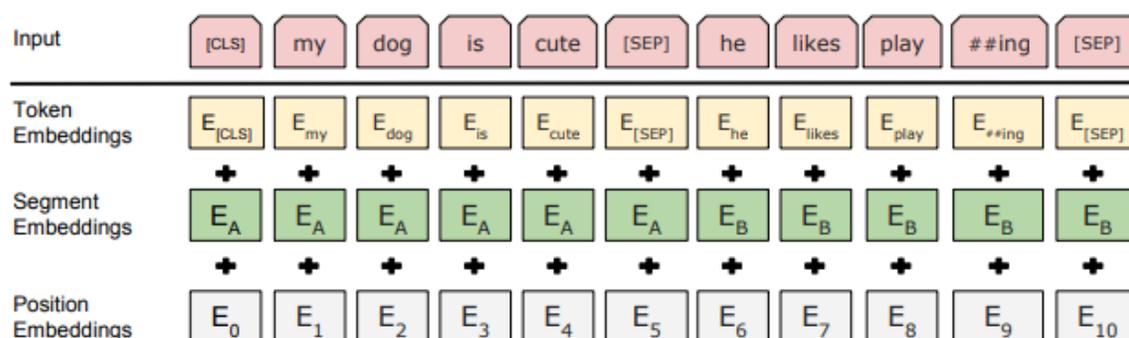
**Figura 9.** Ejemplo de estructuras de incrustaciones del modelo GPT para distintas tareas de procesamiento de lenguaje natural. Tomado de [22].

## BERT

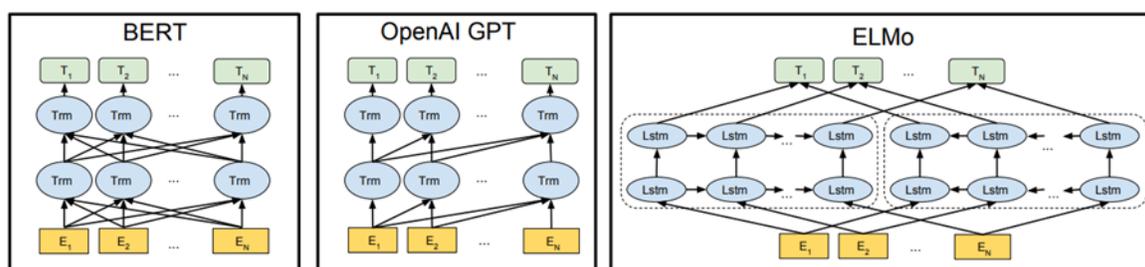
Por último, está el modelo *Bidirectional Encoder Representations from Transformers* (BERT), el cual fue el primer modelo profundo bidireccional de representación de lenguaje no supervisado, este fue entrenado usando solo un corpus lingüístico. Igual que GPT, tiene dos etapas: pre-entrenamiento y ajuste fino.

La principal diferencia entre BERT y GPT, es que los transformadores de BERT utilizan capas de autoatención bidireccionales, mientras que GPT utiliza capas de autoatención donde cada token solo puede entender contexto a la izquierda de la oración [23].

Además, BERT se enfoca en transformar cada token de entrada en una representación compuesta por incrustaciones de posición, segmento y token. Estas representaciones en conjunto permiten saber a qué frase corresponde cada token y además se utilizan tokens de clasificación especial para saber dónde empiezan y terminan las frases. Las incrustaciones de posición permiten obtener información relacionada con la posición a nivel global de la secuencia mientras que las incrustaciones de segmento permiten guardar posiciones relacionadas con cada oración [23]. La representación gráfica de esta descripción se puede ver en la figura 10 y la diferencia entre los modelos de arquitecturas pre entrenadas en la figura 11.



**Figura 10.** Formación de la incrustación de entrada a partir de tokens individuales para el modelo BERT con dos frases A (my dog is cute) y B (he likes playing). Tomado de [23].



**Figura 11.** Comparación de modelos de arquitecturas pre-entradas de BERT, GPT y ELMo (de izquierda a derecha). Adaptada de [23].

Ya tenemos un conocimiento general sobre cómo trabaja cada modelo y los fundamentos básicos sobre los cuales están contruidos para poder ser utilizados en distintas tareas de procesamiento de lenguaje natural. Pero ¿qué tan buenos son?, ¿cuál es su desempeño? Actualmente existe una prueba de rendimiento llamada *General Language Understanding Evaluation* (GLUE) [24], la cual fue construida para entrenar, evaluar y analizar sistemas de procesamiento de lenguaje natural respecto a una gran cantidad de fenómenos lingüísticos y a un diverso rango de tamaño de bases de datos, tipos de texto y grados de dificultad. Los resultados para estos tres algoritmos en las diversas tareas de la prueba GLUE se pueden observar en la figura 12.

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	Average
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT <sub>BASE</sub>	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT <sub>LARGE</sub>	<b>86.7/85.9</b>	<b>72.1</b>	<b>92.7</b>	<b>94.9</b>	<b>60.5</b>	<b>86.5</b>	<b>89.3</b>	<b>70.1</b>	<b>82.1</b>

**Figura 12.** Comparación de desempeño de modelos la prueba de rendimiento GLUE. ELMo (recuadro rojo), GPT (recuadro naranja) y BERT (recuadro verde). Adaptada de [20].

Como se puede observar en la figura 12, por cada prueba individual incluida en la prueba de rendimiento GLUE se observan las métricas de desempeño y el promedio sobre todas estas métricas para los modelos ELMo, GPT y BERT (base y large). Las tareas sobre las cuales trabaja GLUE y específicamente las que aparecen en la figura 12 de izquierda a derecha desde *Multi-Genre Natural Language Inference Corpus* (MNLI-(m/mm)) hasta *Recognizing Textual Entailment* (RTE) se explican de forma resumida a continuación:

**MNLI:** Esta tarea consiste en predecir si una premisa implica una hipótesis, es decir si la relación entre la verdad de la premisa implica que la hipótesis sea verdadera. La salida del modelo debe predecir si existe implicación, contradicción o ninguna de las dos [24].

**QQP:** Esta tarea consiste en determinar si dos oraciones son semánticamente equivalentes [24].

**QNLI:** Esta tarea consiste en determinar si el contexto dado por una oración contiene la respuesta a una pregunta dentro de un conjunto de preguntas. Cabe resaltar, que esta tarea fue modificada respecto a la original mediante la eliminación del requisito de que el modelo seleccione la respuesta exacta, de que la respuesta siempre está presente en la entrada y que la superposición léxica entre oraciones es una señal confiable [24].

**SST-2:** Esta tarea consiste en predecir el sentimiento de una oración usando la separación de clases positivo y negativo, usando etiquetas de clase en vez de un modelo de regresión [24].

**CoLA:** Esta tarea consiste en determinar juicios de aceptabilidad extraídos de libros y artículos de revistas sobre teoría lingüística etiquetados como textos gramaticales o no gramaticales [24].

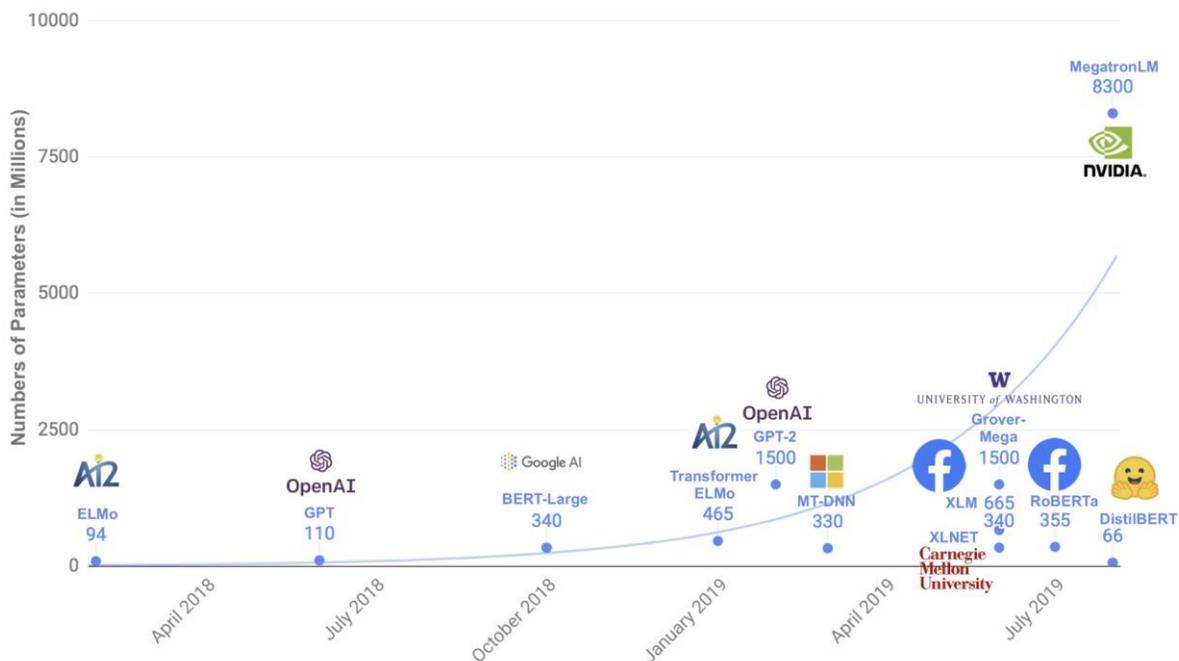
**STS-B:** Esta tarea consiste en identificar en un rango de 1-5 la similitud que tienen dos frases [24].

**MRPC:** Esta tarea consiste en identificar si una oración es parafraseo de otra oración [24].

**RTE:** Esta tarea consiste en reconocer, a partir de dos oraciones, si el significado de un texto se puede inferir del otro texto, es similar a MNLI [24].

Cabe resaltar que cada tarea tiene un número de datos diferente y también una métrica sobre la cual se analiza el desempeño del modelo la cual depende del balanceo de clases dentro de la base de datos y de la naturaleza de la tarea. Sin embargo, de forma general el valor promedio del desempeño de todas las tareas nos brinda una referencia sobre la cual entre más cercano esté a 100, mejor desempeño tiene el modelo en promedio para resolver tareas de procesamiento de lenguaje natural, en este caso el modelo “BERT large” tiene el mejor desempeño como se observa en la figura 12.

Cada arquitectura tiene una base sobre la cual trabaja e interactúa con los datos para poder obtener la salida deseada. A partir de estas arquitecturas, empresas como Microsoft, Facebook y NVIDIA también han empezado a investigar y generar nuevos modelos entrenados cada vez con más datos y parámetros con el fin de obtener mejores resultados. La relación entre el número de parámetros y cada modelo en función del tiempo se puede ver en la figura 13.



**Figura 13.** Representación gráfica del número de parámetros en relación con algoritmos y modelos de procesamiento de lenguaje natural. Tomado de [25].

La tendencia, a medida que se avanza en investigación y desarrollo de soluciones con PLN, nos indica que cada vez se utiliza un conjunto más grande de datos. Esto se debe a que gran parte de los modelos que hacen parte de la figura 13, están contruidos con arquitecturas profundas las cuales por lo general mejoran su desempeño y calidad a medida que se entrena con más datos. Sin embargo, esto no siempre es verdad debido a que la variabilidad de carácter semántico, sintáctico y el idioma en los textos influye de forma significativa en el entrenamiento de estos modelos [26]. Además, se puede ver que en relación con los modelos explicados existen variaciones que han sido implementadas de los modelos originales ELMo, GPT y BERT.

La motivación de este proyecto surge a partir de la necesidad de generar aplicaciones que permitan resolver problemas de procesamiento de lenguaje natural para las distintas demandas de la industria e investigación, específicamente en el área de clasificación de emociones en texto haciendo uso de diversas bases de datos documentadas en el estado del arte con el fin de obtener un modelo más robusto y un prototipo de interfaz gráfica que

permita realizar el análisis del sentimiento y clasificación de la emoción que le permita al usuario tener un mejor entendimiento y comprensión de la clasificación de emociones.

El proyecto dirigido se basa en el área de procesamiento de lenguaje natural y se desarrolla gracias a las herramientas en Google Cloud Platform brindadas por la empresa Grupodot<sup>3</sup>, junto con la colaboración y asesoría del doctor en ingeniería Juan Manuel López y la líder del grupo de ciencia de datos de Grupodot, Mónica Pineda Vargas.

El documento consta de la siguiente distribución de capítulos:

- El capítulo dos presenta los objetivos, uno general y tres específicos del proyecto.
- El capítulo tres contiene la metodología, en la cual se describe ampliamente los datos utilizados para llevar a cabo el proyecto, junto con el esquema de procesamiento, el entrenamiento del modelo de inteligencia artificial y por último la integración del modelo en una interfaz gráfica de usuario web.
- El capítulo cuatro presenta los resultados obtenidos a partir de la metodología empleada.
- El capítulo cinco explica el análisis de los resultados obtenidos en los distintos modelos de inteligencia artificial entrenados.
- El capítulo seis se enfoca en las conclusiones a las cuales se llega en esta investigación.
- Finalizando con el capítulo siete, el cual realiza recomendaciones para futuros trabajos en el área de procesamiento de señales.

---

<sup>3</sup> <https://www.grupodot.com/en/>

## **2. OBJETIVOS**

### **2.1 General**

Desarrollar un prototipo de interfaz que brinde apoyo en la interpretación de contenido emocional en textos, permitiendo al usuario entender de forma cuantitativa y gráfica la emoción de un escrito.

### **2.2 Específicos**

1. Proponer y aplicar una metodología que permita la utilización de diversas bases de datos de textos, para la detección de emociones básicas, utilizando técnicas de PLN.
2. Desarrollar y evaluar modelos de clasificación de emociones en texto, utilizando técnicas modernas de PLN.
3. Integrar, a través de una Interfaz Gráfica de Usuario, el modelo de clasificación de emociones, indicando gráficamente parámetros de valencia y activación emocional.

### 3. METODOLOGÍA

En esta sección se describe detalladamente el proceso que se siguió para la implementación de la interfaz de análisis emocional de texto. Se explican las bases de datos encontradas para entrenar el modelo, la consolidación y unificación de estas bases de datos, preprocesamiento y procesamiento de los textos que conforman las bases de datos, aproximación a un modelo de regresión y modelos de clasificación usando BERT, y por último la integración del modelo generado con interfaz gráfica de usuario web.

#### 3.1 Bases de datos

El reconocimiento de emociones en texto es de gran utilidad en varios campos como la minería de datos, psicología y sistemas de interacción humano-computadora, entre otros [8]. Sin embargo, para poder generar modelos que sirvan de apoyo en la solución de estos problemas, se necesitan datos para poder realizar el entrenamiento, validación y prueba del modelo. Teniendo en cuenta esta consideración, se realizó una búsqueda bibliográfica del estado del arte en bases de datos de uso libre para clasificación de emociones. A continuación, se explica de forma resumida cada base de datos encontrada.

##### EMOBANK

Esta base de datos fue creada por el laboratorio JULIE (Jena University Language & Information Engineering laboratory) y consiste en un registro de 10.062 frases tomadas de blogs, ensayos, cartas, guías de viaje, periódicos y libros de ficción las cuales fueron anotadas respecto a las dimensiones de valencia, activación y dominancia haciendo uso de CrowdFlower. Por cada frase dentro de la base de datos, 5 evaluadores califican la frase por cada dimensión de la emoción (valencia, activación y dominancia) y también desde dos perspectivas, la del lector y la del escritor con base en el modelo SAM (**Self Assessment Manikin**) en una escala de 1 a 5, siendo uno el menor valor y cinco el mayor valor dentro de cada dimensión. La base de datos fue evaluada por personas del Reino Unido, Estados Unidos, Irlanda, Canadá o Australia. Luego se genera un promedio ponderado de las calificaciones por frase y se obtienen los valores finales de valencia, activación y dominancia por frase [27].

##### SEMEVAL 2018

Esta base de datos fue creada con el fin de generar un reto compartido con el fin de inferir el estado emocional de una persona a partir de tweets. El reto consiste en 5 tareas las cuales son: estimar la intensidad de la emoción individual y por clases ordinales, estimar la valencia de la emoción por clases ordinales e individual y por último clasificación de emociones. Para la anotación de la base de datos se hace uso de la técnica Best-Worst Scaling (BWS) con el fin de evitar inconsistencias inter e intra sujeto. Alrededor de 700,000 anotaciones fueron generadas aproximadamente para 22,000 tweets en los idiomas inglés, árabe y español a partir de términos asociados con las emociones a distintas intensidades haciendo el uso del tesoro Roget<sup>4</sup> [28].

---

<sup>4</sup> <http://www.roget.org/>

## **WASSA 2017**

Esta base de datos es similar a SEMEVAL 2018 desde el aspecto en que también fue lanzada con el fin de generar un reto para detectar la intensidad de la emoción en tweets y también hacen uso de la técnica Best-Worst Scaling (BWS). De hecho, esta base de datos fue creada por el mismo grupo de investigadores el año anterior, pero está conformada únicamente por tweets en inglés de las emociones miedo, ira, tristeza y alegría. Con la construcción de esta base de datos se creó un sistema para determinar que los léxicos afectivos relacionados con la asociación entre palabras y emociones son útiles para determinar la intensidad de las emociones [29].

## **EMOTION DATASET FOR NLP**

Esta base de datos es de fácil acceso y de uso libre bajo la licencia Creative Commons [30]. Si bien, no tiene un artículo asociado desde la descripción de la base de datos en Kaggle se menciona que la técnica para preparar la base de datos es basada en el artículo de conferencia CARER (Contextualized Affect Representations for Emotion Recognition) [31]. El cual es un algoritmo semi supervisado que permite generar descriptores estructurales que permiten crear textos con representaciones contextualizadas. Sin embargo, las frases utilizadas para este algoritmo fueron obtenidas utilizando la API de Twitter, generando conjuntos de tweets subjetivos (clasificados a partir de hashtags) y objetivos (a partir de publicaciones de cuentas de Twitter de noticias).

## **GO EMOTIONS**

Esta base de datos está compuesta por comentarios de la red social Reddit tomados desde 2005 hasta enero de 2019. Para obtener la base de datos final, los investigadores de esta base generaron una lista que tiene frases y palabras con contenido adulto y ofensivo, vulgar, de identidad y términos religiosos y basados en comentarios que tuvieran este contenido filtraban y eliminaban las frases de forma automática y no las tenían en cuenta. Además, por cada frase se le asignaron 3 calificadores (hablantes nativos de inglés) para poder anotar la base de datos la cual tiene cerca 58000 registros anotados para una o más de 27 emociones e inclusive neutral [32].

En la tabla I se puede observar la cantidad total de frases por cada base de datos junto con las emociones que la componen y las fuentes bajo las cuales fueron tomadas.

**Tabla I**  
DESCRIPCIÓN BASE DE DATOS ENCONTRADAS

<b>Nombre</b>	<b>Fuente</b>	<b>Emociones</b>	<b>Cantidad total de datos</b>
<b>EMOBANK</b> [24]	Frases tomadas de blogs, ensayos, cartas, guías de viaje, periódicos y libros de ficción	Ira, alegría, tristeza, calma	10.000
<b>SEMEVAL 2018</b> [25]	Tweets en inglés	Ira, anticipación, asco, miedo, alegría, amor, optimismo, pesimismo, tristeza, sorpresa, confianza	10.983
<b>WASSA 2017</b> [26]	Tweets en inglés	Ira, miedo, alegría, tristeza	8.076
<b>GO emotions</b> [29]	Comentarios de Reddit	Admiración, aprobación, molestia, gratitud, desaprobación, diversión, curiosidad, amor, optimismo, decepción, alegría, realización, ira, tristeza, confusión, cariño, emoción, sorpresa, disgusto, deseo, miedo, remordimiento, vergüenza, nerviosismo, orgullo, alivio, dolor	58000
<b>Emotions dataset for NLP</b> [27]	Tweets en inglés	Ira, alegría, tristeza, miedo, amor, sorpresa	16.000
<b>Total</b>			<b>103.059</b>

### 3.2 Consolidación de múltiples bases de datos

Si bien, fue posible obtener varias bases de datos que permiten entrenar modelos de inteligencia artificial enfocados a las tareas de clasificación de emociones, es importante mencionar que cada base de datos fue creada a partir de diferentes fuentes como lo son: redes sociales, ensayos, viajes etc. Cada base de datos viene en un formato diferente, es decir no tienen un estándar general para almacenar estas frases, sino que cada base de datos tiene un número independiente de columnas. Por ejemplo, la base de datos EMOBANK está almacenada en formato CSV (separados por comas) y las columnas que

la componen son: identificador de la frase, división, valor de valencia, valor de activación, valor de dominancia y la frase, mientras que la base de datos SEMEVAL 2018 esta almacenada en formato TXT (texto plano) y tiene las columnas: identificador, tweet (frase) y una columna por cada emoción donde aparece el número 1 si está presente la emoción y 0 donde no está presente.

Teniendo en cuenta que las bases de datos no solo difieren en su contenido sino en la forma y formato en la cual están almacenadas y que existe una mayor cantidad de bases de datos para problemas de clasificación [33], es necesario realizar un proceso de consolidación para poder obtener una base de datos general con un mismo formato que incluya todas las bases de datos encontradas para poder entrenar el modelo de inteligencia artificial. Para este proceso se transforma cada base de datos en formato CSV (archivo separado por comas) y se deja únicamente las columnas que corresponden a texto y emoción a clasificar teniendo en cuenta solamente las emociones básicas que hacen parte del modelo de emociones de Russell, es decir alegría, tristeza, calma e ira.

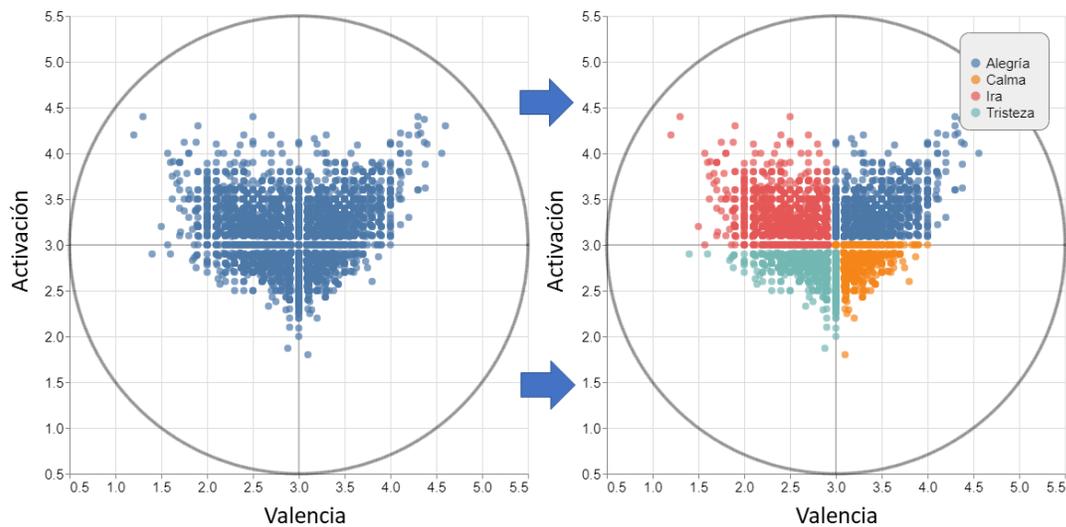
Si bien, el modelo de emociones planteado por Russell ubica las emociones en cuadrantes, basados en los valores de valencia y activación, es posible utilizar los valores cuantitativos dados por la escala SAM de 1 a 5 de la base de datos EMOBANK para poder convertir el paradigma en un problema de clasificación y no descartar esta base de datos. Para poder lograr esto es necesario tener un conocimiento previo acerca de la escala SAM para poder transformar los valores de valencia y activación en emociones el cual ha sido obtenido a partir de mi participación como estudiante investigador en proyectos interdisciplinarios como el desarrollado en conjunto entre la Escuela Colombiana de Ingeniería Julio Garavito y la Corporación Universitaria Minuto de Dios, llamado “Reconocimiento Emocional, Memoria Emocional y respuesta Fisiológicas en Víctimas de Violencia de Pareja”<sup>5</sup> sobre el cual también se han realizado trabajos relacionados con la señal de electroencefalografía para determinar si existen diferencias en bandas de potencia de esta señal para mujeres que han vivido experiencias de violencia de pareja y mujeres que no [43]. El eje central de la escala se tomó como valor de valencia igual a 3 y activación igual a 3 y las emociones se clasificaron de la siguiente forma:

- **Alegría:** Valores de valencia mayores e iguales a 3, con activación mayor a 3.
- **Tristeza:** Valores de valencia menores e iguales a 3, con activación menor a 3.
- **Calma:** Valores de valencia mayores que 3, con activación menor o igual que 3.
- **Ira:** Valores de valencia menores que 3, con una activación mayor o igual que 3.

La representación gráfica de esta transformación de datos de regresión a categorías se puede ver en la figura 14.

---

<sup>5</sup> <https://sites.google.com/view/semilleropromise/proyectos/proyviolencia>



**Figura 14.** Transformación de datos numéricos a categóricos de la base de datos EMOBANK.

Respecto a las otras bases de datos se hace un filtrado para poder obtener únicamente las frases que tienen contenido emocional basado en el modelo de emociones de Russell. En la tabla II se puede observar la descripción de la base de datos unificada con la cantidad de frases relacionada a cada base de datos.

**Tabla II**  
DESCRIPCIÓN BASE DE DATOS UNIFICADA

Nombre	Ira	Alegría	Tristeza	Calma	Total
<b>EMOBANK</b>	2.549	2.594	2.630	2.289	10.062
<b>SEMEVAL 2018</b>	3.960	4.319	3.233	0	11.512
<b>WASSA 2017</b>	1.701	1.616	1.533	0	4.850
<b>GO emotions</b>	8.242	8.034	6.549	0	22.825
<b>Emotions dataset for NLP</b>	2.709	6.761	5.797	0	15.267
<b>Suma</b>	<b>19.161</b>	<b>23.324</b>	<b>19.742</b>	<b>2.289</b>	<b>64.516</b>

Con esta base de datos unificada no sólo se obtiene mayor cantidad de datos para entrenar el modelo, sino que se entrena con una mayor variabilidad de textos debido a que están tomados de distintas fuentes como lo son redes sociales, libros, revistas etc. Sin embargo, se debe tener en cuenta que el contenido de redes sociales puede incluir palabras que no deberían ser tenidas en cuenta para la clasificación de la emoción como lo son hashtags, menciones y caracteres especiales como emoticones y puntuación, para poder eliminar estos contenidos se utilizan expresiones regulares, también conocidas como regex. Cabe resaltar, que no existe una combinación única de expresiones regulares o de técnicas de preprocesamiento para cada clase de texto y se debe hacer un análisis exploratorio de estas técnicas para hallar la que mejor se ajuste a la tarea [34].

### 3.3 Procesamiento

Debido a que cada frase dentro de la base datos tiene una estructura semántica y sintáctica diferente es necesario realizar un preprocesamiento y procesamiento a estas oraciones con el fin de uniformizar los datos que se le entregan al modelo de inteligencia artificial para clasificación de emociones. El esquema de preprocesamiento de los textos que conforman las bases de datos se divide en los siguientes pasos:

1. Remover las menciones, hashtags y enlaces del texto.
2. Quitar los signos de puntuación
3. Remover los emoticones

Después de realizar este preprocesamiento a todas las frases de la base de datos unificada e individual se propone un esquema de procesamiento el cual es aplicado a estas frases preprocesadas para poder analizar cuál técnica de procesamiento se ajusta mejor teniendo en cuenta que no hay una técnica general [34]. Para esta parte se hace uso de dos técnicas de procesamiento en texto las cuales son lematización y radicalización también conocidas como **lemmatization** y **stemming** en inglés las cuales son técnicas que se utilizan para generar la forma base de un conjunto de palabras independientemente de sus tiempos verbales.

La lematización se enfoca en el uso correcto del vocabulario junto con el análisis morfológico de las palabras. El objetivo de realizar este proceso es eliminar solo las terminaciones de las palabras que hacen que cambie el número o el tiempo verbal de una palabra base (terminaciones flexivas) [35] y devolver la forma base o de diccionario de una palabra, lo que se conoce como **lema** [36].

Por otra parte, la radicalización se enfoca en un proceso cuyo objetivo es cortar los extremos de las palabras con el fin de reducir la cantidad de veces en la cual se puede presentar una palabra [36].

Otra técnica de procesamiento es remover las palabras de parada, también conocidas como **stopwords** en inglés. El objetivo de utilizar esta técnica es disminuir la cantidad de tokens que no son útiles para determinar el contexto o el significado de una oración. Cabe resaltar que, estas frases pueden variar dependiendo del corpus que se utilice y el problema que se esté analizando. Sin embargo, los problemas enfocados a la clasificación de textos por lo general no requieren palabras de parada debido a que estas no brindan información relevante para entender el tema general de un texto [37]. Por último, se tokenizan las frases

para poder entrenar los modelos de inteligencia artificial mencionados en la siguiente sección.

### **3.4 Modelos de clasificación y regresión**

En esta fase se propuso hacer una aproximación a modelos de regresión y clasificación mediante el ajuste fino a la arquitectura “BERT base uncased”<sup>6</sup> proporcionada por la comunidad Hugging Face<sup>7</sup>. Esta arquitectura fue utilizada debido al buen rendimiento que tiene respecto a otras arquitecturas como ELMO y GPT [23]. Además, es de acceso libre y permite crear, entrenar y desplegar modelos de procesamiento de lenguaje natural y actualmente es utilizada por más de 5000 organizaciones entre las cuales se encuentran Google AI, Facebook y Amazon.

#### **3.4.1 Modelo de regresión**

En esta fase se realizó una primera aproximación a un modelo de regresión para poder estimar la emoción a partir de un modelo que estime los valores de valencia y activación haciendo únicamente uso de la base de datos EMOBANK (debido a que es la única base de datos encontrada que tiene estas características). Se realizó el ajuste fino de la arquitectura BERT con los hiper parámetros mencionados anteriormente. Sin embargo, fue necesario realizar cambios en esta arquitectura para que la salida sea un valor de valencia y activación por frase en vez de una etiqueta de emoción (ira, alegría, tristeza o calma). Se añadió una capa de abandono con un porcentaje de abandono del 50%, una función de activación tipo ReLu y dos capas densas totalmente conectadas para que el modelo tuviera dos valores de salida, uno para estimar la valencia y otro la activación. La función de pérdida utilizada fue el error cuadrático medio y se utilizó la distribución original de la base de datos para el conjunto de frases de entrenamiento, validación y prueba por una cantidad de 15 épocas. Los resultados del desempeño de este modelo se pueden ver en la sección de resultados.

#### **3.4.2 Modelo de clasificación**

Para esta fase se realizó un análisis y creación de modelos de clasificación para cada base de datos y la base de datos unificada teniendo en cuenta las distribuciones de datos de entrenamiento, prueba y validación de cada base de datos individual y una distribución de 80% de datos para entrenamiento, 10% para validación y 10% para prueba en la base de datos unificada guardando la misma proporción de frases para cada categoría para lidiar con el desbalance de clases. Además, cada base de datos fue entrenada por 15 épocas con un tamaño de lote (batch size) igual a 32, haciendo uso del optimizador Adam y un programador lineal de calentamiento (warmup linear scheduler) para variar la tasa de aprendizaje con el fin de reducir el impacto del entrenamiento del modelo cuando se le muestran nuevos datos de validación a medida que avanza el número de épocas [38]. La métrica para medir el desempeño escogida fue el puntaje F1 ponderado debido a que como las clases no están balanceadas (en especial por la clase calma), esta métrica nos brinda una magnitud ponderada en la cual todas las clases son importantes, pero se tiene en

---

<sup>6</sup> <https://huggingface.co/bert-base-uncased>

<sup>7</sup> <https://huggingface.co/>

cuenta la cantidad de datos asociados para cada clase [39]. Además, los datos de entrenamiento, prueba y validación fueron estratificados con el fin de tener la misma proporción de datos por clase durante el entrenamiento del modelo y así tratar de balancear la diferencia entre la cantidad de datos por clase. Los resultados de los entrenamientos por base de datos individual y base de datos unificada se pueden ver en la sección de resultados.

### 3.5 Integración del modelo generado con interfaz gráfica de usuario web

Luego de entrenar los modelos de inteligencia artificial, hacer un análisis exploratorio y obtener el mejor modelo de clasificación de emociones se procede a realizar el diseño, implementación e integración del modelo en una interfaz web haciendo uso de un servidor local creado a partir del framework FastAPI<sup>8</sup>. Cabe resaltar que la fase de análisis exploratorio de datos, unificación de las bases de datos y el entrenamiento de BERT fue hecho en Python 3.8, el entrenamiento del modelo se hizo con PyTorch<sup>9</sup> y la librería Transformers<sup>10</sup> de Hugging Face.

Para esta fase y con el fin de tener un código más estructurado, legible y fácil de integrar con FastAPI, se realizó una refactorización de código en programación orientada a objetos generando la escritura de una clase que se encarga de cargar el modelo final junto con los pesos base de BERT, cargar el conjunto de palabras vacías a remover usando la librería NLKT<sup>11</sup>, crear una instancia cliente de la API de traducción de Google<sup>12</sup> para poder hacer el análisis de emociones traduciendo desde cualquier lenguaje a inglés (el idioma en el cual está entrenado el modelo de clasificación) y por último otra instancia cliente de procesamiento de lenguaje natural de Google<sup>13</sup> para poder hacer análisis de sentimiento en texto y reconocimiento de entidades. Cabe resaltar que para hacer uso de las APIs que proporciona Google es necesario crear una cuenta de servicio que tenga permisos para estas instancias con el fin de autenticar el servicio y poder usarlas. Además, se hizo uso de la librería Altair<sup>14</sup> para generar gráficas interactivas con el fin de mostrar análisis de sentimiento por frases haciendo uso de la API de procesamiento de lenguaje natural de Google. Para la parte visible por el usuario se usó HTML, CSS, Bootstrap y Javascript. El diagrama de la figura 15 muestra como es el proceso mediante el cual se transmite toda la información desde la vista del usuario (azul) pasando por el servidor local hasta la parte interna donde se hace el procesamiento (naranja) y cómo esta información se transmite de vuelta al usuario. Cabe resaltar que cuando se crea el servidor local, se cargan todas las funciones de Python para poder hacer el procesamiento y predicción del texto, incluyendo el modelo de inteligencia artificial y las credenciales de la cuenta de servicio de Google. El proceso consta de 5 pasos principales:

---

<sup>8</sup> <https://fastapi.tiangolo.com/>

<sup>9</sup> <https://pytorch.org/>

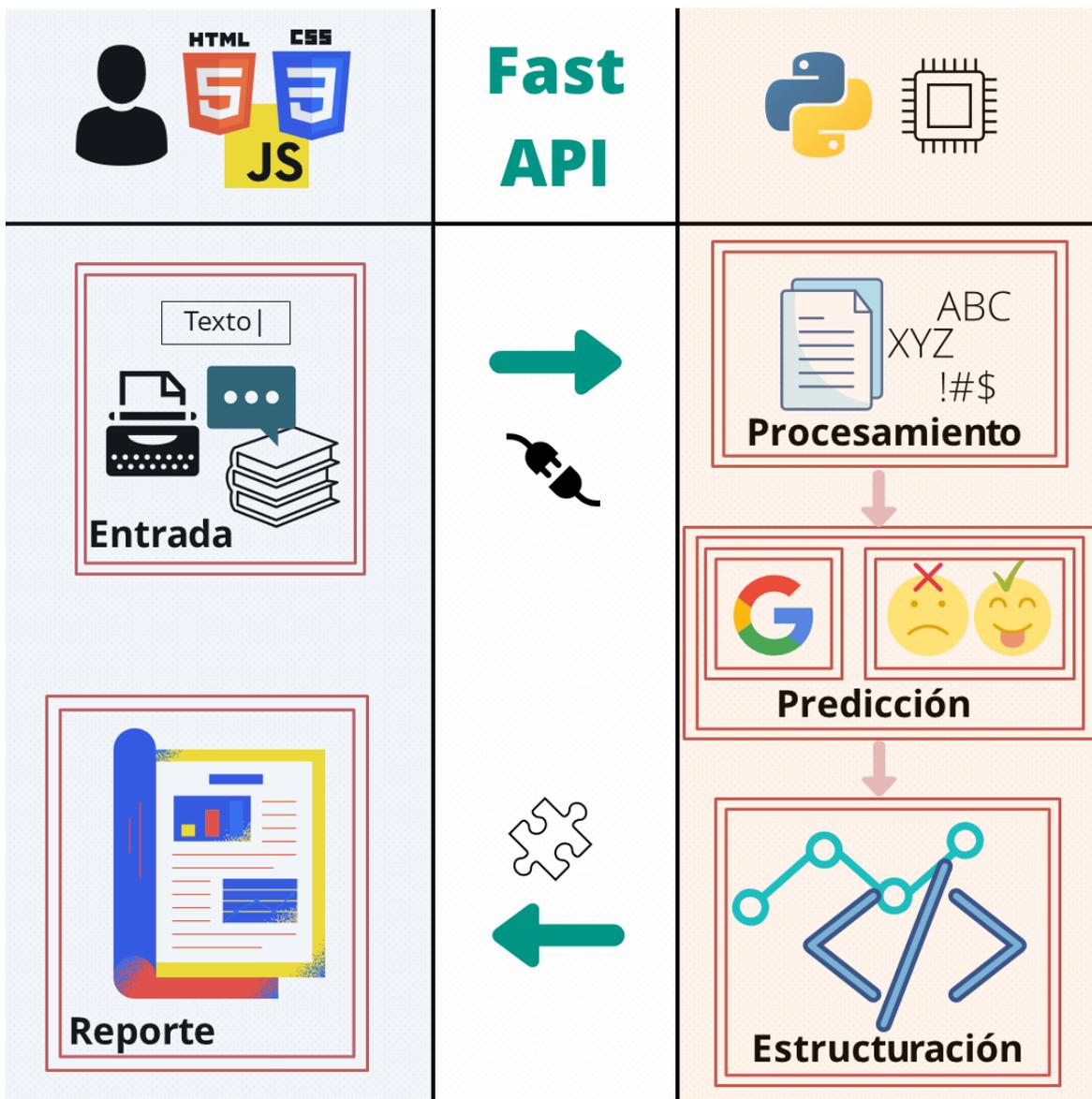
<sup>10</sup> <https://huggingface.co/transformers/>

<sup>11</sup> <https://www.nltk.org/book/ch02.html>

<sup>12</sup> <https://googleapis.dev/python/translation/latest/client.html>

<sup>13</sup> <https://googleapis.dev/python/language/1.2.0/gapic/v1/api.html>

<sup>14</sup> <https://altair-viz.github.io/>



**Figura 15.** Esquema de funcionamiento interfaz gráfica de usuario web.

1. **Entrada:** En esta parte del proceso el usuario ingresa el texto al cual quiere que se le haga el análisis y clasificación de emoción en un recuadro. Luego de esto el usuario hace clic en el botón “predecir” y esta información del texto es enviada mediante una petición de tipo POST para poder hacer el procesamiento.
2. **Procesamiento:** Una vez recibido el texto se hace uso de la API de traducción de Google para pasar el texto desde cualquier idioma que pueda ser detectado por la API a inglés, luego se realiza el procesamiento mencionado en la sección 4.3, de este proceso sale el texto tokenizado para poder hacer la predicción de la emoción en la siguiente etapa.

3. **Predicción:** En esta etapa se utiliza el texto tokenizado de la etapa anterior como insumo del modelo para predecir la emoción general del texto en las categorías: ira, alegría, tristeza y calma. Además, se hace un proceso paralelo en el cual se usa la API de procesamiento de lenguaje natural de Google para hacer análisis de sentimiento por oraciones que están separadas por puntos y también análisis de sentimiento en entidades. Para este análisis se puede utilizar el texto original de la entrada siempre y cuando el lenguaje del texto<sup>15</sup> esté soportado por la API. De lo contrario se traduce también a inglés. Como resultado de esta etapa se obtienen respuestas en formato JSON de la API de procesamiento de lenguaje natural de Google por análisis de sentimiento y análisis de sentimiento en entidades, así como la emoción predicha en el texto por el modelo de inteligencia artificial (BERT).
4. **Estructuración:** Para poder mostrar de forma interactiva el análisis de sentimiento y entidades al usuario, los datos que retornan las respuestas tipo JSON son transformados a tablas organizadas y estructuradas para poder hacer gráficas interactivas haciendo uso de la librería Altair. Es importante mencionar que la API calcula valores de magnitud y puntaje<sup>16</sup> pero para la creación de la gráfica se toma un único valor a partir de normalizar el valor de magnitud de cada frase dividiendo por la magnitud más grande dentro del texto y luego se multiplica por el valor de puntaje de cada oración individual para tener una escala normalizada con colores siendo rojo un sentimiento negativo, gris un sentimiento neutro y verde un sentimiento positivo. Por otra parte, cada entidad encontrada en el análisis de sentimiento en entidades tiene una categoría de entidad asociada<sup>17</sup>, un valor de magnitud y puntaje. Para cada entidad se genera un estilo en CSS con el color asociado al sentimiento y el texto de la categoría de entidad asociada y se genera un código HTML para el texto en general con estas entidades y un JSON con la gráfica de Altair. Toda la información, el HTML con las entidades, el JSON con la gráfica de Altair y la predicción de la emoción del modelo de inteligencia artificial son enviadas como a la interfaz gráfica.
5. **Reporte:** En este paso, se recibe la información de la etapa de estructuración y se construye la vista de usuario haciendo uso de JavaScript y la librería de visualización Vega Embed<sup>18</sup> para mostrar correctamente la gráfica de análisis de sentimiento en texto. Además, se crean imágenes para cada emoción clasificada de acuerdo con el modelo de emociones de Russell y se muestran en la interfaz gráfica de usuario de acuerdo a la emoción clasificada por el modelo BERT.

Por último, en esta sección se muestra el diagrama de Gantt del proyecto en el cual se puede ver la destinación de tiempo asignada para cada tarea con el fin de cumplir los objetivos propuestos. De forma general se divide en cuatro secciones: revisión y búsqueda del estado del arte (azul), implementación de algoritmos y desarrollo técnico del esquema de procesamiento (verde), diseño, conexión e implementación con interfaz web (amarillo) y ajustes finales (morado). Es importante mencionar que durante la ejecución de estas tareas se avanza de forma paralela con la escritura del documento final y la recopilación de información. El diagrama detallado se puede observar en la figura 16.

---

<sup>15</sup> <https://cloud.google.com/natural-language/docs/languages?hl=en>

<sup>16</sup> <https://cloud.google.com/natural-language/docs/reference/rest/v1/Sentiment>

<sup>17</sup> <https://cloud.google.com/natural-language/docs/reference/rest/v1/Entity#Type>

<sup>18</sup> <https://github.com/vega/vega-embed>

Tareas	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10	Semana 11	Semana 12	Semana 13	Semana 14	Sustentación Final
	Feb 15-19	Feb 22-26	Mar 1-5	Mar 8-12	Mar 15-19	Mar 22-26	Mar 29- Abr 2	Abr 5-9	Abr 12-16	Abr 19-23	Abr 26-30	May 3-7	May 10- 14	May 17- 21	Mayo 24
Busqueda bibliográfica estado del arte	█	█	█	█											
Análisis exploratorio base de datos y búsqueda de algoritmos de estado del arte			█	█	█	█									
Propuesta de esquema de procesamiento de datos				█	█	█	█								
Desarrollo tecnico e implementación de algoritmos del estado del arte					█	█	█	█	█						
Escritura en documento final de introducción, objetivos metodología y resultados							█	█	█						
Integración interfaz interactiva en aplicación web								█	█	█	█				
Escritura en documento final de recomendaciones y trabajos futuros. Conclusiones, bibliografía y anexos.										█	█				
Correcciones documento final												█	█		
Preparación sustentación final														█	
Sustentación final															█

Figura 16. Diagrama de Gantt proyecto.

## 4. RESULTADOS

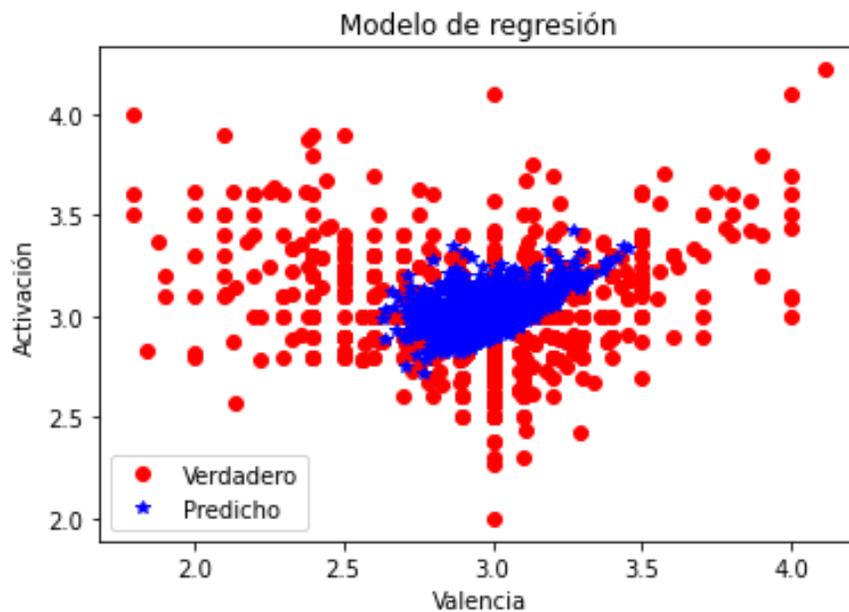
En esta sección se muestran los resultados obtenidos de la sección 3.4 (modelos de clasificación y regresión) y 3.5 (integración del modelo generado con interfaz gráfica de usuario web).

### 4.1 Modelo de regresión

Como se mencionó anteriormente se realizó una aproximación a un modelo de regresión cuyo objetivo es estimar los valores de valencia y activación haciendo uso únicamente de la base de datos EMOBANK, la cual como fue indicado en la sección 3.1 y 3.2 contiene los valores de valencia, activación y dominancia para cada frase en el conjunto de datos lo cual permite tratar esta base de datos con el modelo dimensional de emociones de Russell al únicamente escoger los valores de valencia y activación para obtener el modelo circunplejo explicado en el capítulo 1 [12].

La métrica bajo la cual se valida el desempeño del modelo es el error absoluto medio, esta métrica es utilizada debido a que describe mejor el desempeño del modelo a comparación del error cuadrático medio, teniendo en cuenta que la escala sobre la cual hace predicciones el modelo tiene valores pequeños (de 1 a 5 según la escala SAM), y la diferencia entre los valores predichos y reales puede tener una diferencia máxima de 4 unidades, por lo cual no es relevante penalizar errores grandes usando error cuadrático medio [40].

Como se puede observar en la figura 18 de forma gráfica, se obtiene el gráfico de dispersión de valores de valencia y activación, verdaderos y predichos de cada frase de prueba dentro del conjunto de datos EMOBANK. Cabe resaltar, que este modelo tiene un error absoluto medio de 0.9, es decir, que en promedio los valores de valencia y activación predichos por el modelo pueden estar 0.9 unidades por debajo o por encima del valor real [40].

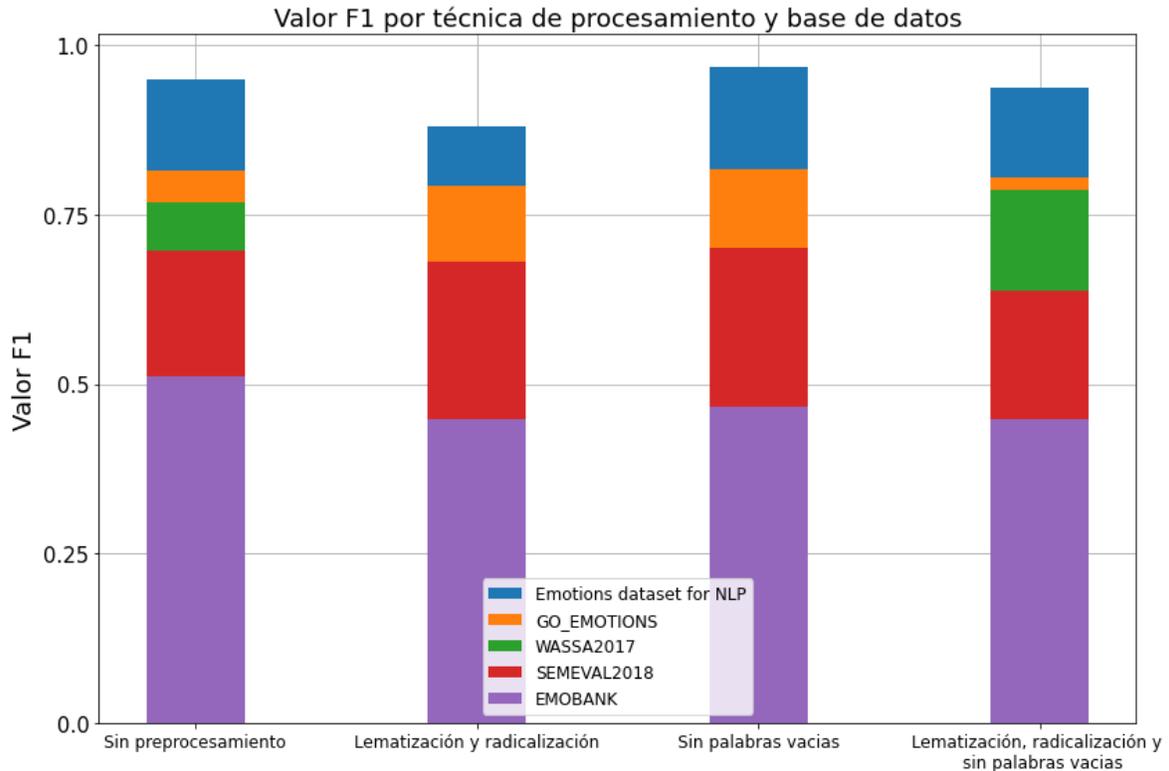


**Figura 17.** Valores verdaderos y predichos de valencia y activación juntos para el conjunto de datos de prueba EMOBANK.

Como se puede observar en la figura 17, los valores verdaderos de valencia y activación de la base de datos están distribuidos entre un rango aproximado de 1 y 5 para ambas dimensiones. Sin embargo, los valores predichos están distribuidos hacia valores cercanos a valencia y activación neutra entre 2.5 y 3.5 por lo que visualmente se puede observar que el modelo no es capaz de estimar valores de valencia y activación cercanos a los extremos de la escala SAM.

## 4.2 Modelo de clasificación

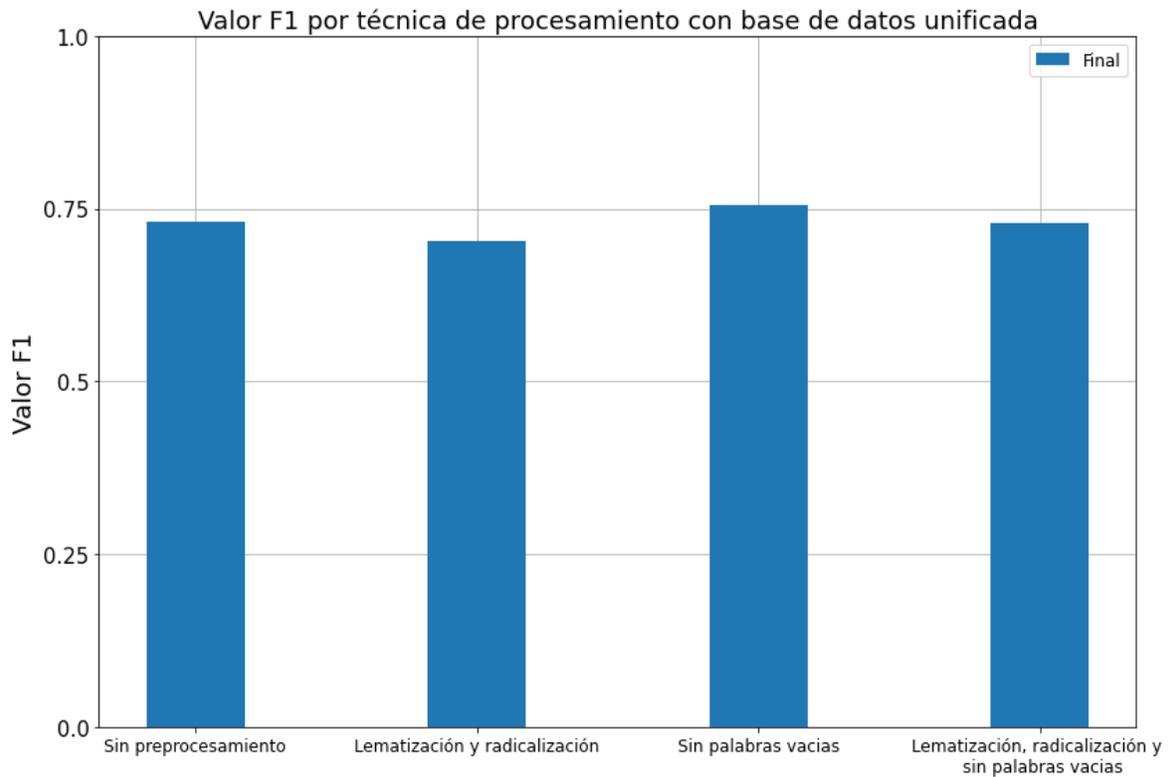
Por otra parte, los resultados de los mejores modelos en el conjunto de prueba para cada base de datos con la tarea de clasificación de emociones discretas, tomando como base el modelo de emociones de Russell, se puede observar en la figura 18.



**Figura 18.** Valor F1 ponderado para cada una de las bases de datos utilizadas con cada método de preprocesamiento.

Como se puede observar cada base de datos, tiene un comportamiento distinto debido a sus características sintácticas y semánticas. Por ejemplo, el modelo entrenado con la base de datos WASSA 2017 obtiene un mejor desempeño sin preprocesamiento o empleando las técnicas de lematización, radicalización y remoción de palabras vacías, mientras que en la base de datos SEMEVAL 2018, se obtiene el mejor desempeño con la técnica de remoción de palabras vacías.

De igual manera, los resultados del mejor modelo para cada técnica de procesamiento con la base de datos unificada se pueden observar en la figura 19 y en la tabla III se puede observar el desempeño de los modelos con las métricas F1 ponderado, macro y micro.



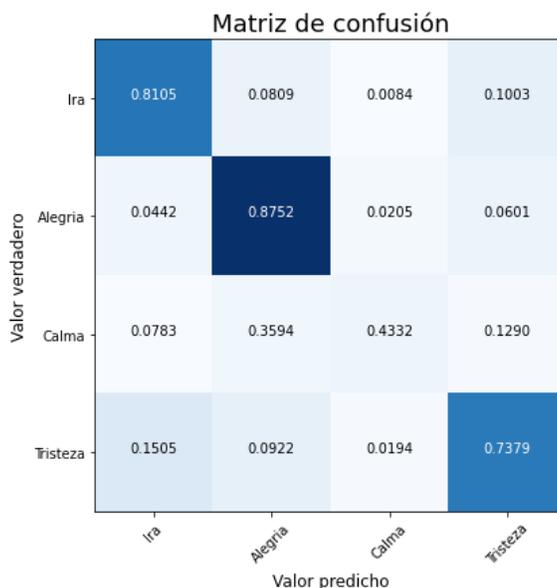
**Figura 19.** Valor F1 ponderado para la base de datos unificada con cada método de preprocesamiento.

**Tabla III**  
DESEMPEÑO MEJORES MODELOS DE CLASIFICACIÓN DE EMOCIONES

Técnica procesamiento	F1 ponderado	F1 macro	F1 micro
Sin preprocesamiento	0,7586	0,6837	0,7592
Lematización y radicalización	0,7359	0,6561	0,7382
Sin palabras vacías	<b>0,7941</b>	<b>0,7139</b>	<b>0,7946</b>
Lematización, radicalización y sin palabras vacías	0,7583	0,6847	0,7604

Al calcular diferentes métricas basadas en el puntaje F1 se tienen en cuenta la sensibilidad y exactitud del modelo de igual forma, es decir que la proporción de las emociones identificadas de forma correcta en general tiene el mismo peso que la proporción de emociones clasificadas de forma correcta por clase lo cual permite evaluar correctamente el desempeño del modelo al tener clases desbalanceadas. También, es importante mencionar que en nuestro caso la métrica F1 ponderada, es la más relevante en la

escogencia del modelo debido a que esta tiene en cuenta el promedio ponderado para cada una de las etiquetas dentro del conjunto de datos. Como podemos observar, el modelo que tiene el mejor resultado es obtenido a partir de remover las palabras vacías en el cual cada métrica individual del valor F1 tanto a nivel macro (no teniendo en cuenta la proporción de cada etiqueta en la base de datos), micro (teniendo en cuenta de forma general el conteo de verdaderos positivos, falsos negativos y falsos positivos) y ponderado tiene los valores más altos en comparación con los otros modelos obtenidos a partir de las otras técnicas de procesamiento. Por esta razón, este modelo fue utilizado como modelo final para poder clasificar las emociones en la interfaz gráfica de usuario web. La matriz de confusión para este modelo se puede observar en la figura 20.



**Figura 20.** Matriz de confusión para el conjunto de prueba del modelo entrenado con la técnica de remoción de palabras vacías.

Con base a la matriz de confusión, se puede observar que el modelo tiene su mejor desempeño en la emoción de alegría obteniendo una exactitud del 87.52% y debido al desbalanceo de clases se obtiene un rendimiento mas bajo en la emoción de calma. Sin embargo, para esta clase se sigue obteniendo un resultado 18% mejor que el azar. Además, se puede observar que el modelo puede discriminar en gran medida la emoción calma de las emociones tristeza e ira, pero le cuesta discriminar correctamente la emoción de alegría. Sin embargo, con el fin de tener un modelo basado en las emociones que se pueden ubicar el modelo de Russell, esta emoción no es descartada.

### 4.3 Interfaz gráfica de usuario

Con el modelo final de clasificación de emociones, se generó una interfaz gráfica de usuario que está compuesta de 2 vistas principales: la página principal donde el usuario puede ver el logo de la empresa Grupodot, de la Escuela Colombiana de Ingeniería y la Universidad del Rosario en la parte superior en la barra de navegación, seguido del nombre de la interfaz gráfica y una breve descripción del contenido de la interfaz y su objetivo principal. Cabe resaltar que la parte visual y la disposición de los elementos dentro de la interfaz gráfica

fueron creadas desde cero haciendo uso de HTML, CSS y Bootstrap como se mencionó en la sección 3.5. En la figura 21 se puede observar la vista principal que ve el usuario cuando abre la aplicación.



Figura 21. Vista principal de usuario de la interfaz gráfica de usuario.

La vista principal de la interfaz gráfica tiene tres divisiones las cuales son:

- **¡Clasifica tu emoción!:** En esta sección se encuentra un recuadro en el cual se puede escribir el texto el cual se quiere analizar, un botón para enviar esta información haciendo uso de fastAPI a Python para realizar el posterior procesamiento, clasificación de la emoción según el modelo de emociones de Russell y empaquetar la información para poder visualizar el reporte de análisis de emociones. También, esta división tiene una subdivisión en la cual hay un ejemplo el cual se puede copiar y pegar en el recuadro correspondiente para realizar el análisis. Este texto fue tomado de un comentario público de la red social LinkedIn y corresponde a la emoción de alegría.
- **Cómo se clasifican las emociones:** En esta sección se puede encontrar un texto en el cual se explica el modelo de emociones de Russell y su relación con la clasificación de emociones, que dimensiones componen el modelo circumplejo afectivo de Russell, las cuales son valencia y activación y por último una representación gráfica adaptada del modelo dimensional en la cual se pueden ubicar las emociones que puede clasificar el modelo de inteligencia artificial entrenado.

- **Por qué es importante analizar emociones:** En esta última sección de la vista principal se puede encontrar un texto informativo en el cual se describe la importancia de analizar y reconocer emociones, el papel que estas juegan en la toma de decisiones, resolución de conflictos y supervivencia [13][14].

La representación gráfica de cada división de la vista principal se puede observar en las figuras 22, 23 y 24.

## ¡Clasifica tu emoción!

Escribe el texto que quieres clasificar en el recuadro:

El éxito no es la clave de la felicidad. La felicidad es la clave del éxito. Si te gusta lo que estás haciendo, tendrás éxito.

Herman Cain

Clasifica mi emoción

**Ejemplo:** Esta semana conocí la ofi 🍷🍷🍷. Después de un año de haber ingresado a Mercado Libre por fin pude conocer este hermoso espacio de laburo, aunque por protocolo sin mi dream team. Feliz con el #RetornoOptativo 😊😊 #orgullomeli #adnmeli #libredeserquiensoy #elhomeofficenoeparatodos

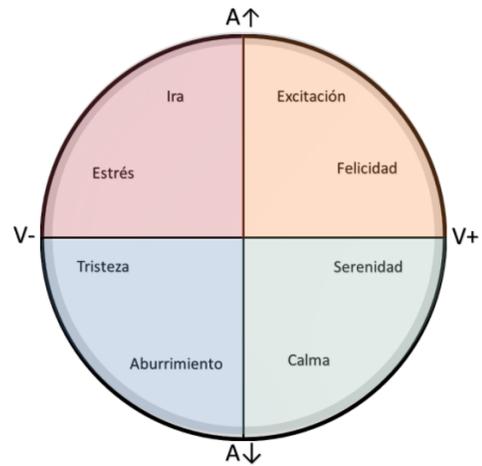
**Figura 22.** Sección “¡Clasifica tu emoción!” de la vista principal de la interfaz gráfica de usuario web.

## Cómo se clasifican las emociones

El modelo de emociones de Rusell plantea que existen dos dimensiones bajo las cuales se pueden ubicar las emociones en relación con el nivel de placer y alerta asociado con la respuesta del sistema neurofisiológico y que la combinación lineal de estos niveles permite la clasificación de emociones, estos niveles se pueden ubicar de forma gráfica en un modelo circunplejo compuesto de dos ejes principales sobre los cuales se puede ubicar qué tan agradable o desagradable es la respuesta de los sistemas neurofisiológicos ante un estímulo, lo cual se conoce como **valencia** y el incremento o la disminución de la respuesta de estos sistemas conocido como **activación**.

La valencia nos indica qué tan positivo o negativo es la sensación que estamos sintiendo y su va desde una valencia negativa hacia valencia positiva. Por otra parte, la activación está relacionada con el nivel de calma o alerta que tengo frente a un estímulo y su escala va desde baja activación hasta alta activación

Por ejemplo, las emociones relacionadas con tristeza tienen asociada una valencia negativa y una activación baja, mientras que las emociones relacionadas con alegría tienen una valencia positiva y una activación alta.



Modelo dimensional de emociones planteado por Rusell. El eje horizontal corresponde a **valencia** y el eje vertical representa la **activación**.

**Figura 23.** Sección “Cómo se clasifican las emociones” de la vista principal de la interfaz gráfica de usuario web.

## Por qué es importante analizar emociones

Para los seres humanos, entender las emociones no es una tarea fácil y a menudo es muy complejo describir, interpretar y evaluar las emociones que estamos sintiendo. Además, reconocer las emociones es un proceso de vital importancia debido a que este proceso está relacionado con la toma de decisiones lo cual continuamente nos lleva a evaluar situaciones e identificar potenciales fuentes de conflicto y tener control al iniciar la respuesta a dichas acciones con el fin de resolver el conflicto y generar una respuesta óptima frente al estímulo.

Las emociones tienen un papel muy importante tanto en la manera en la que pensamos como en la forma en la que nos mostramos delante de los demás. Tanto es así que la importancia de las emociones radica fundamentalmente a como estas nos llegan a dominar en nuestro día a día.

**Figura 24.** Sección “Por qué es importante analizar emociones” de la vista principal de la interfaz gráfica de usuario web.

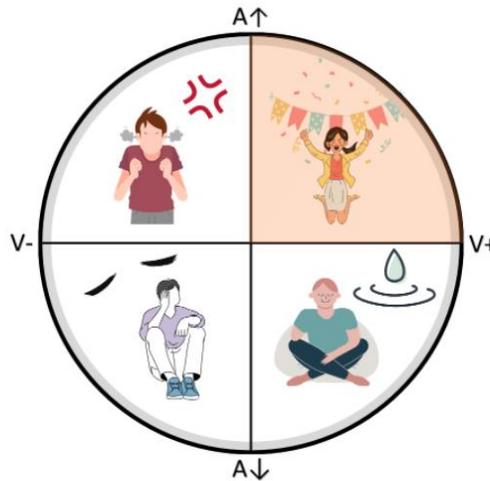
Una vez el recuadro en el cual se puede escribir el texto que se quiere analizar y se hace clic en el botón “Clasifica mi emoción”, se realiza el procesamiento, clasificación y empaquetamiento de los datos de forma interna y una vez este proceso finaliza, se renderiza la página web que tiene la información de la emoción clasificada por el modelo de inteligencia artificial, junto con una explicación breve de la ubicación de la emoción en relación con las dimensiones de valencia y activación, el análisis de sentimiento por frases en una gráfica interactiva de la librería Altair, y por último el análisis de sentimiento por entidades con un diseño personalizado. En las figuras 25, 26 y 27 se puede observar cada una de estas secciones de la vista de análisis para el texto de entrada de ejemplo:

*Esta semana conocí la ofi 🤍🤍🤍. Después de un año de haber ingresado a Mercado Libre por fin pude conocer este hermoso espacio de laburo, aunque por protocolo sin mi dream team. Feliz con el #RetornoOptativo 😊😊 #orgullomeli #adnmeli #libredeserquiensoy #elhomeofficenoesperatodos*

## ¡Analiza y entiende tu resultado!

La emoción que se detectó en el texto que escribiste es:

Alegría 😄



Es decir que la emoción predominante del texto tiene una **valencia positiva y alta activación**

**Figura 25.** Sección “¡Analiza y entiende tu resultado!” de la vista de análisis de la interfaz gráfica de usuario web.

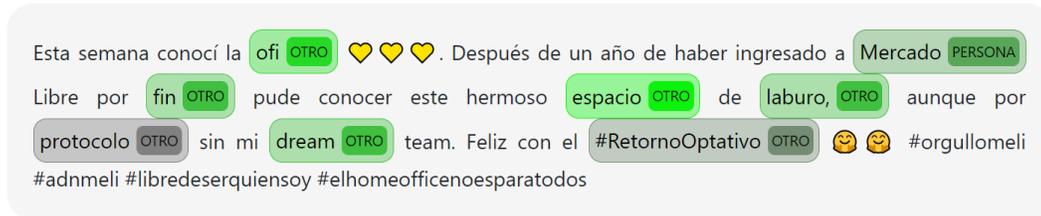
### Análisis por frase y entidades

A continuación encontrarás un gráfico de línea en el cual puedes ver el puntaje asociado con cada oración. El puntaje es un valor que relaciona directamente los valores de calificación y magnitud de las gráficas anteriores y tiene una escala entre -10 y 10. Siendo -10 lo más negativo y 10 lo más positivo. Cada punto indica una oración y su color nos brinda información acerca del puntaje, si es rojo es negativa, verde positiva y gris neutro



**Figura 26.** Sección “Análisis por frase y entidades” de la vista de análisis de la interfaz, donde se puede hacer un análisis de sentimiento (valencia) del texto por frase.

A continuación verás algunas palabras enmarcadas en rectángulos con distintos colores las cuales tienen un identificador (personas, direcciones, entre otras). Estas palabras son entidades que tienen una intensidad de color asociada dependiendo de la emoción que genera dentro del texto. Si tiene tonos de color rojo significa que es negativa, si tiene tonos de color gris significa que es neutra y si tiene tonos de color verde significa que es positiva.



**Figura 27.** Sección “Análisis por frase y entidades” de la vista de análisis de la interfaz, donde se puede observar el análisis de entidades encontradas en el texto.

## 5. DISCUSIÓN

En esta sección se realiza un análisis de los resultados obtenidos, la contribución de unificar varias bases de datos, aportes y/o diferencias del trabajo realizado y una breve comparación del estado del arte con el modelo generado.

En primer lugar, al comparar los resultados de entrenamiento y desempeño de los modelos entrenados (regresión y clasificación), se puede observar que el ajuste fino de la arquitectura BERT brinda mejores resultados para tareas de clasificación que modelos de regresión. En gran medida, el desempeño del modelo de regresión se puede ver reflejado en la poca cantidad de datos que componen a la base de datos EMOBANK en comparación con la base de datos unificada. Además, para la tarea de clasificación se busca en la salida el valor del vector que tiene una mayor magnitud para hacer la clasificación de la emoción en las categorías según el modelo de emociones de Russell mientras que en el modelo de regresión se necesita tener un valor de valencia y activación para poder hacer su posterior clasificación.

Por otra parte, se puede confirmar la teoría en la cual no existe una única técnica de procesamiento que brinde buenas métricas de desempeño para todas las bases de datos, sino que cada base de datos individual tiene una técnica con la cual se obtiene un mejor desempeño debido a la diferencia de características semánticas y sintácticas dentro de los textos [34]. Es importante mencionar que, debido a la naturaleza de la arquitectura BERT, al ser principalmente basada en un modelo de aprendizaje profundo, por lo general requiere una mayor cantidad de datos para obtener un mejor desempeño [33].

Por esta razón, uno de los puntos a favor del proyecto realizado es la recopilación y unificación de bases de datos para clasificación de emociones, al realizar este proceso no solo se obtiene un mayor número de datos con los cuales trabajar (lo cual es importante dada la naturaleza de la arquitectura BERT) sino que se obtiene una base de datos con gran diversidad de ejemplos con los cuales se abarca un espacio más amplio de variabilidad sintáctica y semántica. Es importante mencionar que, si bien se trabaja con bases de datos anotadas por seres humanos y la capacidad de hablar y razonar es inherente a nosotros,

aún es difícil describir, interpretar y evaluar las emociones. Tener una herramienta que permita orientar y dar una clasificación de emociones en texto es de gran utilidad para el apoyo no únicamente en la mejora de procesos en empresas para la evaluación de métricas y desempeños [6], sino también como herramienta de apoyo a profesionales de en el área de lingüística y psicología, para los cuales una interfaz intuitiva y fácil de utilizar es una base fundamental.

Además, en el estado del arte se pueden encontrar modelos de clasificación y detección de emociones que en su mayoría utilizan únicamente una base de datos como el caso de SEMEVAL o el uso de bases de datos creadas por el propio investigador generadas a partir del uso de APIs de redes sociales como Facebook o Twitter [33]. También, se pueden encontrar modelos generados a partir de la misma arquitectura (BERT) para una menor cantidad de clases y usando técnicas híbridas, los cuales tienen desempeños más bajos en comparación con el modelo entrenado en este proyecto [35].

En este caso, el desempeño del modelo sin hacer ajuste de hiper parámetros y haciendo uso de bases de datos con características sintácticas y semánticas diferentes obtiene un resultado aproximado del 80%. Cabe resaltar que como el modelo fue entrenado con distintas bases de datos esto hace difícil su comparación, sin embargo, según el estado del arte el modelo tiene un buen rendimiento teniendo en cuenta el estado del arte y otros modelos entrenados en menos clases con la misma arquitectura BERT. Por ejemplo, modelos basados en LSTM y redes neuronales convolucionales alcanzan valores de exactitud entre 58.61% y 59.2% [33]. También, existen modelos de máquinas de soporte vectorial basados en modelos de emociones como el de Paul Ekman que alcanzan porcentajes de exactitud de hasta 67.86% [33].

Por otra parte, existen modelos híbridos que emplean modelos adicionales como HRLCE, el cual obtiene un puntaje F1 de 0.779 sobre 3 categorías, alegría, tristeza y rabia [33]. En el estado del arte se pudo encontrar que, solo hay un modelo de clasificación de emociones que es mejor que nuestro modelo propuesto con valores ponderados F1 de 0.808 y 0.879 para las bases de datos "Friends" y "EmotionPush" respectivamente de la base de datos EmotionLines [33] [42]. Sin embargo, hay que tener en cuenta que este modelo realiza un modelado de expresiones causales y no es únicamente la arquitectura BERT con ajuste fino la encargada de hacer la clasificación.

Además, otro punto importante a favor de la interfaz gráfica de usuario y el sistema propuesto es la traducción automática al idioma inglés como se describe en la sección 3.5, para poder realizar la clasificación de emociones y análisis de sentimiento. Aunque existen palabras que no pueden ser traducidas de un idioma a otro porque son propias de un idioma, la poca cantidad de datos disponibles en idiomas distintos a inglés pone un obstáculo en la investigación y desarrollo de algoritmos de clasificación en distintos idiomas [33]. Sin embargo, esta es una forma de hacer frente a este problema y poder en gran medida hacer análisis en "otros idiomas".

En la búsqueda bibliográfica realizada y en el estado del arte se pudo observar que la gran mayoría de recursos disponibles para clasificación de emociones en texto están en

repositorios de GitHub almacenados en cuadernos de Jupyter<sup>19</sup> o Google Colaboratory<sup>20</sup> pero hay muy poca presencia de interfaces web (ya sea en servidores locales o en línea) que permitan hacer clasificación de emociones a parte de la página de modelos de HuggingFace<sup>21</sup>, por lo cual generar una interfaz gráfica en un repositorio de Github<sup>22</sup> donde se puede acceder a la interfaz es un valor agregado al proyecto además de que el modelo entrenado tiene un buen rendimiento con las métricas reportadas en el estado del arte [33].

Por último, gran parte del análisis de sentimiento se hace sobre estimación de valencia en valores negativos, neutrales y positivos como lo hace la API de procesamiento de lenguaje natural de Google. El modelo generado permite no solo identificar si la emoción es positiva o negativa debido a la clasificación dada por el eje y del modelo de Russell (valencia), sino que también permite identificar dentro de estas divisiones dos tipos emociones, alegría y calma para valencia positiva, rabia y tristeza para valencia negativa [12] [33].

## 6. RECOMENDACIONES Y TRABAJOS FUTUROS

Entre los aspectos más importantes para trabajar en el futuro cercano se encuentran: exploración de hiper parámetros de la arquitectura BERT con el fin de mejorar el desempeño del modelo, uso de la API de procesamiento de lenguaje natural de Google no solo para crear análisis de sentimiento sino también para generar un modelo híbrido que en conjunto con el modelo generado tenga un mejor desempeño y implementación de técnicas de regularización para mejorar el desempeño del modelo teniendo en cuenta el desbalanceo de clases.

También, surge la necesidad de explorar cuál es el impacto de usar traducción de cualquier lenguaje al inglés o usar bases de datos directamente en el idioma de entrenamiento. Además, con la diversidad de distintos sistemas operativos y librerías que cada usuario utiliza en sus computadoras, es importante crear una imagen contenerizada de Docker para evitar problemas con el versionamiento de los sistemas operativos y paquetes instalados para poder hacer uso de la interfaz.

## 7. CONCLUSIONES

Como resultado principal de este proyecto se cumplió el objetivo de desarrollar un prototipo de interfaz gráfica de usuario, que permite interpretar de forma cuantitativa y gráfica la emoción de un texto mediante la unión e integración de un modelo de inteligencia artificial para la clasificación de emociones con un valor F1 cercano al 80%. También, se propuso una nueva metodología que permite la utilización de diversas bases de datos para la clasificación de emociones, además de una base de datos unificada compuesta de textos de diversos dominios junto con la transformación de una base de datos dimensional a una

---

<sup>19</sup> <https://jupyter.org/>

<sup>20</sup> <https://colab.research.google.com/notebooks/intro.ipynb#recent=true>

<sup>21</sup> [https://huggingface.co/models?pipeline\\_tag=text-classification](https://huggingface.co/models?pipeline_tag=text-classification)

<sup>22</sup> <https://github.com/SebasFN9710/proyecto-dirigido-nlp>

base de datos de clasificación. Al tener como resultado final una interfaz gráfica la cual puede ser utilizada de forma intuitiva y sencilla por cualquier usuario, esta puede ser utilizada para futuros proyectos de investigación en el área de procesamiento de lenguaje natural y clasificación de emociones. Además, como se tiene el modelo base pre-entrenado de clasificación, este puede ser utilizado para realizar un modelo híbrido con imágenes y voz generando una interfaz cerebro-máquina que sirva de apoyo para ayudar a personas con trastorno de espectro autista a reconocer las emociones. También, puede ser utilizado en diversas aplicaciones e industrias como lo son la minería de datos, la psicología y la evaluación de las interacciones del servicio al cliente de una empresa para mejorar procesos internos y curvas de aprendizaje dentro de la compañía.

Es importante resaltar también, que al hacer uso de la API de procesamiento de lenguaje natural y la API de traducción de Google, esta interfaz puede ser utilizada en una gran diversidad de idiomas aumentando el campo de investigación y aplicación.

De igual manera, la generación de esta interfaz en conjunto con las herramientas que brinda Google para análisis de sentimiento por frases y en entidades, permite hacer un análisis dimensional no solo basado en el valor de valencia del texto sino que genera un valor agregado al añadir la dimensión de activación para clasificar emociones, además de estar acompañado de una representación gráfica de la ubicación de la emoción en el modelo de emociones de Russell, para facilitar su interpretación y mostrar los resultados de una forma interactiva.

Finalmente, el campo de detección de emociones en texto haciendo uso de técnicas de procesamiento de lenguaje natural es un área de investigación activa y reciente en la cual cada aporte generado permite la exploración de nuevas técnicas y metodologías, la divulgación de conocimiento y la creación de nuevas herramientas que permiten seguir innovando e investigando con el fin de resolver problemáticas de la vida social tanto nacionales como internacionales, teniendo en cuenta el gran potencial no solo a nivel social y tecnológico sino a nivel económico, que el procesamiento de lenguaje natural tiene y tendrá en los próximos años generando ganancias del orden de billones de dólares.

## 8. BIBLIOGRAFÍA

- [1] WIRE, B., 2021. Global Revenue for Enterprise Applications is Expected to Reach \$107.3 Billion by 2025, as Enterprises Move From Trials to Deployments, According to Tractica. [online] Businesswire.com. Available at: <<https://www.businesswire.com/news/home/20191014005213/en/Global-Revenue-for-Enterprise-Applications-is-Expected-to-Rreach-107.3-Billion-by-2025-as-Enterprises-Move-From-Trials-to-Deployments-According-to-Tractica>> [Accessed 12 April 2021].
- [2] Madhavan, R., 2021. Natural Language Processing - Current Applications and Future Possibilities. [online] Emerj. Available at: <<https://emerj.com/partner-content/nlp-current-applications-and-future-possibilities/>> [Accessed 12 April 2021].
- [3] Medium. 2021. Medium. [online] Available at: <<https://becominghuman.ai/alternative-nlp-method-9f94165802ed>> [Accessed 10 April 2021].
- [4] Enterpriseproject.com. 2021. Artificial intelligence (AI) vs. natural language processing (NLP): What are the differences?. [online] Available at: <<https://enterpriseproject.com/article/2020/2/artificial-intelligence-ai-vs-natural-language-processing-nlp-differences>> [Accessed 10 April 2021].
- [5] Beccue, M. and Kaul, A., 2021. Tractica Report: Natural Language Processing for the Enterprise. [online] IT Pro. Available at: <<https://www.itprotoday.com/artificial-intelligence/tractica-report-natural-language-processing-enterprise>> [Accessed 10 April 2021].
- [6] Market, N., 2021. Natural Language Processing Market Size, Share and Global Market Forecast to 2026 | MarketsandMarkets. [online] Marketsandmarkets.com. Available at: <<https://www.marketsandmarkets.com/Market-Reports/natural-language-processing-nlp-825.html>> [Accessed 12 April 2021].
- [7] B. WIRE, "Natural Language Processing Market to Reach \$22.3 Billion by 2025, According to Tractica", Businesswire.com, 2021. [Online]. Available: <https://www.businesswire.com/news/home/20170821005088/en/Natural-Language-Processing-Market-to-Rreach-22.3-Billion-by-2025-According-to-Tractica>. [Accessed: 03-May- 2021].
- [8] N. Alswaidan and M. Menai, "A survey of state-of-the-art approaches for emotion recognition in text", Knowledge and Information Systems, vol. 62, no. 8, pp. 2937-2987, 2020. Available: 10.1007/s10115-020-01449-0.
- [9]"The Passion of the Soul - Early Emotion Theories", Sagepub.com, 2021. [Online]. Available: [https://www.sagepub.com/sites/default/files/upm-binaries/63133\\_Schirmer\\_Chapter\\_1.pdf](https://www.sagepub.com/sites/default/files/upm-binaries/63133_Schirmer_Chapter_1.pdf). [Accessed: 03- May- 2021].
- [10]M. Gendron and L. Feldman Barrett, "Reconstructing the Past: A Century of Ideas About Emotion in Psychology", Emotion Review, vol. 1, no. 4, pp. 316-339, 2009. Available: 10.1177/1754073909338877.

- [11]D. Rubin and J. Talarico, "A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words", *Memory*, vol. 17, no. 8, pp. 802-808, 2009. Available: 10.1080/09658210903130764.
- [12]J. POSNER, J. RUSSELL and B. PETERSON, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", *Development and Psychopathology*, vol. 17, no. 03, 2005. Available: 10.1017/s0954579405050340.
- [13]J. Lerner, Y. Li, P. Valdesolo and K. Kassam, "Emotion and Decision Making", *Annual Review of Psychology*, vol. 66, no. 1, pp. 799-823, 2015. Available: 10.1146/annurev-psych-010213-115043 [Accessed 3 May 2021].
- [14]"Investigating the Physiology of Human Decision-Making | Dana Foundation", Dana Foundation, 2021. [Online]. Available: <https://www.dana.org/grant/investigating-the-physiology-of-human-decision-making/>. [Accessed: 03- May- 2021].
- [15] FloydHub Blog. 2021. Tokenizers: How machines read. [online] Available at: <<https://blog.floydhub.com/tokenization-nlp/>> [Accessed 12 April 2021].
- [16]Y. Zhang, R. Jin and Z. Zhou, "Understanding bag-of-words model: a statistical framework", *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1-4, pp. 43-52, 2010. Available: 10.1007/s13042-010-0001-0 [Accessed 10 May 2021].
- [17]J. Ramos, "Using tf-idf to determine word relevance in document queries," 01 2003
- [18]D. Jurafsky and J. Martin, *Speech and language processing*. Uttar Pradesh (India): Pearson, 2020.
- [19] S. Ruder, "Nlp's imagenet moment has arrived,"*The Gradient*, 2018. [online] Available at: <<https://blog.floydhub.com/tokenization-nlp/>> [Accessed 12 April 2021].
- [20]M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep Contextualized Word Representations,"*CoRR*, vol. abs/1802.05365, 2018. [Online]. Available: <http://arxiv.org/abs/1802.05365>
- [21] Alammr, J., 2018. The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning). [online] [Jalammar.github.io](http://jalammar.github.io). Available at: <<http://jalammar.github.io/illustrated-bert/>> [Accessed 12 April 2021].
- [22] A. Radford and K. Narasimhan, "Improving language understanding by generative pre-training," 2018

- [23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: <https://www.aclweb.org/anthology/N19-1423>
- [24] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," CoRR, vol. abs/1804.07461, 2018. [Online]. Available: <http://arxiv.org/abs/1804.07461>
- [25] Latysheva, N., 2019. 2019: The Year of BERT. [online] Medium. Available at: <https://towardsdatascience.com/2019-the-year-of-bert-354e8106f7ba> [Accessed 13 April 2021].
- [26] D. Kondratyuk, T. Gavenciak, M. Straka, and J. Hajic, "Lemmatag: Jointly tagging and lemmatizing for morphologically-rich languages with BRRNs," CoRR, vol. abs/1808.03703, 2018. [Online]. Available: <http://arxiv.org/abs/1808.03703>
- [27] S. Buechel and U. Hahn, "EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis", Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017. Available: 10.18653/v1/e17-2092 [Accessed 3 May 2021].
- [28] S. Mohammad, F. Bravo-Marquez, M. Salameh and S. Kiritchenko, "SemEval-2018 Task 1: Affect in Tweets", Proceedings of The 12th International Workshop on Semantic Evaluation, 2018. Available: 10.18653/v1/s18-1001 [Accessed 3 May 2021].
- [29] S. Mohammad and F. Bravo-Marquez, "WASSA-2017 Shared Task on Emotion Intensity", arXiv.org, 2021. [Online]. Available: <https://arxiv.org/abs/1708.03700>. [Accessed: 03- May- 2021].
- [30] P. Govindaraj, "Emotions dataset for NLP", Kaggle.com, 2021. [Online]. Available: <https://www.kaggle.com/praveengovi/emotions-dataset-for-nlp>. [Accessed: 03- May- 2021].
- [31] E. Saravia, H. Liu, Y. Huang, J. Wu and Y. Chen, "CARER: Contextualized Affect Representations for Emotion Recognition", Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018. Available: 10.18653/v1/d18-1404 [Accessed 3 May 2021].
- [32] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade and S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions", arXiv.org, 2020. [Online]. Available: <https://arxiv.org/abs/2005.00547>. [Accessed: 03- May- 2021].
- [33] F. Acheampong, C. Wenyu and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities", Engineering Reports, vol. 2, no. 7, 2020. Available: 10.1002/eng2.12189 [Accessed 3 May 2021].

- [34]A. Uysal and S. Gunal, "The impact of preprocessing on text classification", *Information Processing & Management*, vol. 50, no. 1, pp. 104-112, 2014. Available: 10.1016/j.ipm.2013.08.006 [Accessed 3 May 2021].
- [35]C. Huang, A. Trabelsi and O. Zaïane, "ANA at SemEval-2019 Task 3: Contextual Emotion detection in Conversations through hierarchical LSTMs and BERT", *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019. Available: 10.18653/v1/s19-2006 [Accessed 3 May 2021].
- [36]"Base Words and Inflectional Ending", Institute of Education Sciences. [Online]. Available: [https://ies.ed.gov/ncee/edlabs/regions/southeast/foundations/resources/secondgrade/rec3/3.3\\_Act\\_17\\_Base\\_Words\\_and\\_Inflectional\\_Endings.pdf](https://ies.ed.gov/ncee/edlabs/regions/southeast/foundations/resources/secondgrade/rec3/3.3_Act_17_Base_Words_and_Inflectional_Endings.pdf). [Accessed: 03- May- 2021].
- [37]"Stemming and lemmatization", Nlp.stanford.edu, 2008. [Online]. Available: <https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>. [Accessed: 03- May- 2021].
- [38]L. Vallantin, "Why is removing stop words not always a good idea", Medium, 2019. [Online]. Available: <https://medium.com/@limavallantin/why-is-removing-stop-words-not-always-a-good-idea-c8d35bd77214>. [Accessed: 03- May- 2021].
- [39]J. Ma and D. Yarats, "On the adequacy of untuned warmup for adaptive optimization", arXiv.org, 2021. [Online]. Available: <https://arxiv.org/abs/1910.04209>. [Accessed: 03- May- 2021].
- [40]M. Grandini, E. Bagli and G. Visani, "Metrics for Multi-Class Classification: an Overview", arXiv.org, 2020. [Online]. Available: <https://arxiv.org/abs/2008.05756>. [Accessed: 03- May- 2021].
- [41]J. J, "MAE and RMSE — Which Metric is Better?", Medium, 2016. [Online]. Available: <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>. [Accessed: 11- May- 2021].
- [42]Y. Huang, S. Lee, M. Ma, Y. Chen, Y. Yu and Y. Chen, "EmotionX-IDEA: Emotion BERT -- an Affectional Model for Conversation", arXiv.org, 2019. [Online]. Available: <http://arxiv.org/abs/1908.06264>. [Accessed: 12- May- 2021].
- [43]J. López et al., "Induced EEG activity during the IAPS tests and avEMT in intimate partner violence against women", *14th International Symposium on Medical Information Processing and Analysis*, 2018. Available: 10.1117/12.2511600 [Accessed 23 May 2021].