



**Análisis de emociones en entornos digitales: Una herramienta para la identificación de
estudiantes que requieren asistencia psicológica y social**

Cabezas Corral, Jhoel Esteban y Nolasco Casamen, Jefferson Alexander

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en Tecnologías de
la Información

Ing. Jerez Villota, Eleana Inés, MSc.

Sangolquí, 29 de febrero del 2024

Resultados de la herramienta para verificación y/o análisis de similitud de contenidos



Certified by
Copyleaks

About this report
help.copyleaks.com

copyleaks.com
in f @



Ing. Jerez Villota, Eleana Inés, MSc.

C. C 1717225039



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: **"Análisis de emociones en entornos digitales: Una herramienta para la identificación de estudiantes que requieren asistencia psicológica y social"** fue realizado por los señores **Cabezas Corral Jhoel Esteban** y **Nolasco Casamen Jefferson Alexander**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 29 de febrero de 2024



Firmado digitalmente por:
ELEANA INES JEREZ
VILLOTA

Ing. Jerez Villota, Eleana Inés, MSc.

C. C 1717225039



Departamento de Ciencias de la Computación
Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Nosotros, **Cabezas Corral Jhoel Esteban**, con cédula de ciudadanía n°1754440962, y **Nolasco Casamen Jefferson Alexander**, con cédula de ciudadanía n°1727504589 declaramos que el contenido, ideas y criterios del trabajo de integración curricular: **“Análisis de emociones en entornos digitales: Una herramienta para la identificación de estudiantes que requieren asistencia psicológica y social”** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 29 de febrero de 2024

Cabezas Corral Jhoel Esteban

C.C.: 1754440962

**Nolasco Casamen Jefferson
Alexander**

C.C.: 1727504589



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Nosotros **Cabezas Corral Jhoel Esteban**, con cédula de ciudadanía n°1754440962, y **Nolasco Casamen Jefferson Alexander**, con cédula de ciudadanía n°1727504589, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **"Análisis de emociones en entornos digitales: Una herramienta para la identificación de estudiantes que requieren asistencia psicológica y social"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 29 de febrero de 2024

Cabezas Corral Jhoel Esteban

C.C.: 1754440962

**Nolasco Casamen Jefferson
Alexander**

C.C.: 1727504589

Dedicatoria

A mis queridos padres, quienes han sido mi roca y mi faro a lo largo de mi viaje universitario. Su amor incondicional y apoyo constante han sido el motor que me impulsa a alcanzar mis metas. Este logro es tan suyo como mío, y dedico este trabajo a ustedes con profundo agradecimiento.

A mis profesores, ahora convertidos en amigos, quienes no solo me instruyeron académicamente, sino que también guiaron mi crecimiento personal. Su sabiduría y amistad han iluminado mi camino hacia la excelencia profesional y personal.

Jhoel Cabezas

Dedicatoria

A Dios y a mis amados padres, quienes cada día, con su amor y confianza, me enseñaron a creer siempre en mí. Sus enseñanzas son las semillas de mi vida y hoy el árbol ha dado frutos, donde el esfuerzo, perseverancia y dedicación son la clave para cumplir cualquier sueño por el que vale la pena luchar. Además, a mi hermano, quien siempre estuvo a mi lado para alegrar mi día y apoyarme en todo momento.

A mis profesores, ya que no solo formaron a buenos profesionales, sino a buenas personas. Cada uno de ustedes demostraron que la enseñanza no se trata solamente de compartir conocimiento, sino también de impartir valores. Evidenciaron que todo estudiante está destinado a la grandeza y que, para triunfar, es necesario enfrentar al fracaso sin miedo y con determinación.

Una dedicatoria especial para un ser querido que hoy me está apoyando desde el cielo. Te dedico este logro, abuelito Luis Felipe Nolasco. Me acompañaste por muchos años en este largo camino, siendo tu mano la ayuda para levantarme cuando más lo necesitaba, a dar pelea con un paso firme y nunca darme por vencido. Hoy he llegado a la meta y sé que tu espíritu siempre me acompañó en cada paso, en los momentos donde mis hombros tenían una gran carga y tus alas fueron ese apoyo que me impulsó a seguir adelante. Siempre serás mi fortaleza papito Luis.

Jefferson Nolasco

Agradecimiento

A mi compañero de tesis, Jefferson Nolasco, cuya colaboración ha sido invaluable. A través de los desafíos y triunfos, su apoyo constante ha sido un recordatorio crucial de que juntos podemos superar cualquier obstáculo. Confío en él más allá de las palabras y agradezco su amistad duradera.

A mi grupo de amigos, quienes han sido más que simples compañeros de estudio. Han sido mi red de apoyo emocional, inspirándome con su positividad y motivándome a seguir adelante. Gracias por enriquecer mi vida con nuevas experiencias y hacerme una mejor persona.

A nuestra dedicada tutora de tesis, Ing.Eleana Jerez, quien nos acompañó en cada paso de este desafiante viaje académico. Su orientación experta y apoyo inquebrantable fueron fundamentales para superar obstáculos y llevar a buen término este proyecto. Estoy agradecido por su compromiso y paciencia.

Cada uno de ustedes ha dejado una huella imborrable en mi corazón y en mi camino hacia el éxito. Este logro no solo es mío, sino también de aquellos que han iluminado mi camino. Muchas Gracias.

Jhoel Cabezas

Agradecimiento

A mi compañero de tesis, Jhoel Cabezas. Tu apoyo incondicional, sabiduría y generosidad fueron la luz en los momentos de oscuridad. Siempre serás una inspiración para mí porque demostraste que la verdadera amistad nace cuando te imaginas a ese amigo junto a ti viviendo momentos únicos y alcanzando metas que por siempre quedarán marcadas en nuestras vidas.

A mi grupo de amigos, quienes me acompañaron a lo largo de estos años en cada tarea, examen o proyecto. Su amistad, tal como un refugio, es mi abrigo en cada desafío que estuvo presente hasta culminar con el objetivo deseado.

A nuestra querida tutora de tesis, Ing. Eleana Jerez. Su pasión por el tema queda reflejada hoy con los grandes éxitos obtenidos. Gracias por no ser solo nuestra tutora, sino también una mentora. Su conocimiento es una fuente de inspiración para nosotros y esperamos, en un futuro, que nuestros caminos se crucen para trabajar juntos en un nuevo proyecto.

Gracias a todos los profesionales que, de alguna u otra manera, fueron nuestra guía para el desarrollo y culminación de este trabajo de titulación. Un agradecimiento infinito por su ayuda.

Jefferson Nolasco

Índice de Contenidos

Dedicatoria.....	6
Agradecimiento.....	8
Índice de Contenidos	10
Índice de tablas.....	14
Índice de figuras	15
Resumen	17
Abstract	18
Capítulo I	19
Introducción	19
Antecedentes	19
Planteamiento del problema.....	20
Objetivos	21
<i>Objetivo General</i>	21
<i>Objetivos Específicos</i>	21
Alcance	21
Hipótesis de Trabajo	23
Estructura del trabajo	23
Capítulo II	24
Estado del arte.....	24
Cadena de búsqueda y palabras claves.....	24

Criterios de Inclusión y Exclusión	25
<i>Criterios de inclusión</i>	25
<i>Criterios de exclusión</i>	25
Resultados de la revisión	29
Capítulo III	32
Desarrollo	32
Introducción al análisis de sentimientos	32
Herramientas y Técnicas.....	32
<i>Fuentes de Datos</i>	32
<i>Web Scraping</i>	33
<i>Octoparse</i>	33
<i>ParseHub</i>	34
<i>Apify</i>	34
<i>Helium Scraper</i>	34
<i>Bright Data</i>	35
Selección de la Herramienta.....	35
<i>Microsoft Power BI para limpieza de datos</i>	36
<i>Google Forms</i>	36
<i>Google Colab</i>	36
<i>Google Cloud</i>	36

<i>Vertex AI</i>	37
<i>Python</i>	37
<i>Flet</i>	37
<i>Selenium</i>	37
<i>Técnica de muestreo aleatorio simple</i>	38
Solución Propuesta	39
<i>Recolección de Datos</i>	40
<i>Proceso ETL</i>	41
<i>Limpieza de Publicaciones</i>	41
<i>Limpieza de Comentarios</i>	43
<i>Obtención del conjunto para el etiquetado</i>	46
<i>Proceso de Etiquetado</i>	47
<i>Análisis Exploratorio de los Datos</i>	51
<i>Desarrollo del modelo entrenado para el análisis de emociones</i>	63
<i>Desarrollo de la herramienta para análisis de emociones</i>	69
Capitulo IV	78
Evaluación y Resultados.....	78
Evaluación.....	78
<i>Evaluación del modelo de aprendizaje Vertex AI</i>	78
<i>Evaluación de la extracción de la información automatizada</i>	79

Resultados	81
<i>Análisis de casos específicos con el modelo de prueba Vertex AI</i>	<i>81</i>
<i>Resultados acerca del uso de la herramienta desarrollada</i>	<i>81</i>
<i>Resultados del análisis de las emociones.....</i>	<i>82</i>
<i>Resultados de la relación entre emociones y contenido mediante la herramienta</i>	<i>82</i>
Capítulo V Conclusiones y Recomendaciones.....	85
Conclusiones.....	85
Recomendaciones.....	86
Trabajos Futuros	86
Bibliografía.....	88
Apéndices	90

Índice de tablas

Tabla 1	<i>Alcance del trabajo de titulación</i>	21
Tabla 2	<i>Artículos Relacionados</i>	25
Tabla 3	<i>Artículos Primarios</i>	26
Tabla 4	<i>Datasets y Referencias</i>	27
Tabla 5	<i>Tipos de Contribuciones</i>	28
Tabla 6	<i>Técnicas de Machine Learning</i>	28
Tabla 7	<i>Tabla comparativa de Herramientas Web Scraping</i>	35
Tabla 8	Distribución de Datos	48
Tabla 9	Distribución de Formularios a Psicólogos.....	48
Tabla 10	Escala Visual Analógica (EVA).....	63
Tabla 11	Escala Visual Analógica (EVA).....	79
Tabla 12	Métricas analizadas para la extracción de datos	80
Tabla 13	Comentarios extraídos con la herramienta	80
Tabla 14	Resultados del análisis de emociones extraídos con la herramienta ..	83

Índice de figuras

Figura 1 <i>Gráfico de Barras de Datasets</i>	28
Figura 2 <i>Consola de Apify</i>	40
Figura 3 <i>Consulta para extraer texto mediante Power Query</i>	43
Figura 4 <i>Consulta eliminación de texto mediante Power Query</i>	44
Figura 5 <i>Dataset de Publicaciones</i>	45
Figura 6 <i>Dataset de Comentarios</i>	45
Figura 7 <i>Consulta Tabla Aleatoria de Comentarios</i>	47
Figura 8 <i>Primera sección del formulario</i>	49
Figura 9 <i>Segunda sección del formulario</i>	50
Figura 10 <i>Respuestas del formulario A</i>	50
Figura 11 <i>Respuestas del formulario A con el formato adecuado</i>	51
Figura 12 <i>Importación de librerías</i>	53
Figura 13 <i>Carga del dataset</i>	54
Figura 14 <i>Creación de la columna LongitudTexto y visualización de variables</i>	55
Figura 15 <i>Frecuencia Absoluta y Relativa</i>	55
Figura 16 <i>Gráfico de la Distribución de Emociones</i>	56
Figura 17 <i>Histograma del análisis del Texto</i>	57
Figura 18 <i>Código para crear el Diagrama de Violín</i>	58
Figura 19 <i>Diagrama de Violín</i>	59
Figura 20 <i>Código de Tokenización del Texto</i>	60
Figura 21 <i>Tokenización del Texto</i>	61
Figura 22 <i>Código para visualizar las palabras más comunes en los textos</i>	62
Figura 23 <i>Palabras más comunes en los textos</i>	62

Figura 24	Dataset para el entrenamiento del modelo.....	64
Figura 25	Crear conjunto de datos en Vertex AI	65
Figura 26	Carga exitosa del conjunto de datos	66
Figura 27	Creación del modelo en Vertex AI.....	67
Figura 28	Precisión del modelo.....	67
Figura 29	Prueba del modelo	68
Figura 30	Permisos y roles de acceso al modelo	69
Figura 31	Elemento XML para extracción de comentario	70
Figura 32	Función para la obtención de publicaciones.....	71
Figura 33	Elemento de inicio de sesión de Facebook	72
Figura 34	Autor del comentario	72
Figura 35	Elemento para mostrar el comentario completo	73
Figura 36	Script de arranque Flet.....	74
Figura 37	Funciones de control.....	74
Figura 38	Interfaz de usuario de la herramienta.....	75
Figura 39	Análisis de emociones en CrushEspeoficial	76

Resumen

La utilización de entornos digitales ha ido aumentando con el paso del tiempo, especialmente desde el surgimiento de la pandemia global causada por la COVID-19. Las personas, al estar aisladas, buscaron alternativas en entornos digitales para continuar con sus actividades diarias en el ámbito familiar, laboral, social y académico. Entornos digitales como las redes sociales ganaron popularidad en la población joven. Los jóvenes comparten todo tipo de información en sus perfiles de redes sociales, en muchas ocasiones la información que comparten puede reflejar comportamientos inusuales como depresión, soledad o ansiedad y que comúnmente pasan desapercibidos para sus familiares o los miembros de unidades de bienestar estudiantil. Si estos comportamientos no son identificados y tratados a tiempo, pueden llegar a consecuencias catastróficas para los estudiantes, como intentos de suicidio y casos graves de depresión. Creemos que es importante identificar a tiempo aquellos casos de estudiantes que requieran asistencia psicológica y social para su posterior tratamiento. Por lo tanto, en este trabajo se desarrolla una herramienta que aborda esta problemática y que toma en consideración lo siguiente: i) los tipos de contribución de otros autores con propuestas similares, ii) un modelo capaz de analizar emociones en entornos digitales, y iii) la implementación y despliegue de la herramienta, siendo el caso de estudio los estudiantes de la Universidad de las Fuerzas Armadas ESPE. La solución desarrollada obtuvo buenos resultados, teniendo la capacidad de identificar aquellos casos de estudiantes que necesitan asistencia psicológica o social a través de la identificación de emociones como la depresión, tristeza o ansiedad en sus publicaciones en entornos digitales.

Palabras claves: Análisis, emociones, sentimientos, entornos digitales, herramienta de procesamiento de lenguaje natural.

Abstract

The use of digital environments has been increasing over time, especially since the emergence of the global pandemic caused by COVID-19. People, being isolated, sought alternatives in digital environments to continue their daily activities in family, work, social and academic settings. Digital environments such as social networks gained popularity in the young population. Young people share all kinds of information on their social networking profiles, in many occasions the information they share may reflect unusual behaviors such as depression, loneliness or anxiety that commonly go unnoticed by their family members or members of student welfare units. If these behaviors are not identified and treated early, they can lead to catastrophic consequences for students, such as suicide attempts and severe cases of depression. We believe that it is important to identify in time those cases of students who require psychological and social assistance for further treatment. Therefore, in this work we develop a tool that addresses this problem and takes into consideration the following: i) the types of contribution of other authors with similar proposals, ii) a model capable of analyzing emotions in digital environments, and iii) the implementation and deployment of the tool, being the case study the students of the University of the Armed Forces ESPE. The developed solution obtained good results, having the ability to identify those cases of students who need psychological or social assistance through the identification of emotions such as depression, sadness or anxiety in their publications in digital environments.

Keywords: Analysis, emotions, feelings, digital environments, natural language processing tool.

Capítulo I

Introducción

El primer capítulo del presente trabajo de titulación aborda diferentes puntos relevantes para su desarrollo, entre los que constan los antecedentes, planteamiento del problema, justificación del trabajo, alcance, objetivo (tanto general como específicos) y, por último, la estructura que el trabajo va a contemplar en los posteriores capítulos.

Antecedentes

Desde el inicio de la pandemia global causada por la COVID-19, las herramientas tecnológicas han ganado mayor relevancia en la vida cotidiana de las personas, siendo muy utilizadas en el campo educativo, laboral y social. La sociedad, al tener que estar aislada para evitar los contagios del coronavirus, tuvo que buscar otros medios para poder interactuar y convivir hasta que la situación mundial mejorara. Es por esta razón que las redes sociales se fortalecieron notablemente, tanto en número de usuarios como en la cantidad de horas de consumo de estos entornos digitales, generando que millones de personas expresaran sus opiniones entre sí (Ochoa Quispe & Barragán Condori, 2022).

En la actualidad, estas plataformas son objeto de estudio por parte de investigadores en el campo de la educación, de la psicología y de la tecnología, debido a la valiosa información que almacenan, de la cual se pueden inferir opiniones, sentimientos, emociones y comportamientos de cualquier usuario que utilice estos entornos digitales, siendo la salud mental un tema de gran interés debido a la situación mundial de los últimos años y, al existir varios factores que ponen en riesgo el bienestar social y psicológico de una persona, es necesario enfocarse en aquellos que son causados en el ambiente de las redes sociales (Confederación Salud Mental España, 2022).

Planteamiento del problema

El avance tecnológico ha marcado un cambio significativo en la vida cotidiana, especialmente con la adopción de herramientas digitales que se han vuelto indispensables, más aún durante la pandemia. Sin embargo, este incremento en el uso de tecnología también ha revelado la complejidad de las interacciones humanas en entornos digitales. A pesar de la omnipresencia de las redes sociales, plataformas de comunicación y otros medios digitales, aún carecemos de una comprensión completa sobre cómo las personas interactúan en estos espacios y cómo las emociones influyen en su comportamiento. Esta falta de comprensión puede tener consecuencias significativas, especialmente en contextos educativos donde la identificación de estudiantes que necesitan apoyo psicológico y social se ha vuelto más desafiante. Además, la falta de comprensión sobre en qué medida influyen las emociones en el comportamiento de los estudiantes en las redes sociales añade una capa adicional de complejidad a esta problemática.

La detección de estudiantes que requieren asistencia psicológica y social es crucial para garantizar su bienestar y éxito académico. Sin embargo, en el contexto digital actual, esta tarea se ha vuelto aún más compleja. La falta de contacto físico y las barreras tecnológicas pueden dificultar la identificación temprana de señales de depresión o dificultades emocionales en los estudiantes. Además, el anonimato y la facilidad para expresar una idea en los entornos digitales pueden llevar a que problemas y comportamientos inusuales pasen desapercibidos o sean subestimados. Por lo tanto, es fundamental desarrollar estrategias y herramientas específicas para identificar y apoyar a los estudiantes que enfrentan desafíos psicológicos y sociales en este contexto digital.

Objetivos

Objetivo General

Desarrollar una herramienta para identificar casos de estudiantes que requieren asistencia psicológica y social dentro de un entorno digital.

Objetivos Específicos

- Realizar la revisión de literatura académica y científica relacionada con el análisis de emociones en entornos digitales, así como la detección de señales de estudiantes en riesgo psicológico y social.
- Seleccionar herramientas para llevar a cabo el análisis de emociones en entornos digitales.
- Recolectar datos de redes sociales para determinar el entrenamiento del modelo.
- Desarrollar una herramienta para la extracción y análisis de emociones en un entorno digital.

Alcance

El alcance del presente proyecto de titulación se encuentra dividido como se ilustra en la siguiente tabla.

Tabla 1

Alcance del trabajo de titulación

Objetivos	Preguntas de investigación	Alcance
<p>Realizar la revisión de literatura académica y científica relacionada con el análisis de emociones en entornos digitales, así como la detección de señales de estudiantes en riesgo psicológico y social.</p>	<p>a. ¿Cuál es el estado del arte del análisis de emociones en los entornos digitales? b. ¿Cuáles son los tipos de contribución que otros autores presentan?</p>	<p>a. Se investigarán trabajos desde el 2019 hasta el 2023. b. Se consultarán trabajos relacionados con el análisis de emociones en entornos digitales donde se presenten modelos, propuestas, procesos, herramientas, experimentos, métodos, evaluaciones o validaciones.</p>
<p>Seleccionar herramientas para llevar a cabo el análisis de emociones en entornos digitales.</p>	<p>a. ¿Cuáles son las técnicas, métodos o herramientas que utilizan otros autores para identificar emociones?</p>	<p>a. Para la selección de la herramienta se tomará en cuenta su confiabilidad y veracidad.</p>
<p>Recolectar datos de redes sociales para determinar el entrenamiento del modelo.</p>	<p>a. ¿Qué recursos (Datasets) utilizaron otros autores en los experimentos?</p>	<p>a. La obtención de datos debe hacerse en estudiantes que usan los entornos digitales dentro del contexto del caso de estudio, un entorno de interacción interpersonal o red social.</p>

Nota. Esta tabla muestra los objetivos y el alcance del trabajo de titulación.

Hipótesis de Trabajo

Este proyecto usa una herramienta tecnológica basada inteligencia artificial para analizar y detectar las emociones de los estudiantes en riesgo de sufrir problemas psicológicos o sociales. Con ello se puede mejorar la salud mental y el bienestar de los estudiantes.

Estructura del trabajo

A lo largo de este documento se detallarán los aspectos relevantes sobre el desarrollo y evaluación de la herramienta de software, y de la investigación correspondiente para apoyar este trabajo de integración curricular.

El segundo capítulo cuenta con una revisión de la literatura relacionada y las etapas correspondientes para llevar a cabo el estado del arte.

En el tercer capítulo se habla del proceso de desarrollo de la herramienta, con el objetivo de informar sobre la estructura de la aplicación y el proceso de su construcción y despliegue.

En el cuarto capítulo se habla sobre la evaluación del software y de los resultados obtenidos y, por último, el quinto capítulo contiene las conclusiones, recomendaciones y trabajos futuros obtenidos posteriormente al desarrollo y evaluación del proyecto.

Capítulo II

Estado del arte

El principal objetivo de este trabajo de integración curricular es desarrollar una herramienta para la identificación de estudiantes que requieren asistencia psicológica y social dentro de un entorno digital y, para lograr este propósito, hemos planteado las siguientes preguntas de investigación (research questions):

- **RQ1:** ¿Cuál es el estado del arte del análisis de emociones en los entornos digitales que pueden estar enfocados en el campo laboral, educativo o interacción interpersonal?
- **RQ2:** ¿Cuáles son los tipos de contribución (modelos, propuestas, procesos, herramientas, experimentos, métodos, evaluaciones o validaciones) que otros autores presentan?
- **RQ3:** ¿Cuáles son las técnicas, métodos o herramientas que utilizan otros autores para identificar emociones?
- **RQ4:** ¿Qué recursos (Datasets) utilizaron otros autores en los experimentos?
- **RQ5:** ¿En qué medida el análisis de emociones puede identificar a estudiantes que enfrentan desafíos psicológicos y sociales dentro de los entornos digitales?
- **RQ6:** ¿Cómo puede optimizarse el proceso de notificación dentro de los entornos digitales para garantizar una respuesta a las necesidades psicológicas y sociales de los estudiantes?

Cadena de búsqueda y palabras claves

Con las preguntas de investigación planteadas, se realizaron las consultas necesarias para la búsqueda de información dentro de las siguientes bibliotecas digitales: IEEE Xplore, Scopus y Google Scholar.

(Analysis) and (Emotions or Sentiment) and (digital environment or virtual learning or MOOCs or EVEAs) and (Natural Process Language)

Al aplicar la cadena de búsqueda en las bibliotecas digitales, obtuvimos los resultados que se muestran en la Tabla 2.

Tabla 2
Artículos Relacionados.

Repositorios (2019-2023)	Cantidad De Artículos
IEEE Xplore	13
Scopus	11
Google Scholar	1
TOTAL	25
TOTAL (SIN DUPLICADOS)	0
SELECCIONADOS	25

Nota. Esta tabla refleja la cantidad de artículos por cada biblioteca digital.

Criterios de Inclusión y Exclusión

La cantidad de artículos obtenidos anteriormente no representa la cantidad de artículos utilizados para el presente trabajo de titulación, ya que existe la posibilidad de presentación de artículos poco relevantes al tema. Razón por la que se determinaron los siguientes criterios:

Criterios de inclusión

- **I1.** Artículos en los que el Título o Abstract reporten análisis de emociones en entornos digitales que pueden estar enfocados en el campo laboral, educativo o interacción interpersonal.
- **I2.** Artículos que pertenezcan al campo de la Informática (Computer Science).

Criterios de exclusión

- **E1.** Artículos que no estén escritos en inglés.
- **E2.** Artículos que no fueron publicados en Revistas.

Luego de aplicar los criterios de inclusión y exclusión obtuvimos 11 artículos primarios (ver Tabla 3.).

Tabla 3
Artículos Primarios.

N°	AÑO	TITULO	AUTOR(ES)	DOI	REPOSITORIO
1	2019	A Content Analysis System That Supports Sentiment Analysis for Subjectivity and Polarity Detection in Online Courses	Ruth Cobos; Francisco Jurado; Alberto Blázquez-Herranz	https://doi.org/10.1109/RIT-A.2019.2952298	IEEE XPLORE
2	2023	Aspect-Based Sentiment Analysis With Heterogeneous Graph Neural Network	Wenbin An; Feng Tian; Ping Chen; Qinghua Zheng	https://doi.org/10.1109/TCSS.2022.3148866	IEEE XPLORE
3	2022	Data Analytics on Online Student Engagement Data for Academic Performance Modeling	Xiaohui Tao; Aaron Shannon-Honson; Patrick Delaney; Lin Li; Christopher Dann; Yan Li; Haoran Xie	https://doi.org/10.1109/ACCESS.2022.3208953	IEEE XPLORE
4	2022	A Hybrid Recommendation Integrating Semantic Learner Modelling and Sentiment Multi-Classification	Rawaa Alatrash; Rojalina Priyadarshini; Hadi Ezaldeen; Akram Alhinnawi	https://doi.org/10.13052/jwe1540-9589.2141	IEEE XPLORE
5	2023	Child Tracking and Prediction of Violence on Children In Social Media Using Natural Language Processing and Machine Learning	Nallakaruppan, M.K.; Srivastava, G.; Gadekallu, T.R.; Krishnan, S.; Polap, D.	10.1007/978-3-031-42505-9_47	SCOPUS
6	2022	Sentiment analysis through twitter as a mechanism for assessing university satisfaction	Chamorro-Atalaya, O.; Arce-Santillan, D.; Morales-Romero, G.; Auqui-Ramos, E., Levano-Stella, M.	10.11591/ijee.cs.v28.i1.pp430-440	SCOPUS

N°	AÑO	TITULO	AUTOR(ES)	DOI	REPOSITORIO
7	2022	Human context in Sentiment Analysis symbolic technique	Amo-Filvà, D.; Usart, M.; Grimalt-Álvaro, C., Chen, J.		SCOPUS
8	2022	A Deep Learning Approach for Public Sentiment Analysis in COVID-19 Pandemic	Sakhawat Hossain, G.M.; Asaduzzaman, S., Mynoddin, M., Sarker, I.H.	10.1109/CO NIT55038.20 22.9847839	SCOPUS
9	2022	IoT-Based Pervasive Sentiment Analysis: A Fine-Grained Text Normalization Framework for Context Aware Hybrid Applications	Habib, A.; Raza, A.A.	10.1007/978- 3-030-75123- 4_10	SCOPUS
10	2020	Sentiment Analysis of Tamil-English Code-Switched Text on Social Media Using Sub-Word Level LSTM	Raveendrarasa, V.; Amalraj, C.R.J.	10.1109/ICIT R51448.2020 .9310817	SCOPUS
11	2023	User eXperience evaluation on university virtual learning through sentiment analysis	Sanchis Rosario	Font, http://hdl.handle.net/10251/194059	GOOGLE SCHOLAR

Nota. Esta tabla muestra cada uno de los artículos primarios investigados.

Tabla 4

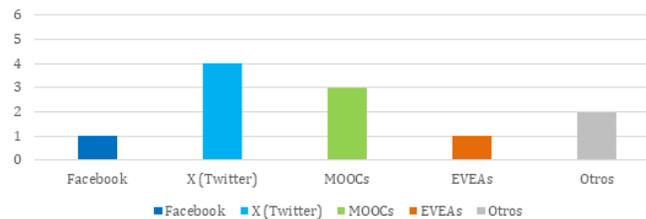
Datasets y Referencias.

Dataset	Referencias
Facebook	[10]
X (Twitter)	[5], [6], [8], [9]
MOOCs	[1], [3], [11]
Entorno Virtual de Aprendizaje (EVEAs)	[7]
Otros	[2], [4]

Nota. Esta tabla muestra cada uno de los dataset utilizados en los artículos.

Figura 1

Gráfico de Barras de Datasets.



Nota. Esta figura muestra la cantidad de datasets utilizados en cada artículo donde la mayoría se enfoca en X.

Tabla 5

Tipos de Contribuciones.

Contribución	Referencias
Herramienta	[1], [7],[10]
Propuesta	[2], [4], [5], [6], [8], [9], [10], [11]
Modelo	[2], [3], [4], [5], [11]
Método	[3]

Nota. Esta tabla muestra la asignación de artículos por tipo de contribución presentada.

Tabla 6

Técnicas de Machine Learning

Técnicas De Machine Learning	Referencias
NLP	[1], [4], [6], [7], [8], [9], [10]
Redes Neuronales	[2], [4]
Árboles de Decisión	[5]
Método Apilado (Stacking)	[3]
Minería de Opiniones	[9]
Aprendizaje Supervisado (Clasificación)	[11]

Nota. Esta tabla muestra las técnicas de Machine Learning utilizadas en los artículos para la detección de emociones.

Resultados de la revisión

Los resultados obtenidos a partir de las preguntas de investigación planteadas en este estudio proporcionan una visión integral y detallada sobre el estado del arte del análisis de emociones en entornos digitales. Se han identificado tendencias, enfoques y herramientas relevantes que contribuyen a la comprensión de la detección de emociones en el contexto laboral, educativo o de interacción personal. Por ello toda la información recolectada dará respuesta a las preguntas de investigación previamente planteadas:

- **RQ1: ¿Cuál es el estado del arte del análisis de emociones en los entornos digitales que pueden estar enfocados en el campo laboral, educativo o interacción interpersonal?**

La revisión de literatura, según la Tabla 2, revela que, si bien se han recopilado 25 artículos relacionados con el análisis de emociones en entornos digitales en el período de 2019 a 2023, esta cantidad puede considerarse moderada en comparación con la totalidad de la investigación disponible en este campo. Aunque la cantidad de artículos recopilados no es extensa, sugiere que el tema del análisis de emociones en entornos digitales ha sido abordado, en cierta medida, en la literatura académica reciente, lo que indica interés e investigación en este ámbito específico.

- **RQ2: ¿Cuáles son los tipos de contribución (modelos, propuestas, procesos, herramientas, experimentos, métodos, evaluaciones o validaciones) que otros autores presentan?**

A través de la revisión de literatura, se identifican diferentes tipos de contribuciones presentadas por otros autores en el campo del análisis de emociones en entornos digitales, estas contribuciones, según lo reflejado en la Tabla 5, demuestra que la mayoría de los artículos se

centran en propuestas como marcos de trabajo para la evaluación de la satisfacción del cliente a través del análisis de emociones en interacciones digitales, seguidas de modelos como la clasificación de emociones basado en redes neuronales convolucionales y herramientas como análisis de sentimientos con lenguaje natural, lo que indica una diversidad de enfoques y aportes en este campo de investigación.

- **RQ3: ¿Cuáles son las técnicas, métodos o herramientas que utilizan otros autores para identificar emociones?**

Según la Tabla 6, se observa que la gran mayoría de artículos se centran en técnicas de análisis de sentimientos tales como el procesamiento del lenguaje natural y modelos de aprendizaje automático o machine learning para llevar a cabo la identificación de emociones. Estas herramientas y enfoques han sido aplicados en diferentes contextos, incluyendo entornos laborales, educativos e interpersonales, lo que demuestra la diversidad de métodos utilizados en la investigación sobre el análisis de emociones en entornos digitales.

- **RQ4: ¿Qué recursos (datasets) utilizaron otros autores en los experimentos?**

Existe una gran variedad de fuentes de datos para realizar estos experimentos, incluyendo datos de redes sociales, interacciones en entornos digitales y estudios de casos específicos. Estos conjuntos de datos han sido fundamentales para la realización y validación de modelos en el campo del análisis de emociones en entornos digitales, donde se identificó que X (Twitter) es la fuente de datos que ha sido más utilizada en los artículos relacionados con el análisis de emociones en entornos digitales. Por ello, se puede inferir que esta plataforma ha sido una fuente de datos significativa para la investigación en este campo.

- **RQ5: ¿En qué medida el análisis de emociones puede identificar a estudiantes que enfrentan desafíos psicológicos y sociales dentro de los entornos digitales?**

El análisis de emociones en entornos digitales ha demostrado ser efectiva para identificar a estudiantes que enfrentan desafíos psicológicos y sociales. Algunos de los artículos revisados en la literatura académica han destacado la capacidad del análisis de emociones para detectar signos de estrés, ansiedad, depresión y otros problemas o trastornos emocionales en los estudiantes, a través de las interacciones en entornos digitales tales como las redes sociales.

Estos estudios han resaltado la importancia de utilizar técnicas de procesamiento de lenguaje natural, análisis de sentimientos y modelos de aprendizaje automático para identificar patrones emocionales que puedan indicar la necesidad de apoyo psicológico y social en los estudiantes.

- **RQ6: ¿Cómo puede optimizarse el proceso de notificación dentro de los entornos digitales para garantizar una respuesta a las necesidades psicológicas y sociales de los estudiantes?**

La optimización del proceso de notificación dentro de los entornos digitales para garantizar una respuesta efectiva a las necesidades psicológicas y sociales de los estudiantes puede lograrse mediante la integración de herramientas de monitoreo de emociones. Estas herramientas, basadas en análisis de sentimientos y procesamiento de lenguaje natural, permiten supervisar, de manera continua, las interacciones y expresiones emocionales de los estudiantes en plataformas digitales, identificando patrones de comportamiento emocional que puedan indicar la necesidad de una intervención psicológica o social.

Capítulo III

Desarrollo

La presente sección detalla con precisión las técnicas y herramientas utilizadas para la recolección de datos, así como la solución propuesta, que abarca de manera integral tanto el proceso de limpieza, el entrenamiento del modelo y el desarrollo de la herramienta.

Introducción al análisis de sentimientos

Dentro de la técnica para el análisis de sentimientos se puede observar que las emociones no solo están presentes dentro de una comunicación oral, sino también de una comunicación escrita por ello esta técnica puede aplicarse a una gran variedad de propósitos, incluyendo la evaluación de la psicología humana tal y como los son las emociones.

El análisis de sentimientos puede utilizarse para comprender el estado emocional de un profesional, estudiantes o de cualquier usuario que interactúe en el entorno digital. Esto puede ser útil para un directivo educativo, profesores o profesionales que se preocupen por la salud mental de la sociedad ya que les puede proporcionar información sobre cómo las personas están respondiendo a su experiencia ya sea laboral, educativa o social.

Herramientas y Técnicas

Fuentes de Datos

Dentro del estado del arte, la mayoría de los trabajos relacionados con técnicas de análisis de sentimientos se aplican en X, antes conocido como Twitter tal y como lo podemos observar en los artículos estudiados (Chamorro-Atalaya et al., 2022). X es una plataforma social que permite a los usuarios expresar sus opiniones y sentimientos de forma rápida y sencilla mediante posts.

Luego de analizar las redes sociales de las principales universidades del Ecuador, observamos que existen pocos usuarios activos en X. Por lo tanto, se debe encontrar otras fuentes de datos que puedan utilizarse para analizar sentimientos y emociones de los estudiantes.

En este trabajo de integración curricular, se propone usar Facebook como fuente de datos para analizar sentimientos y emociones de los estudiantes, ya que Facebook es una red social utilizada por los estudiantes de la Universidad de las Fuerzas Armadas ESPE, siendo la Universidad que utilizaremos como caso de estudio, donde la mayoría de los estudiantes usan publicaciones dentro de los grupos para comunicarse entre sí, compartir información y expresar sus opiniones y sentimientos. Estos grupos permiten a los estudiantes comunicarse sobre una variedad de temas, incluyendo sus experiencias académicas, sus relaciones personales y sus opiniones sobre el mundo.

Web Scraping

Es el proceso de extracción y almacenamiento de datos o contenido de un sitio web a través de un software (IONOS, 2024). Se pueden obtener datos como correos electrónicos, enlaces web, números de teléfono, comentarios de usuarios, etc.

Debido a que los estudiantes de la ESPE tienen una mayor interacción con las páginas de Facebook anteriormente mencionadas, fue necesario buscar una herramienta que logre la extracción de datos en Facebook.

Se encontraron varias alternativas que logran realizar web scraping:

Octoparse

Es un software de extracción de datos que se destaca por su facilidad de uso. Puede extraer información masiva de sitios web sin necesidad de una codificación previa, al igual que

tiene una variedad de plantillas para su uso de las cuales se destacan el scraping de X (Octoparse, 2024).

ParseHub

Es una herramienta visual de extracción de datos de la web. Se caracteriza por extraer fácilmente datos masivos de cualquier página web mediante el uso de plantillas, tal cual como lo realiza Octoparse (ParseHub, 2024).

Apify

Es una solución integral para web scraping que permite recopilar todo tipo de información de Internet a través de su extracción de datos. La herramienta tiene una prueba gratuita de 7 días con un límite de uso de cinco dólares estadounidenses (\$5). Se destaca principalmente de otras alternativas de scraping debido a que brinda una gran cantidad de opciones para realizar extracción de datos, ya sea en Facebook, X, Instagram, TikTok, incluso a chatbots como ChatGPT ya que posee una consola con diversas opciones y plantillas, incluyendo una sección donde muestra la información del consumo de la licencia gratuita. La extracción puede realizarse de acuerdo con varios parámetros personalizados y cuenta con varios formatos de descarga para el dataset obtenido (CSV, XLSX, JSON, etc.) (Apify, 2024).

Helium Scraper

Al igual que las herramientas ya mencionadas Helium Scraper ayuda extraer datos de sitios web de forma rápida y sencilla. La herramienta se basa en un enfoque visual, es decir, se pueden definir los datos que desean extraer con solo señalar y clicar en los elementos de la página web. Luego, estos datos se pueden exportar a una base de datos o a un archivo de hoja de cálculo, como CSV o Excel (Helium Scraper, 2024).

Bright Data

Es una plataforma de recolección de datos utilizada en el ámbito empresarial. Se pueden obtener grandes conjuntos de datos de cualquier sitio web y cuenta con un plan gratuito para su utilización, sin embargo, es necesaria una entrevista previa con uno de sus representantes de ventas para poder acceder al plan gratuito (Bright Data, 2024).

Selección de la Herramienta

Se consideraron 4 aspectos importantes para la elección de la herramienta de scraping. La herramienta debe tener una licencia gratuita que logre obtener los resultados deseados sin necesidad de costos adicionales. La facilidad de uso es primordial para el entendimiento y utilización de la herramienta. Debe contar con documentación para solventar dudas sobre su funcionamiento y con funcional se hace referencia al cumplimiento exitoso de su función, es decir, la obtención del dataset después de haberse realizado el scraping para Facebook.

Tabla 7

Tabla comparativa de Herramientas Web Scraping.

Herramienta	Gratuita	Facilidad de Uso	Documentación	Funcional
Octoparse	✓	✓	✓	x
ParseHub	✓	✓	x	x
Apify	✓	✓	✓	✓
Helium Scraper	✓	x	x	x
Bright Data	✓	x	✓	x

Nota. Esta tabla muestra las características importantes del por qué se tomó en cuenta la herramienta de Apify para la recolección de datos.

Microsoft Power BI para limpieza de datos

Es una herramienta empresarial que permite el análisis de datos, transformación y modelado de los mismos para facilitar la toma de decisiones en base a ellos.

Cuenta con diferentes herramientas que nos ayudan con la transformación y visualización de datos, para ello se utiliza una mesa de trabajo a la cual se denomina “dashboard” en donde se pueden incrustar varios elementos gráficos, tales como: barras, círculos, tarjetas y medidores de datos.

Dentro de la interfaz Power BI admite una amplia gama de orígenes de datos, incluidos SQL Server, Oracle, Azure SQL Database, Excel, archivos CSV y JSON, donde la interfaz gráfica hace que sea fácil crear consultas complejas gracias a la herramienta de Query Builder ya que con ella se crea datasets personalizados que se adapten a cualquier necesidad (Microsoft, 2024).

Google Forms

Es un software de administración de encuestas basado en el web creado por Google. Se destaca por su facilidad de crear formularios personalizados que se adapten a las necesidades del usuario (Google, 2024).

Google Colab

Es una herramienta de Google que brinda la posibilidad de escribir y ejecutar código en lenguaje Python desde cualquier navegador, preferiblemente Google Chrome. Es gratuita y su facilidad de uso es gracias a que se encuentra alojada en la nube (Google, 2024).

Google Cloud

Son servicios de Cloud Computing de Google basados en el almacenamiento, computación, redes e inteligencia artificial, para ayudar a las empresas y organizaciones a construir, implementar y administrar sus aplicaciones o datos (Google Cloud, 2024).

Vertex AI

Servicio de inteligencia artificial ofrecido por Google Cloud que permite a los usuarios crear, entrenar y desplegar modelos de aprendizaje automático de manera eficiente. Proporciona una plataforma completa para el desarrollo de modelos de IA, con herramientas y servicios que simplifican el proceso de construcción, entrenamiento y despliegue de modelos (Google Cloud, 2024).

Python

Este lenguaje de programación destaca por su versatilidad, ya que se puede utilizar tanto en el desarrollo de aplicaciones web como en la creación de software tradicional. Incluso es una herramienta poderosa para el aprendizaje automático. Además, es totalmente gratuito y se integra sin problemas en cualquier sistema, lo que lo convierte en una opción flexible y adaptable a diversos proyectos (Amazon, 2024).

Flet

FLET se presenta como una herramienta innovadora para el desarrollo de aplicaciones web, móviles y de escritorio en tiempo real utilizando Python. Su principal atractivo reside en la posibilidad de crear interfaces intuitivas sin necesidad de experiencia previa en front-end, lo que la convierte en una opción accesible tanto para principiantes como para programadores experimentados (Valentine Enedah, 2024).

Selenium

Es una herramienta ideal para crear suites de pruebas automatizadas sólidas basadas en navegadores web. Permite simular la interacción del usuario en diferentes entornos y escalar la ejecución de pruebas en múltiples máquinas. Funciona mediante "bibliotecas" específicos para

cada lenguaje de programación, como Python, que brindan control sobre el navegador web para simular acciones como clics, escritura de texto y navegación (Selenium, 2024).

Análisis exploratorios de los datos

El análisis exploratorio es una herramienta fundamental para el desarrollo de un modelo que permita detectar emociones en los comentarios y encabezados al analizar patrones en el lenguaje, para ayudar en la detección temprana de trastornos emocionales como la depresión, la ansiedad, entre otros.

Se busca visualizar los datos para descubrir patrones y tendencias que permitan comprender mejor el comportamiento de los usuarios y la relación entre los términos y las emociones expresadas en los comentarios, además de identificar patrones y características relevantes para la detección de emociones, tales como: las palabras más comunes asociadas con cada emoción y las tendencias de frecuencia de las emociones a lo largo del tiempo (IBM, 2024).

Técnica de muestreo aleatorio simple

Un muestreo aleatorio simple consiste en determinar el tamaño de la muestra necesario para obtener una estimación precisa de una población, donde se sigue la siguiente fórmula:

$$n = \frac{z^2 * p * q}{e^2}$$

Donde:

- **n** es el tamaño de la muestra
- **z** es el valor z correspondiente al nivel de confianza deseado

- **p** es la proporción esperada de datos que se espera que sean etiquetados correctamente
- **q** es la proporción esperada de datos que se espera que sean etiquetados incorrectamente
- **e** es el error aceptable

Con ello obtendremos una selección de muestra al azar de modo que cada elemento de la población tenga la misma probabilidad de ser seleccionado, logrando un etiquetado preciso que demuestre la emoción presente en el conjunto de datos (Kelmansky & Diana M, 2009).

Solución Propuesta

En la Tabla 7, se nota que muchas herramientas enfrentan dificultades al realizar web scraping en Facebook debido a las restricciones de protección de datos. Sin embargo, Apify destaca al ejecutar el web scraping en Facebook con éxito por ello es la mejor opción para la recolección de datos en Facebook, cumpliendo con todos los criterios de selección establecidos.

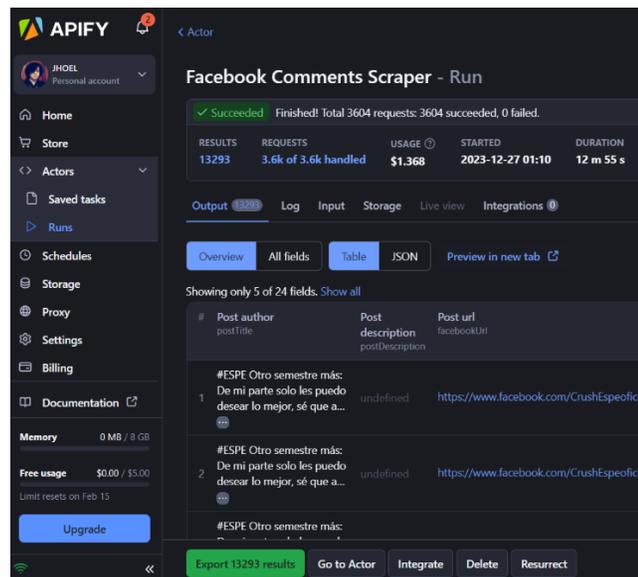
Para la recolección de datos, se utilizaron las siguientes plantillas disponibles de la tienda de Apify:

- Facebook Post Scraper
- Facebook Comments Scraper

Facebook Post Scraper es una plantilla de extracción de datos que permite extraer datos de publicaciones de Facebook. Para obtener esos datos, simplemente se necesita de la URL de la página de Facebook o el perfil de Facebook (Apify, 2024)

Facebook Comments Scraper es una plantilla de extracción de datos que puede extraer el texto del comentario en sí, la publicación a la que se hace referencia y la cantidad de reacciones a cada comentario (Apify, 2024).

Figura 2
Consola de Apify.



Nota. En esta figura se puede observar el área de trabajo que maneja Apify y una de las plantillas descritas anteriormente.

Recolección de Datos

Las páginas de Facebook a las cuales se les vas a realizar la extracción de datos son:

- Momos ESPE
- Crush Espe
- Memes ESPE

Para cada página se siguen los siguientes parámetros para la extracción:

- Desde la fecha de creación de la página hasta el 18 de diciembre del 2023.

- Se extraen todos los posts y comentarios disponibles.

Gracias a que Apify solamente necesita de un correo personal verificado, fue posible la creación de varias cuentas para solventar las limitaciones de la prueba gratuita, por lo que el uso de las plantillas tanto para obtener los comentarios y las publicaciones no generaron ningún inconveniente.

Al tener dos plantillas para la extracción de los datos se tuvieron que dividir en dos datasets donde se extrajeron un total de 5827 publicaciones y un total de 16055 comentarios para posteriormente realizar su transformación y carga.

Para la limpieza de los datos se utilizó Power BI debido a las grandes ventajas que esta herramienta ofrece en los procesos ETL.

Proceso ETL

Son las 3 fases de un proceso de gestión de datos: Extracción (Extract), Transformación (Transform), y Carga (Load), esto con el fin de garantizar la calidad de los datos y para prepararlos de manera adecuada para su análisis (AWS, 2024).

La carga de datos a la herramienta se facilitó ya que el uso del formato estándar csv proporciona una alineación rápida con cada una de las columnas obtenidas, tomando en cuenta que varias columnas no son necesarias para realizar el análisis ya que no contienen información irrelevante para el análisis (Ganti & Das Sarma, 2013).

Limpieza de Publicaciones

Para la limpieza de las publicaciones se realizó la eliminación de columnas como un primer paso para poder limpiar la data, donde de las 565 columnas que se encuentran en cada dataset de publicaciones se tomaron en cuenta los siguientes criterios:

- No contienen información relevante para el análisis. Por ejemplo, una columna que contiene el identificador de la publicación no es necesaria para analizar el número de publicaciones realizadas.
- No son consistentes en los datos. Por ejemplo, una columna que contiene links para medios como videos o imágenes de una publicación puede ser inconsistente si algunas publicaciones no tienen links de publicación.
- No aportan información adicional al análisis. Por ejemplo, una columna que contiene el número de comentarios de una publicación no aporta información adicional al análisis de la popularidad de la publicación.

Tomando en cuenta estos criterios, se eliminaron las columnas hasta obtener 10 únicas las cuales determinarán las métricas para su posterior observación.

El siguiente paso que se realizó es el reordenamiento de cada una de las columnas, al igual que su cambio de nombre, esto se realizó para tener una vista adecuada de la data, donde se obtuvieron los siguientes encabezados a utilizar.

- Tiempo (Fecha y hora en la que se realizó la publicación)
- Página (Página de Facebook de dónde se obtuvieron los datos)
- Encabezado (Texto del encabezado de cada una de las publicaciones)
- Imagen Texto (Texto transformado solo en caso de que la publicación tenga una imagen con algún texto escrito)
- Likes (Número de me gusta de cada publicación)
- Comentarios (Número de comentarios de cada publicación)
- Compartidos (Número de veces en las que se compartió la publicación)
- Video (Etiqueta para saber si existe un video dentro de la publicación)

- Fotos (Etiqueta para saber si existe una foto dentro de la publicación)
- Url (Enlace que contiene el post id de cada una de las publicaciones)

Tras ordenar las columnas, se procedió a la limpieza de la información. Un ejemplo de ello es dentro de la columna "imagenTexto", la cual contenía una etiqueta que describía la imagen en forma de texto de forma que existía un texto, que no proporcionaba información relevante para el análisis. Por ello, dentro de Power BI se utilizó la herramienta de consultas para extraer mediante delimitadores la etiqueta que no contenía texto relevante.

Figura 3

Consulta para extraer texto mediante Power Query.

```
= Table.TransformColumns("#Columnas reordenadas1",  
{{"imagenTexto", each Text.BetweenDelimiters(_, "'", "'"), type text}})
```

Nota. En la figura se observa la consulta para sacar el texto mediante delimitadores, logrando descartar la etiqueta que es información innecesaria.

Otra de las modificaciones a realizar es la eliminación de duplicados y encabezados vacíos para ello la herramienta de Power Bi nos facilita el trabajo ya que una de las funciones que tiene es que elimina automáticamente los elementos duplicados según la columna que se seleccione, para limpieza de otros datos con valores nulos se realizó la transformación de los mismos a una etiqueta booleana esto nos ayudó a determinar de forma más sencilla la verificación de columnas como lo son las de Foto y Video, así obteniendo una mejor visualización del dashboard.

Limpieza de Comentarios

Para la limpieza de los comentarios se realizó la eliminación de columnas como un primer paso para poder limpiar la data, donde, de las 23 columnas que se encuentran en cada dataset

de comentarios se eliminaron, hasta obtener 4 únicas columnas tomando en cuenta los siguientes criterios:

- No contienen información relevante para el análisis. Por ejemplo, una columna que contiene el identificador de los comentarios no es necesaria para analizar el número de comentarios realizados.
- No son consistentes en los datos. Por ejemplo, una columna que contiene el url del perfil de la persona que realizó el comentario puede ser inconsistente si algunos comentarios no tienen estos url de perfil.
- No aportan información adicional al análisis. Por ejemplo, una columna que contiene el número de likes de un comentario no aporta información adicional al análisis del texto del comentario.

Donde se obtuvieron los siguientes encabezados a utilizar.

- Url (Enlace que contiene el post id de cada una de las publicaciones)
- NombrePerfil (Nombre del usuario que realizó el comentario)
- Comentario (Texto del comentario)
- Página (Página de Facebook de dónde se obtuvieron los datos)

Con las columnas ordenadas se procede a realizar la limpieza de la información tomando en cuenta los criterios de exclusión, una de ellas es la filtración de filas las cuales contienen información innecesaria tales como lo son campos donde se observan etiquetas de perfiles al igual que filas que contienen datos vacíos o nulos.

Figura 4

Consulta eliminación de texto mediante Power Query.

```

= Table.SelectRows(
  #"Columnas quitadas",
  each ([text] <> null and [text] <> "" and Text.Length([text]) > 24)
)

```

Nota. En la figura se observa la consulta para filtrar filas mediante condiciones, logrando descartar filas con información innecesaria.

Esta consulta ayudó a eliminar y filtrar las filas que tuvieran información que no sea relevante tales como caracteres y valores nulos o vacíos. Además, se agregó una nueva columna para la identificación de la página de Facebook a la que pertenece el comentario.

Para finalizar la limpieza de la data, se realizó la unión de cada uno de los dataset que se van a utilizar obteniendo dos datasets finales: uno de las publicaciones y otro para los comentarios.

Figura 5
Dataset de Publicaciones.

	Tiempo	Página	Encabezado	imagenTexto.1
1	13/12/2023 20:17:43	Momos ESPE	Como esta people esto le pasó a un amigo cuiden sus perrenencias ya ... calidad pero es lo único que se hay de evidencia. Hola Chicos Como sabrán el martes 28 de noviembre hubo la feria de emprendimient	
2	12/12/2023 07:33:21	Momos ESPE	Yo a las 7:00 am, con toda la actitud. Segunda es todo	
3	11/12/2023 08:08:59	Momos ESPE		Buenos días a los que ya despertaron, que les vaya bien en la e
4	5/12/2023 08:04:12	Momos ESPE	#URGENTE ACTIVIDADES DEL VIERNES 08 DE DICIEMBRE	ESPE UNIVERSIDAD FUERZAS ARMADAS INNOVACION PARA EX
5	4/12/2023 22:20:31	Momos ESPE	Comenten nombres de BANDIDAS UNIVERSITARIAS	
6	4/12/2023 07:14:55	Momos ESPE	Lunes fiestas de Quito Martes y miércoles conciertos Viernes feriado Semanita corta y activa	
7	2/12/2023 14:33:07	Momos ESPE	Si son	7> Más recientes Jonathan Jaguaco Los consentidos de la ESPE

Nota. En la figura se observa el dataset final para las publicaciones.

Figura 6
Dataset de Comentarios.

URL	NombrePerfil	Comentario	Página
1 https://www.facebook.com/CrushEspecifical/posts/pfbid0WhWYING...	Miguel Espinoza	Que te tome diez años graduarte en esa universidad es absurdo no ser...	CrushEspecifical
2 https://www.facebook.com/CrushEspecifical/posts/pfbid0WhWYING...	Miguel Espinoza	Y a las pruebas me remito	CrushEspecifical
3 https://www.facebook.com/CrushEspecifical/posts/pfbid0WhWYING...	Iván Gabriel	Mi opinión: yo mejor agradecería por estar en una Universidad Pública...	CrushEspecifical
4 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Fabrizio Mendoza	Si vas a ser el moso, asegúrate de ser el mejor jaja	CrushEspecifical
5 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Tomas Asaí	El que come callado come dos veces	CrushEspecifical
6 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Santiago Andrés	Ahora estos niños se quejan de ser mosos jajajaja en mis tiempos era l...	CrushEspecifical
7 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Dennis Carrillo Fenix	Muy aparte de las bromas y todo deberían estudiar y no dedicarse a b...	CrushEspecifical
8 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Diego Illanes	Una pndjada estos de prepo hahaha así como switchito va, switchito v...	CrushEspecifical
9 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Andrea Josselyn	Las de prepo no son de Biotecnología todavía ni van a pasar ajajaja	CrushEspecifical
10 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Ingrid Riera	Mela Quiñonez que buen chisme	CrushEspecifical
11 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Nicolas Arcos	No te quejes, panas panas con el novio no eres brother si o que Steph...	CrushEspecifical
12 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Andres Oña	Jhennifer Tatiana dales consejos tu que sabes del tema	CrushEspecifical
13 https://www.facebook.com/CrushEspecifical/posts/pfbid02wkzWpoE...	Adanis Cruz	Henry Túquerrez oiga usted no es el milico?	CrushEspecifical
14 https://www.facebook.com/CrushEspecifical/posts/pfbid0PjshesFMB...	Stefy Sam	No soy de la Espe, pero quiero amigas :c jaja	CrushEspecifical
15 https://www.facebook.com/CrushEspecifical/posts/pfbid0PjshesFMB...	Blass Collaguazo	Elton Alex pensé que vo era el único que pasa solito ☹️	CrushEspecifical

Nota. En la figura se observa el dataset final para los comentarios.

Obtención del conjunto para el etiquetado

Una vez obtenido el conjunto limpio de datos finales, tanto de publicaciones como de comentarios, se procede a aplicar la técnica de muestreo simple. Esta técnica consiste en llevar a cabo una serie de fórmulas para obtener un etiquetado preciso para el entrenamiento. Para ello, se calcula el tamaño de la muestra para cada conjunto de datos por separado.

$$n = \frac{z^2 * p * q}{e^2}$$

- Para el conjunto de datos de comentarios, que tiene un total de 11450 filas, se obtuvo un tamaño de muestra de 481.92.
- Para el conjunto de datos de publicaciones, que tiene un total de 4251 filas, se obtuvo un tamaño de muestra de 177.21.

Como se observa, el tamaño de la muestra para obtener un resultado preciso es muy diferente, por lo que se aplicó una comparación de tamaños. Esto asegurará que ambas muestras representen adecuadamente a la población total.

$$r = \frac{n1}{n2}$$

Se observó que el conjunto de datos de 11450 comentarios es aproximadamente 2,72 veces más grande que el conjunto de datos de 4241 publicaciones. Por lo tanto, la propuesta para repartir las muestras es la siguiente: 250 muestras del conjunto de datos de comentarios y 100 muestras del conjunto de datos de publicaciones. Esta distribución se considera adecuada si se espera que ambas muestras contribuyan de manera significativa al conjunto de datos.

Con ello, nos aseguramos de que, al realizar las encuestas, tengamos una mayor cantidad de respuestas por parte de expertos, lo que garantizará la obtención de respuestas confiables para el conjunto de datos.

Conociendo ahora el número de muestras que debemos obtener, se procedió a realizar una consulta dentro de Power Query para la obtención de una tabla que toma filas de forma aleatoria de cada uno de los conjuntos de datos.

Figura 7

Consulta Tabla Aleatoria de Comentarios

```
Tabla Aleatoria Comments =  
VAR TotalFilas = COUNTROWS(datasetComment)  
VAR FilasAExtraer = 250  
RETURN  
SUMMARIZE(FILTER(datasetComment, RAND() <= FilasAExtraer / TotalFilas), datasetComment[Pagina],datasetComment[Comentario])
```

Nota. En la figura se observa la consulta para crear una nueva tabla, logrando obtener una muestra aleatoria del dataset de comentarios.

Proceso de Etiquetado

Para la creación de cada formulario se tuvieron en cuenta las siguientes distribuciones, considerando que cada formulario no debe contar con más de 50 ítems para que la persona encuestada no llegue a agotarse por una cantidad excesiva de ítems que debe responder. Para

recopilar información más detallada, se consideró dividir el formulario en secciones para facilitar la interacción del usuario y obtener una mejor calidad de las respuestas.

Tabla 8

Distribución de Datos

Formulario	Dato	Ítems
A	Comentario	50
B	Publicación	50
C	Comentario	50
D	Publicación	50
E	Comentario	50
F	Comentario	50
G	Comentario	50

Nota. Esta tabla muestra la distribución de datos para cada uno de los formularios.

Para la validación de los datos, es necesario del apoyo de profesionales expertos en el campo de la psicología, quienes fueron encuestados de la siguiente manera:

Tabla 9

Distribución de Formularios a Psicólogos

Psicólogo	Formulario	Ítems
1	A-B	100
2	C-D	100
3	E-F	100
4	G-A	100
5	B-C	100
6	D-E	100
7	F-G	100

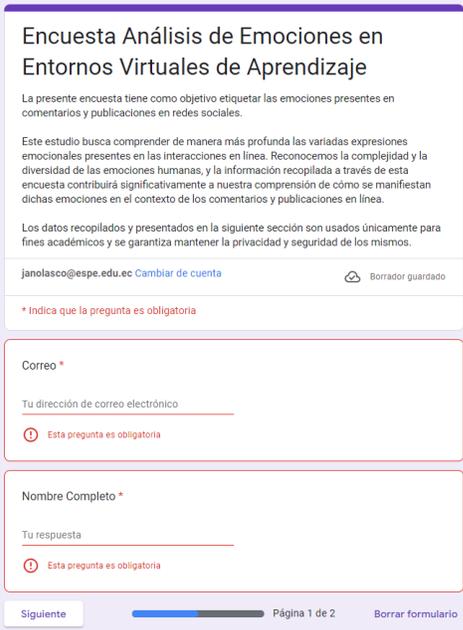
Nota. Esta tabla muestra la distribución de los formularios en orden alfabético para cada uno de los psicólogos, los cuales deben validar 100 ítems (publicaciones y comentarios de Facebook).

Los formularios fueron realizados en Google Forms, los cuales cuentan con 2 secciones:

- La primera sección donde se presenta el objetivo de la encuesta con una explicación del por qué se la realiza, además de garantizar la confidencialidad de los datos expuestos y solicitar el correo y nombre completo de la persona encuestada.
- La segunda sección donde se muestra con un ejemplo el proceso de etiquetado por parte del encuestado, presentando después los 50 ítems, cada uno con 10 emociones como opción de respuesta.

Figura 8

Primera sección del formulario



The image shows a preview of a survey form. The title is "Encuesta Análisis de Emociones en Entornos Virtuales de Aprendizaje". Below the title, there is a paragraph explaining the survey's purpose: to analyze emotions in online interactions. It mentions that the study aims to understand various emotional expressions and that the information collected will be used for academic purposes, with a guarantee of privacy and security. At the bottom of the form, there are two input fields: "Correo *" and "Nombre Completo *". Both fields have a red border and a red icon indicating they are mandatory. The "Correo *" field has a placeholder text "Tu dirección de correo electrónico" and a red error message "Esta pregunta es obligatoria". The "Nombre Completo *" field has a placeholder text "Tu respuesta" and a red error message "Esta pregunta es obligatoria". At the bottom of the form, there is a "Siguiete" button, a progress bar, and the text "Página 1 de 2" and "Borrar formulario".

Nota. En la figura se observa la primera sección del formulario en modo vista previa.

Figura 9

Segunda sección del formulario

Proceso de Etiquetado

En cada uno de los ítems, existen 10 tipos de emociones a etiquetar que van desde las primarias como la alegría y el miedo hasta aquellas más complejas como el dolor, donde cada ítem debe ser respondido con la emoción o las emociones que transmita el comentario según su criterio profesional.

Ejemplo:
Comentario:
A 18462
hombres queso les gusta esta publicación jajajaja busquen oficio!

Respuesta:
Alegría - Enojo

Recuerde que puede elegir una o varias respuestas.

Mi opinión: yo mejor agradecería por estar en una Universidad Pública porque en la Universidad Privada toca pagar!

- Alegría
- Tristeza
- Miedo
- Enojo
- Sorpresa
- Decepción
- Vergüenza
- Dolor
- Amor
- Neutral

Nota. En la figura se observa la segunda sección del formulario en modo vista previa.

Después de haber obtenido las respuestas por parte de los psicólogos, se las descargó en una hoja de cálculo para poder adaptarlas a un formato adecuado, con la ayuda de Power BI, para el análisis exploratorio de los datos que ayudará en la detección de emociones.

Figura 10

Respuestas del formulario A

	ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	ABC 123 Column5
1	Marca temporal	Dirección de correo electrónico	Nombre Completo	Mi opinión: yo mejor agradecería por estar en una Universidad Pública...	Muy aparte de las bromas y t
2	20/1/2024 8:35:58	sas-rai@hotmail.com	Saskia Saray Moya Quintero	Decepción, Dolor	Vergüenza, Dolor
3	21/1/2024 21:49:51	vjeimytatiana05@gmail.com	Jeimy Tatiana Villalba Artieda	Sorpresa	Sorpresa

- pandas (pd): Proporciona estructuras de datos de alto rendimiento y herramientas de análisis de datos, siendo útil para trabajar con conjuntos de datos tabulares.
- numpy (np): Biblioteca para realizar operaciones numéricas.
- math: Proporciona funciones matemáticas básicas.
- matplotlib.pyplot (plt): Se utiliza para visualizar datos mediante gráficos y trazados.
- %matplotlib inline: Una "magic command" de Jupyter Notebook que permite mostrar gráficos directamente en el notebook.
- seaborn (sns): Construye gráficos estadísticos sobre matplotlib y facilita la creación de visualizaciones atractivas y descriptivas.
- Punkt: Biblioteca que se utiliza para tokenizar frases en lenguaje natural.
- stopwords: Un conjunto predefinido de palabras comunes que se utilizan como palabras vacías en el análisis de texto.
- Image: Permite mostrar imágenes en un notebook.
- spacy (nlp): Biblioteca de procesamiento de lenguaje natural (NLP) que proporciona herramientas para realizar tareas como tokenización, lematización y análisis de entidades.
- re: Proporciona expresiones regulares para la manipulación de cadenas.
- string: Proporciona funciones y constantes relacionadas con manipulación de cadenas.
- nltk: La Natural Language Toolkit proporciona herramientas para el análisis de texto, tokenización, lematización, entre otras tareas.
- warnings: Permite controlar las advertencias que se generan durante la ejecución del código.
- RegexpTokenizer: Tokenizador basado en expresiones regulares.

Figura 12

Importación de librerías

```
#importación de librerías básicas
import pandas as pd
import numpy as np
import math
import matplotlib.pyplot as plt
from matplotlib import inline
import seaborn as sns
from bs4 import BeautifulSoup
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from IPython.display import Image
import spacy
nlp = spacy.load("en_core_web_sm")

import re
import string

#importación NLTK
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('omw-1.4')
nltk.download('averaged_perceptron_tagger')
nltk.download('vader_lexicon')

from nltk.tokenize import word_tokenize
from nltk.tokenize import sent_tokenize
from nltk.stem.wordnet import WordNetLemmatizer

#importación de librerías de frecuencias
from nltk.probability import FreqDist

#importación de librerías para pre-procesamiento y normalización
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.corpus import wordnet
from nltk.tag import pos_tag

#importación de librerías para Análisis de Sentimientos
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob
from nltk.tokenize import RegexpTokenizer

import warnings
warnings.filterwarnings('ignore')

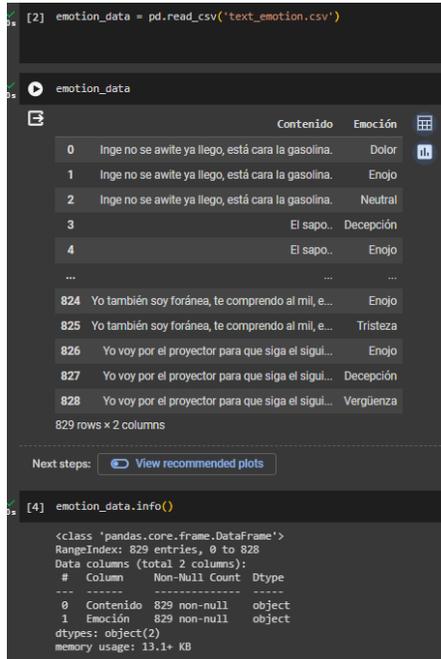
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
```

Nota. En la figura se observa que las librerías fueron importadas con éxito.

Cargamos el dataset, el cual tiene por nombre text_emotion.csv y verificamos la información que posee.

Figura 13
Carga del dataset

```
[2] emotion_data = pd.read_csv('text_emotion.csv')
```



	Contenido	Emoción
0	Inge no se awite ya llego, está cara la gasolina.	Dolor
1	Inge no se awite ya llego, está cara la gasolina.	Enojo
2	Inge no se awite ya llego, está cara la gasolina.	Neutral
3	El sapo..	Decepción
4	El sapo..	Enojo
...
824	Yo también soy foránea, te comprendo al mil, e...	Enojo
825	Yo también soy foránea, te comprendo al mil, e...	Tristeza
826	Yo voy por el proyector para que siga el sigui...	Enojo
827	Yo voy por el proyector para que siga el sigui...	Decepción
828	Yo voy por el proyector para que siga el sigui...	Vergüenza

```
[4] emotion_data.info()

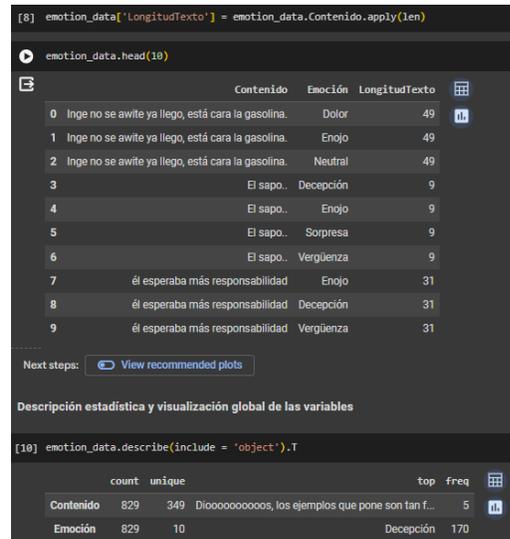
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 829 entries, 0 to 828
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Contenido  829 non-null    object
1   Emoción    829 non-null    object
dtypes: object(2)
memory usage: 13.1+ KB
```

Nota. En la figura se observa que el dataset está formado por dos columnas: Contenido y Emoción.

Para el análisis exploratorio, vamos a crear una nueva columna llamada LongitudTexto para medir la longitud del contenido y verificar si este afecta en el tipo de emoción. Además, calculamos las estadísticas descriptivas para las columnas Contenido y Emoción.

Figura 14

Creación de la columna *LongitudTexto* y visualización de variables

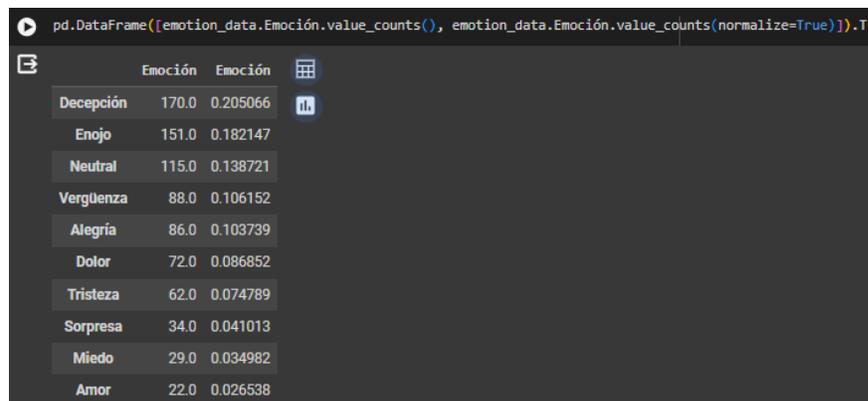


Nota. En la figura se observa que existen textos que pueden llegar a los 170 caracteres o más. También están presentes las variables numéricas que muestran los valores top y frecuencia.

Realizamos el conteo y proporción de la columna Emoción para saber su valor único y mostrar una proporción normalizada (frecuencia relativa) de esos valores.

Figura 15

Frecuencia Absoluta y Relativa

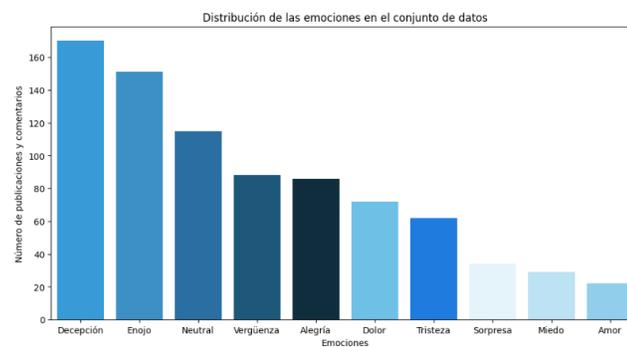


Nota. En la figura se observa que la primera columna representa la frecuencia absoluta de cada valor único donde la emoción Decepción es la que posee un mayor valor y Amor el menor valor.

Para una mejor visualización de los valores numéricos, se crea un gráfico de barras donde se muestra la distribución de las emociones en el conjunto de datos.

Figura 16

Gráfico de la Distribución de Emociones

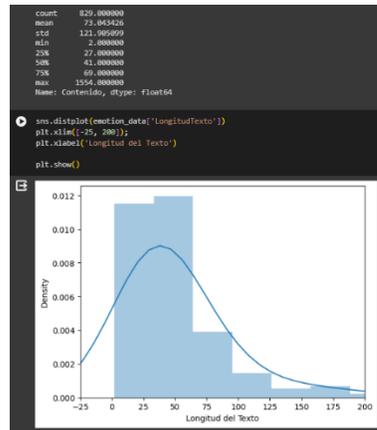


Nota. En la figura se observa la distribución de emociones de mayor a menor, donde la emoción Decepción es la de mayor distribución con respecto a las demás emociones.

Realizamos un análisis del texto en base al número de caracteres presentes en cada publicación o comentario, tal y como se lo muestra en la Figura 17.

Figura 17

Histograma del análisis del Texto



Nota. En la figura se observa un histograma de la longitud de los textos que oscilan entre 0 y 200 caracteres.

Utilizaremos el diagrama de violín para explorar la relación entre la longitud de los textos y la emoción preclasificada y determinar si la longitud del texto es un factor crítico en la identificación de la emoción expresada en el mismo.

Figura 18

Código para crear el Diagrama de Violín

```

# Calcular la media de la longitud de los textos
mean_length = emotion_data['LongitudTexto'].mean()

# Definir la paleta de colores
cmap = sns.color_palette(['#1DA1F2', '#2795D9', '#1771B5', '#0D5C8B', '#073044', '#5AC8FA', '#007AFF', '#E1F5FE', '#BBE5FC',
                        '#81D4FA', '#4FC3F7', '#29B6F6', '#03A9F4'])

# Configurar el estilo de Seaborn
sns.set_style('whitegrid')

# Cambiar el tamaño de la fuente de las etiquetas del eje Y
plt.rc('xtick', labels=14)

# Rotar las etiquetas del eje Y en 45 grados
plt.xticks(rotation=45)

# Generar el violinplot
sns.violinplot(x='Emoción', y='LongitudTexto', data=emotion_data, palette=cmap)

# Configurar el tamaño de la figura
plt.rcParams['figure.figsize'] = (14,14)
plt.rcParams.update({'font.size': 14})

# Agregar la línea horizontal en la posición de la media
plt.axhline(mean_length, color='black', linestyle='dashed', linewidth=1)

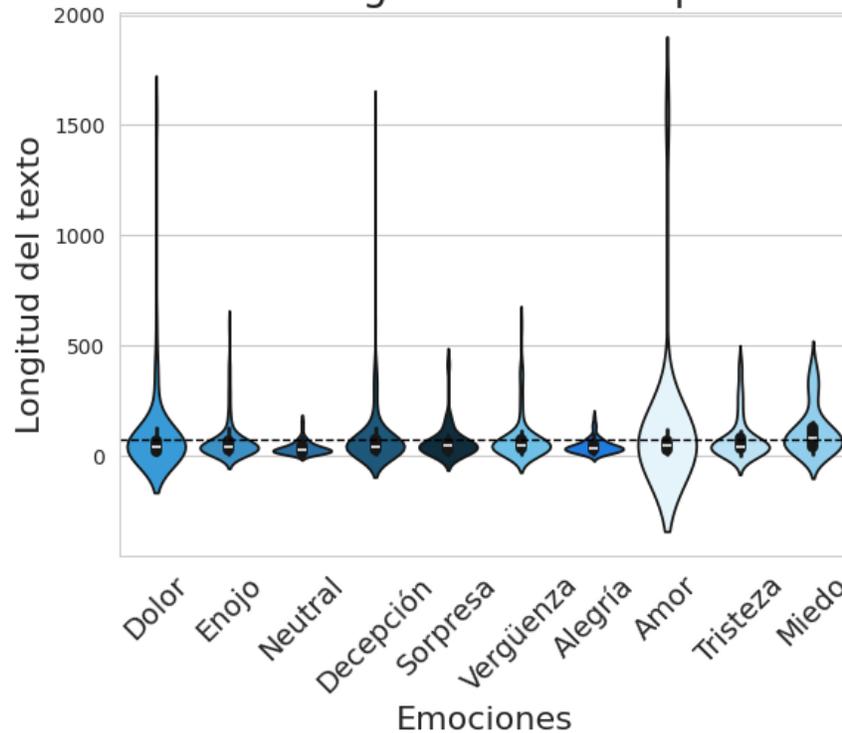
plt.title('Distribución de la longitud del texto por cada emoción', fontsize=20)
plt.xlabel('Emociones', fontsize=16)
plt.ylabel('Longitud del texto', fontsize=16)

# Mostrar el gráfico
plt.show()

```

Nota. En la figura se observa el paso a paso para crear el gráfico de violín en base a la longitud del texto y las emociones.

Para una mejor visualización de los datos con relación a la longitud del texto, utilizaremos el Diagrama de Violín, donde el eje X representa a los diferentes emociones y el eje Y muestra la longitud del texto en caracteres.

Figura 19*Diagrama de Violín***Distribución de la longitud del texto por cada emoción**

Nota. Cada violín representa la distribución de la longitud del texto para una emoción específico. La forma del violín refleja la forma de la distribución de datos. La línea central dentro de cada violín indica la mediana de la longitud del texto para la emoción correspondiente. Los puntos negros dentro de cada violín indican la media de la longitud del texto para la emoción correspondiente. El color blanco dentro de la zona central del violín representa la mediana de la longitud del texto para la emoción correspondiente.

Alegría y tristeza son las emociones con la mayor longitud de texto, con medianas de 1.400 y 1.200 caracteres, respectivamente. Decepción y Sorpresa son las emociones con la menor longitud de texto, con medianas de 800 y 900 caracteres, respectivamente. La variabilidad

en la longitud del texto es mayor para las emociones de Alegría, Tristeza y Enojo. La distribución de la longitud del texto está sesgada hacia la derecha para la mayoría de las emociones, lo que significa que hay más textos largos que cortos.

Realizar el análisis de emociones implica utilizar técnicas de procesamiento de lenguaje natural para identificar y extraer información emocional de un texto.

En el caso particular de las publicaciones y comentarios de Facebook, se puede realizar un análisis de emociones para comparar los resultados obtenidos con la emoción asignada previamente en el conjunto de datos original. Para llevar a cabo este análisis, se utilizan diversas herramientas de procesamiento de lenguaje natural, como la tokenización, lematización y análisis sintáctico, que permiten identificar las emociones expresadas en el texto.

Figura 20

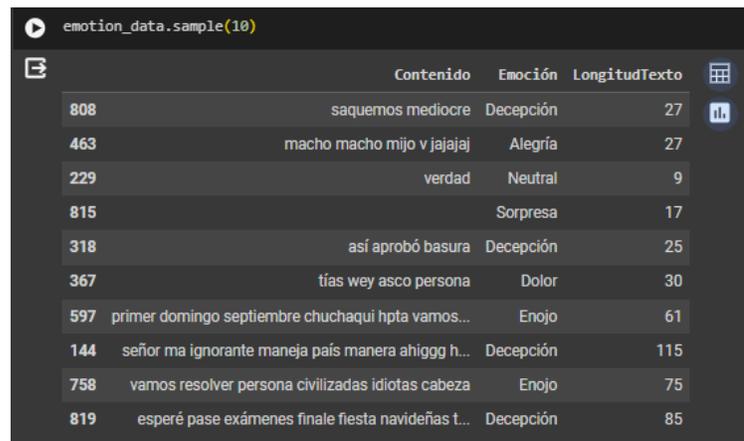
Código de Tokenización del Texto

```
def preproces_texto(texto):  
    # Convertir el texto a minúsculas  
    texto = texto.lower()  
  
    # Eliminar signos de puntuación  
    texto = re.sub(r'^\w\s', '', texto)  
  
    # Eliminar números  
    texto = re.sub(r'^\d+', '', texto)  
  
    # Eliminar palabras comunes (stopwords)  
    stop_words = stopwords.words('spanish')  
  
    words = texto.split()  
    words = [word for word in words if word not in stop_words]  
  
    # Lematización (reducir las palabras a su raíz)  
    lemmatizer = WordNetLemmatizer()  
    words = [lemmatizer.lemmatize(word) for word in words]  
    texto = ' '.join(words)  
  
    return texto  
  
[ ] # Itera a través de cada texto en la columna "Contenido"  
for i, texto in enumerate(emotion_data['Contenido']):  
    # Preprocesa el texto  
    processed_text = preproces_texto(texto)  
  
    # Reemplaza el texto original con el texto preprocesado  
    emotion_data.at[i, 'Contenido'] = processed_text
```

Nota. En la figura se observa la conversión del texto a minúsculas, eliminación de los signos de puntuación, números y palabras comunes, lematización y reemplazo del texto original con el texto preprocesado.

Figura 21

Tokenización del Texto



The screenshot shows a Jupyter Notebook interface with a table titled 'emotion_data.sample(10)'. The table has four columns: 'Contenido', 'Emoción', and 'LongitudTexto'. The 'Contenido' column contains tokenized text, where words are separated by spaces and punctuation is removed. The 'Emoción' column lists the detected emotion for each row, and the 'LongitudTexto' column shows the length of the tokenized text. The table is displayed in a dark theme.

	Contenido	Emoción	LongitudTexto
808	saquemos mediocre	Decepción	27
463	macho macho mijo v jajajaj	Alegría	27
229	verdad	Neutral	9
815		Sorpresa	17
318	así aprobó basura	Decepción	25
367	tías wey asco persona	Dolor	30
597	primer domingo septiembre chuchaqui hpta vamos...	Enojo	61
144	señor ma ignorante maneja país manera ahiggg h...	Decepción	115
758	vamos resolver persona civilizadas idiotas cabeza	Enojo	75
819	esperé pase exámenes finale fiesta navideñas t...	Decepción	85

Nota. En la figura se observa que la columna Contenido ya tiene el texto tokenizado.

Figura 22

Código para visualizar las palabras más comunes en los textos

```

from collections import Counter
import matplotlib.pyplot as plt

# Obtener las palabras y contar su frecuencia
all_words = ' '.join(content for content in emotion_data.Contenido)
word_counts = Counter(all_words.split())

# Mostrar las top words en una lista
top_words = word_counts.most_common(10)
for word, count in top_words:
    print(f'{word}: {count}')

# Mostrar las top words en un gráfico de barras
plt.figure(figsize=(15, 10))
plt.bar([word[0] for word in top_words], [word[1] for word in top_words])
plt.title('Palabras más comunes en los textos', fontdict={'size': 22})
plt.xlabel('Palabras')
plt.ylabel('Frecuencia')
plt.show()

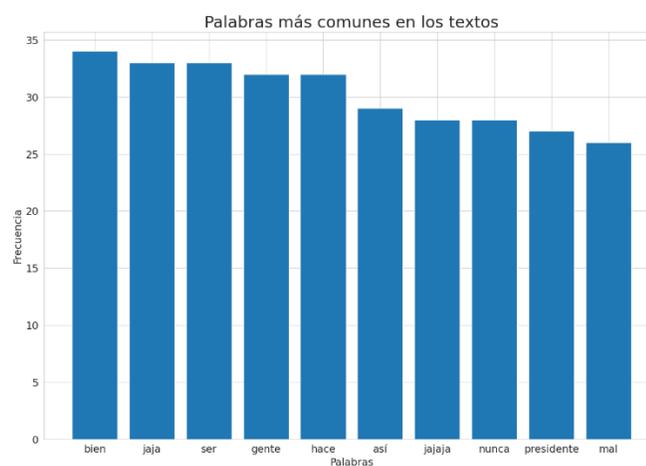
bien: 34
jaja: 33
ser: 33
gente: 32
hace: 32
así: 29
jajaja: 28
nunca: 28
presidente: 27
mal: 26

```

Nota. En la figura se observa que las palabras más comunes son: bien, jaja, ser, etc.

Figura 23

Palabras más comunes en los textos



Nota. En el gráfico de barras podemos observar las palabras más comunes en los textos.

El análisis exploratorio de datos es una etapa esencial en cualquier proyecto o investigación de análisis de datos o ciencia de datos ya que ayuda a comprender los datos, detectar problemas, formular hipótesis, seleccionar técnicas adecuadas, lo que contribuye a la toma de decisiones informadas.

Desarrollo del modelo entrenado para el análisis de emociones

En base al conjunto de datos etiquetados por psicólogos profesionales, se entrenará un modelo que logre predecir la emoción con mayor intensidad presente en el texto, determinando si existe o no depresión con la combinación de varias emociones.

Cada ítem fue calificado con una o varias emociones, sin embargo, es necesario dar un valor numérico a esas emociones para poder realizar un análisis profundo con resultados fundamentados. Con la ayuda de un profesional, se fueron calificando las emociones del 0 al 5 en base a la Escala Visual Analógica (EVA), la cual tiene las siguientes ponderaciones:

Tabla 10

Escala Visual Analógica (EVA)

Calificación	Intensidad Emocional
0	Emoción nula
1	Emoción leve
2	Emoción leve
3	Emoción intensa
4	Emoción intensa
5	Emoción intensa

Nota. Esta tabla muestra de manera general las emociones calificadas del 1 al 5 según la intensidad de la emoción.

Debido a que utilizaremos Vertex AI para el entrenamiento del modelo, es necesario clasificarlo en base a la Intensidad Emocional para una mejor predicción de las emociones.

Figura 24

Dataset para el entrenamiento del modelo

Contenido	Alegría	Tristeza	Miedo	Enojo	Sorpres	Decepción	Vergüenza	Dolor	Amor	Neutral
a eso le decimos... El Hércules	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
A quién te recuerda?	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Aaahhh jajaja	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
Adivina el personaje jajajaja	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
ahh esto que	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Ajajaja la city jajaja confirmo, quise decir usted regala sus cosas jaja disculpas	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
Alguien pásame ese sticker también pllox	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
Alguien que no haya tomado la materia de Proyecto II CENI?	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Amiga alguien en penal jajajaja	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
Aprovehchando el bug jajaja	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
así hoy que ser pa tener 10 JAJAJAJA	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
bien dicho mi frog	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
buen día con todos	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
cualquiera	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
cuando dijeron que la clase les había gustado	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Cuando piensas que el inge es tu papá	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Daniel así hubieras estado	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Diego también pero no últimamente jaja	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
dónde hemos visto esto antes???	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
El spiderverse confirmado encaman jajajaj	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
En la U igual un desmadre pero mas cerca del título pa graduarme de ingenieri	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
entendido	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
es ese wey	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso
Es hoy es hoy! Pilas invitados todos!	Alegría intensa	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral nulo
Es una en específico y todos estamos listos para esa conversación eyy	Alegría leve	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral leve
es verdad	Alegría nula	Tristeza nula	Miedo nulo	Enojo nulo	Sorpres nula	Decepción nula	Vergüenza nula	Dolor nulo	Amor nulo	Neutral intenso

Nota. En la figura se observa el dataset final basado en el campo de Intensidad Emocional de la escala de EVA.

El desarrollo del modelo está compuesto por tres fases: preparación de los datos, entrenamiento e implementación.

Figura 25

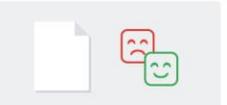
Crear conjunto de datos en Vertex AI

← Crear conjunto de datos

Nombre del conjunto de datos *
untitled_1708672894564
Se pueden usar hasta 128 caracteres.

Selecciona un tipo de datos y un objetivo
Primero, selecciona el tipo de datos que contendrá tu conjunto de datos. Luego, selecciona un objetivo, que es el resultado que deseas lograr con el modelo entrenado. [Más información](#)

IMAGEN TABULARES **TEXTO** VIDEO

 <p><input type="radio"/> Clasificación de texto (una etiqueta) Predice la etiqueta correcta que deseas que se asigne a un documento.</p>	 <p><input checked="" type="radio"/> Clasificación de texto (varias etiquetas) Predice las etiquetas correctas que deseas que se asignen a un documento.</p>	 <p><input type="radio"/> Extracción de entidades de texto Identifica entidades en tus elementos de texto.</p>	 <p><input type="radio"/> Análisis de opiniones de texto Comprende la opinión general expresada en un bloque de texto.</p>
---	--	---	--

Nota. En la figura se observa la interfaz para crear el conjunto de datos en Vertex AI y debido a que la predicción se realiza a partir de un texto, elegimos la opción Clasificación de texto (varias etiquetas) ya que un texto puede transmitir diferentes emociones.

Figura 26

Carga exitosa del conjunto de datos

The screenshot shows a web interface for data exploration. At the top, there is a breadcrumb navigation: [← datos_etiquetados](#) and [DATOS_ETIQUETADOS_TCN ▾](#). Below this are three tabs: **IMPORTAR**, **EXPLORAR** (which is active), and **ANALIZAR**. The main content is split into two columns. The left column, titled 'Etiquetas', shows a list of emotion labels with their corresponding counts. The right column, titled 'Imágenes', shows a list of text samples, each with a checkbox and a dropdown arrow. At the bottom of the right column, there is a pagination control: 'Elementos por página: 10 ▾ 1 - 10 de muchas < >'.

Etiquetas	Cantidad	Imágenes
Todos	349	Filtro Filtrar elementos
Etiquetados	348	<input type="checkbox"/> Texto
Sin etiquetas	1	<input type="checkbox"/> ▼ si wey, esa misma, hecha la influencer.
Filtro	+	<input type="checkbox"/> ▼ entonces querias ser al que de niños todos odiamos
Alegría intensa	70	<input type="checkbox"/> ▼ Ahora le dicen chupabotas a los que tienen modales
Alegría leve	20	<input type="checkbox"/> ▼ saludos
Alegría nula	258	<input type="checkbox"/> ▼ Por culpa de ustedes bandidas hay muchos en el mundo del alcohol por es
Amor intenso	11	<input type="checkbox"/> ▼ mamona es que se vuelve
Amor leve	10	<input type="checkbox"/> ▼ Aparte de castrosa putona
Amor nulo	327	<input type="checkbox"/> ▼ Claro es que muchos asambleístas no terminaron ni el colegio y tuvieron t
Decepción intensa	152	<input type="checkbox"/> ▼ Hable serio inge si mi relación a distancia no funcionó...menos sus clases
Decepción leve	17	<input type="checkbox"/> ▼ Bien!! Ingresaste a la U, parece imposible salir de aqui
Decepción nula	179	
Dolor intenso	63	
Dolor leve	10	

Nota. En la figura se observan todos los datos cargados junto con la cantidad de etiquetas que existe por nivel de emoción.

El modelo se lo crea con el método de entrenamiento AutoML (Aprendizaje Automático Automatizado) debido a su facilidad para la creación y despliegue de modelos.

Figura 27

Creación del modelo en Vertex AI

Entrena el modelo como una versión de un modelo existente

Nombre *
prueba_vertex

Descripción

División de datos

Aleatorio Manual (avanzado)

Tu conjunto de datos se aleatorizará automáticamente y se dividirá en conjuntos de entrenamiento, validación y prueba con las siguientes proporciones. [Más información](#)

Training 80 % Validation 10 % Test 10 %

● Entrenamiento: 80%
● Validación: 10%
● Prueba: 10%

Encriptación

Clave de encriptación administrada por Google
No se requiere configuración

Clave de encriptación administrada por el cliente (CMEK)
Se administra a través de Google Cloud Key Management Service

ID de modelo

Nota. En la figura se observa la interfaz para la creación del modelo, resaltando la importancia de elegir la división de datos en Entrenamiento 80%, Validación 10% y Prueba 10% ya que es lo recomendado para el tamaño del dataset.

Figura 28

Precisión del modelo

datos_etiquetados_v2 > Versión 1 [VER CONJUNTO DE DATOS](#) [EXPORTAR](#)

[EVALUAR](#) [IMPLEMENTA Y PRUEBA](#) [PREDICCIÓN POR LOTES](#) [DETALLES DE LA VERSIÓN](#)

untitled_3330727206731087872 [COMPARAR](#) [CREAR EVALUACIÓN](#)

Etiquetas [Filtro](#) **Detalles de la evaluación** [Limite de confianza](#) 0.5

Etiqueta	Puntuación
Todas las etiquetas	0.835
Alegría intensa	1
Dolor nulo	0.936
Neutral intenso	0.933
Neutral nulo	0.926
Alegría nula	0.906
Decepción nula	0.792
Neutral leve	0.782
Enojo nulo	0.743
Miedo nulo	0.729
Decepción intensa	0.673
Sorpresa nula	0.591

Todas las etiquetas

Precisión promedio	0.835
Precisión	80.1 %
Recuperación	78.3 %
Puntuación de F1	0.7919075
Fecha de creación	19 feb 2024 06:23:28
Cantidad total de elementos	348
Elementos de entrenamiento	279
Elementos de validación	34
Elementos de prueba	35

Nota. En la figura se observa que la precisión del modelo es de 80.1%, tomando en cuenta que para llegar a este valor fue necesario de un tiempo de entrenamiento de alrededor de 4 horas y 30 minutos.

Gracias a las ventajas de Vertex AI en relación al Cloud Computing, podemos validar nuestro modelo entrenado directamente desde la opción de Implementa y Prueba.

Figura 29

Prueba del modelo

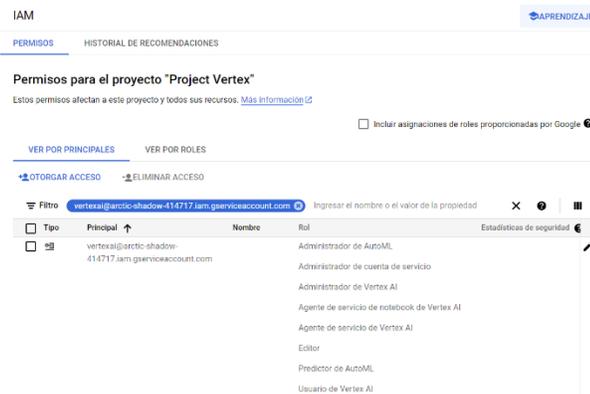


Nota. En la figura se observa que al ingresar el texto “Adivina el personaje jajajaja” nos muestra con porcentaje principal a la emoción Alegría Intensa.

Para realizar la implementación del modelo, es necesario brindar los permisos necesarios y generar una clave que nos descargará un archivo en formato JSON que sirve para autenticarnos, obtener permisos y consumir los servicios del modelo. Este archivo debe ser llamado dentro del código de la herramienta.

Figura 30

Permisos y roles de acceso al modelo



Nota. En la figura se observan todos los permisos necesarios que debe tener la cuenta para poder acceder y consumir los servicios del modelo.

Desarrollo de la herramienta para análisis de emociones

Para el desarrollo de la herramienta se seleccionó el lenguaje de programación Python. Esta elección se basó en su versatilidad, facilidad de uso y amplia comunidad de desarrolladores.

En cuanto a las bibliotecas para poder realizar la extracción de la información se tomó en cuentas las siguientes:

- Selenium: Permite realizar web scraping de forma automatizada.
- BeautifulSoup: Facilita la extracción de datos de documentos HTML.
- XML Path Language: Ofrece herramientas para la manipulación de archivos XML.

Metodología de extracción:

La herramienta implementa un proceso de extracción automatizado que se basa en los siguientes pasos:

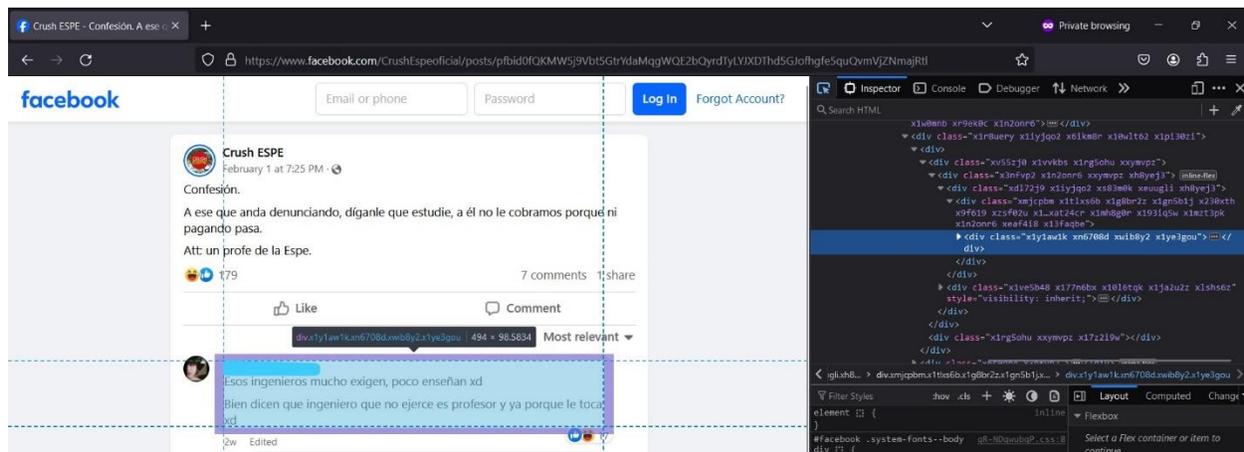
Obtención del código HTML: Se utiliza Selenium para obtener el código HTML de la página de Facebook de la que se desea extraer la información.

Localización de los elementos: Se emplea BeautifulSoup para identificar y localizar los elementos HTML que contienen la información deseada (publicaciones y comentarios).

Extracción de datos: Se utiliza XML Path Language para extraer los datos específicos de cada elemento HTML, como el texto de la publicación, la fecha de publicación, el nombre del autor y los comentarios.

Figura 31

Elemento XML para extracción de comentario



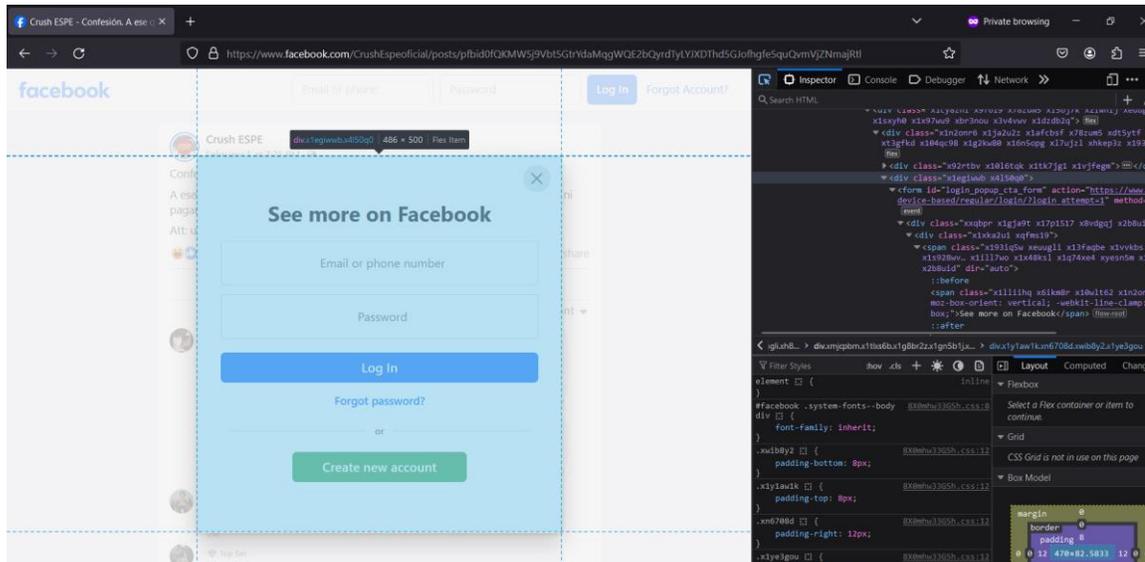
Nota. En la figura se observa el elemento XML del cual se debe extraer la información en este caso dentro de una publicación de Facebook.

Automatización del proceso:

La herramienta implementa un algoritmo que permite automatizar el proceso de extracción de información. Este algoritmo se encarga de desplazarse por la página web de Facebook mediante un método de selenium denominado `window.scroll()`, con ello lograr identificar y seleccionar cada una de las publicaciones tomando en cuenta que elemento se debe

Figura 33

Elemento de inicio de sesión de Facebook

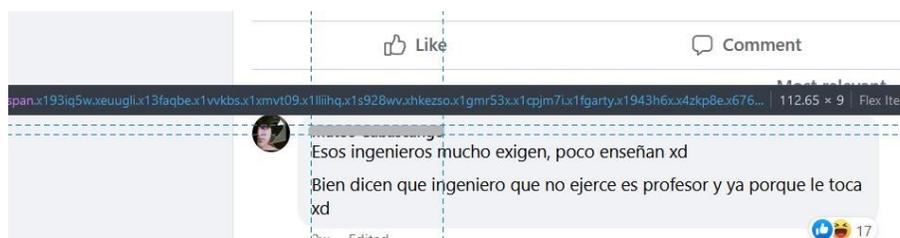


Nota. En la figura se observa el elemento que se genera al no haber realizado el inicio de sesión dentro de la publicación de Facebook donde el proceso principal es cerrarlo para proseguir con la extracción de datos.

Autor del Comentario: Elemento que muestra el nombre de las personas que realizan el comentario.

Figura 34

Autor del comentario

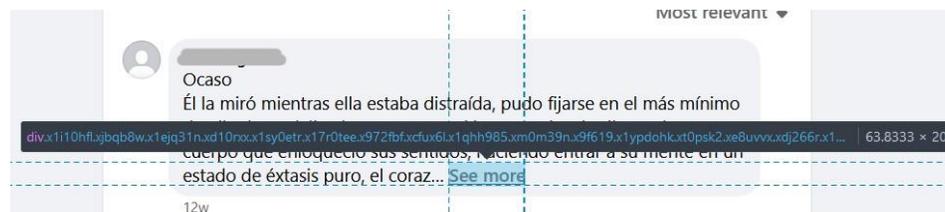


Nota. En la figura se observa un comentario completo donde se puede identificar el autor del comentario al igual que su elemento XML.

Ver más: Elemento que se genera cuando el texto se excede del tamaño dentro del comentario.

Figura 35

Elemento para mostrar el comentario completo



Nota. En la figura se observa el elemento XML para identificar si el comentario tiene una cadena de texto más larga.

Imágenes: Elemento que no genera información relevante dentro del comentario.

Determinando cada uno de estos elementos se logró realizar las validaciones para que la automatización en extracción de cada comentario sea eficaz un ejemplo de ello es la verificación de las imágenes, ya que esta no genera ninguna información en específico se descartan estos comentarios para la extracción.

Creación de la Interfaz de Usuario

En el desarrollo de la interfaz se optó por el uso de Flet ya que es un framework que nos permite crear aplicaciones web o desktop fácilmente y aunque sea relativamente nuevo, sus métodos para desarrollar ventanas adaptables a las necesidades del proyecto son completos y sencillos.

Para implementar Flet, se llevó a cabo un ajuste a la organización del proyecto. Donde cada archivo y función sigue la documentación oficial del framework, incluyendo el proceso de inicio del proyecto para la vista.

Figura 36

Script de arranque Flet

```

inicio.py > ...
You, 1 second ago | 2 authors (frame and others)
1 import flet as ft
2 from flet import Page
3 from analisis_sentimiento.controles import ComentariosApp
4
5 def create_app(page: Page):
6     page.title = "Análisis de Sentimientos"
7     page.horizontal_alignment = "center"
8     page.scroll = "adaptive"
9
10
11     # Crear la Instancia de ComentariosApp:
12     instance = ComentariosApp(page)
13     page.add(instance.initialize_ui())
14
15 def main():
16     ft.app(target=create_app)
17
18 if __name__ == "__main__":
19     main()

```

Nota. En la figura se observa el módulo principal para el arranque de la herramienta en Flet.

Como resultado, se logró obtener una vista principal que renderiza cada una de las funciones alojadas en el archivo de controles. En este directorio se definieron cada una de las funcionalidades para la observación de la herramienta.

Figura 37

Funciones de control

```

from flet import (TextField, ProgressBar, ElevatedButton
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.firefox.options import Options
from bs4 import BeautifulSoup
import time
import torch
from transformers import BertTokenizer, BertForSequenceClassification
import os

frame, 3 days ago | 1 author (frame)
class ComentariosApp(UserControl):
    def __init__(self, page): ...
    def initialize_ui(self): ...

    def predecir_emocion(self, texto): ...

    def get_comentarios(self, url): ...

    def analyze_data(self, url): ...

```

Nota. En la figura se observa cada una de las funciones específicas donde se realizan los procesos de análisis y creación de las vistas de la herramienta.

Cada una de las funciones ayuda a la actualización de la vista en función del proceso que se esté ejecutando. Finalmente, se presenta una interfaz limpia donde el primer valor a ingresar será la URL del grupo público de Facebook en la que se desea procesar los comentarios.

Figura 38

Interfaz de usuario de la herramienta



The image shows a dark-themed user interface for a tool titled "Análisis de Emociones en Entornos Digitales". At the top, the title is displayed in white. Below the title is a large, empty text input field with a light gray border and a placeholder text "Ingresa el nombre de la página de Facebook". Underneath the input field is a horizontal line, followed by a blue button with white text that says "Obtener Comentarios". Below the button is another horizontal line, and the bottom portion of the interface is currently blank.

Nota. En la figura se puede observar la interfaz final de la herramienta la cual cumple con una vista sencilla y fácil de usar.

Análisis de emociones con API de Vertex AI:

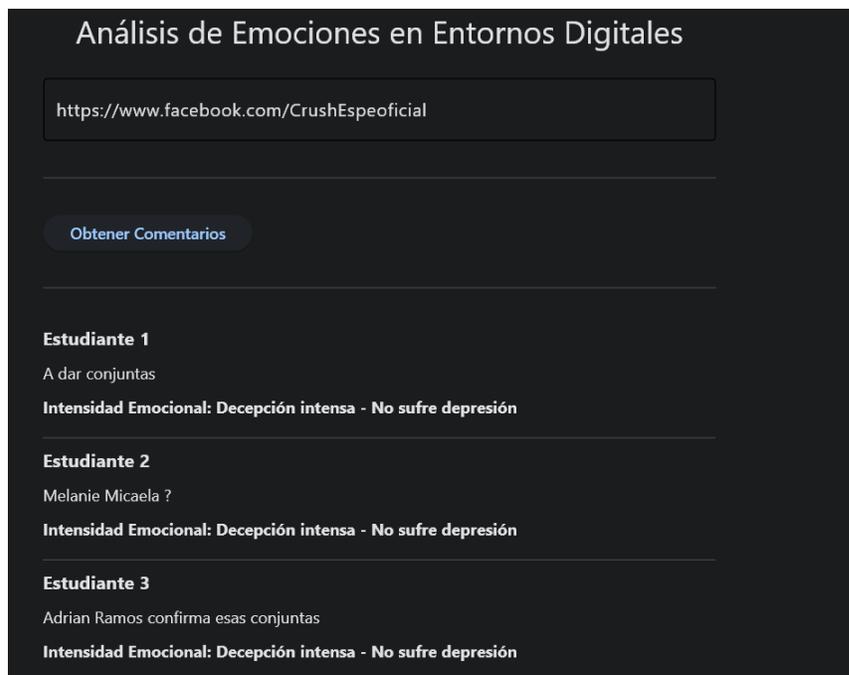
La herramienta luego de lograr obtener correctamente cada uno de los comentarios alojados en las publicaciones se procede con el análisis de emociones la cual consiste en enviar el texto generado en una lista para que analice mediante el modelo generado con Vertex AI y con ello obtener el análisis de las emociones.

Además, para determinar si una persona sufre o no depresión se necesita tomar en consideración que en la depresión intervienen varios factores que generan este trastorno mental, destacándose la combinación de varias emociones: tristeza, miedo, enojo, sorpresa, decepción,

vergüenza y dolor (Pérez-Padilla et al., 2017). Por esta razón, la herramienta utiliza la siguiente condición para realizar el cálculo de la depresión: si la suma de las emociones Miedo intenso, Sorpresa intensa, Tristeza intensa, Dolor intenso, Vergüenza intensa, Enojo intenso, Decepción intensa es mayor a 2, muestra el mensaje “Sufre depresión”, identificando así posibles casos de personas (en este caso estudiantes) que necesiten asistencia psicológica o social. El valor de 2 surge a raíz de un promedio, usando como base 10 textos donde claramente se expresaban palabras relacionadas con la depresión. Estos 10 textos fueron analizados con la herramienta y, al obtener un valor máximo de 2.0157 y un valor mínimo 1.8566, se establece un promedio de 1.9995 que equivale a 2. Al sumar las emociones anteriormente mencionadas, cumplimos con el principio de que la depresión surge a partir de la combinación de ciertas emociones.

Figura 39

Análisis de emociones en CrushEspeoficial



The screenshot displays a web interface for emotion analysis. At the top, the title "Análisis de Emociones en Entornos Digitales" is shown. Below it is a text input field containing the URL "https://www.facebook.com/CrushEspeoficial". A button labeled "Obtener Comentarios" is positioned below the input field. The results are presented in three sections, each for a different student:

- Estudiante 1**
A dar conjuntas
Intensidad Emocional: Decepción intensa - No sufre depresión
- Estudiante 2**
Melanie Micaela ?
Intensidad Emocional: Decepción intensa - No sufre depresión
- Estudiante 3**
Adrian Ramos confirma esas conjuntas
Intensidad Emocional: Decepción intensa - No sufre depresión

Nota. En la figura se observa que la herramienta obtiene las emociones de cada uno de los comentarios al igual el número de estudiantes analizados, identificando casos de asistencia psicológica y social a partir del mensaje “Sufre depresión”.

Capítulo IV

Evaluación y Resultados

En el presente capítulo se evidencia la evaluación de la propuesta presentada en el trabajo de titulación, para ello se han establecido dos secciones: i) Evaluación y ii) Resultados obtenidos. La evaluación del sistema consta de las respectivas pruebas de la herramienta propuesta, que, al resultar exitosa, permitieron obtener si existen comportamientos inusuales que generan los estudiantes dentro de los entornos virtuales y requieran asistencia psicológica o social.

Evaluación

Evaluación del modelo de aprendizaje Vertex AI

Tras la creación del modelo de entrenamiento, es necesario considerar los siguientes aspectos para un correcto funcionamiento del modelo:

- El dataset debe tener una cantidad considerable de datos, y por cada etiqueta debe existir como mínimo 10 datos para el entrenamiento con una proporción equitativa.
- La efectividad del modelo está relacionada con su tiempo de entrenamiento y generalmente es un proceso que puede durar varias horas.
- Para las pruebas realizadas desde la interfaz de Google Cloud, se utilizaron datos que no fueron parte de la muestra de publicaciones y comentarios de Facebook.
- Debido a que utilizamos un modelo desplegado en la nube, es necesario guardar los resultados del despliegue en una lista para poder acceder y trabajar con cada uno de los valores.

Tabla 11*Escala Visual Analógica (EVA)*

Emoción	Porcentaje	Valor sobre 1
Decepción intensa	4%	0.036
Enojo intenso	5%	0.051
Alegría intensa	1%	0.005
Dolor intenso	81%	0.814
Vergüenza intensa	1%	0.012
Tristeza intensa	0%	0.003
Neutral intenso	3%	0.030
Miedo intenso	2%	0.024
Amor intenso	0%	0.003
Sorpresa intensa	1%	0.011

Nota. Esta tabla muestra los resultados del texto “Astrid aún no lo supero”

Evaluación de la extracción de la información automatizada

Tras el desarrollo de la herramienta es necesario contemplar las siguientes consideraciones sobre el funcionamiento:

- Se tomo en cuenta que Facebook y otras redes sociales implementan medidas para detectar y prevenir actividades automatizadas no autorizadas, como el uso de herramientas automatizadas para realizar extracción de datos o acciones no permitidas.
- Las publicaciones privadas están diseñadas para ser vistas solo por un grupo selecto de personas, como amigos o amigos de amigos, por ello la extracción de la información se la realiza en publicaciones públicas ya que están destinadas a ser vistas por cualquier persona en la plataforma.
- En promedio, el tiempo para realizar la extracción de la información depende del desplazamiento que realiza selenium, la identificación de cada publicación por desplazamiento y las validaciones consideradas dentro del código.
- A continuación, se presentan los resultados obtenidos al usar la herramienta.

Tabla 12

Métricas analizadas para la extracción de datos

Métricas	Valor
Publicaciones Omitidas	6
Publicaciones Aceptadas	5
Comentarios totales	18
Comentarios Omitidos	1

Nota. Esta tabla muestra de manera general las métricas analizadas durante el proceso de la extracción de los datos en la herramienta.

Tabla 13

Comentarios extraídos con la herramienta

Autores	Comentarios
Estudiante 1	A dar conjuntas
Estudiante 2	No mames jajaja Es como decir vas a estudiar o a tomar?
Estudiante 3	Cagar
Estudiante 4	Recuerdo que cuando visité a mi abuela en nuestro pequeño pueblo, me dejó caer una gran sabiduría. Ella dijo: "Nieto, algún día algunas personas perderán el tiempo leyendo tu comentario". Hoy es ese día.
Estudiante 5	Hay un poco de foto en tus píxeles jaja
Estudiante 6	Una chica de cabello rojo también estaba linda
Estudiante 7	Y por qué no le preguntaste tu mismo ..?
Estudiante 8	Todo river mall
Estudiante 9	Vingala
Estudiante 10	Por el San Luis
Estudiante 11	Mi casa
Estudiante 12	Confirma Cristhian Gutierrez en tus épocas si hubieses invertido en ese negocio, ahora en Ferrari rodarías mijo
Estudiante 13	Esos ingenieros mucho exigen, poco enseñan xd
Estudiante 14	Razhiel Galvez jajaja están hablando del Lugobelto
Estudiante 15	Henry Hernan pero no ponga en anónimo ññ ya sabemos que es uste'
Estudiante 16	Francisco Barba Flores
Estudiante 17	Tan mal paga la universidad que los docentes tienen que cOim4r 2?

Nota. En la tabla se puede observar la información extraída de la herramienta.

Resultados

En esta sección se presentan los resultados obtenidos de la evaluación de la herramienta propuesta para el análisis de emociones donde se revelan patrones, tendencias y relaciones que determinan la diversidad de experiencias y emociones presentes en entornos digitales, destacando la relevancia de considerar la variabilidad y la complejidad en la interpretación de los comportamientos en línea.

Análisis de casos específicos con el modelo de prueba Vertex AI

Al examinar detenidamente casos específicos dentro del modelo de prueba de Vertex AI con el uso del dataset extraído tanto de comentarios como de publicaciones, se reveló una relación compleja entre la expresión emocional y el contenido. Se observó que los mismos términos o frases podían adquirir significados distintos según el contexto del comentario. Por ello se debe considerar la variabilidad en la interpretación de las emociones en el análisis de entornos digitales.

Resultados acerca del uso de la herramienta desarrollada

Los resultados derivados del uso de la herramienta desarrollada revelan un desempeño alentador, abarcando aproximadamente el 80% de la información necesaria según lo previsto. Se destaca la consideración de posibles contramedidas por parte de plataformas como Facebook para detectar actividades automatizadas no autorizadas. Además, se enfatiza que la extracción se realizó en publicaciones públicas, dado que las privadas están dirigidas a un grupo selecto de personas. El tiempo promedio para la extracción se ajustó según el desplazamiento de Selenium, la identificación de cada publicación y las validaciones del código. Los resultados indican que solo se omitieron seis publicaciones, se aceptaron cinco y se extrajo un total de 18 comentarios,

con tan solo un comentario omitido en el proceso. Estas métricas ofrecen una perspectiva optimista del rendimiento de la herramienta, cumpliendo efectivamente con la mayor parte de la información esperada en el proceso de extracción de datos.

Resultados del análisis de las emociones.

Los datos revelan un panorama emocional diverso en los comentarios analizados, destacando la decepción como la emoción más común. Este resultado sugiere que los comportamientos en entornos digitales tales como las redes sociales están marcadas por experiencias personales y emociones desafiantes para muchos usuarios. Por otro lado, la presencia significativa de la alegría indica que los estudiantes tienden a ser positivos y disfrutar de momentos divertidos. El enojo, aunque menos frecuente, revela la presencia de situaciones frustrantes que generan respuestas emotivas intensas, mientras que el miedo refleja la preocupación por eventos inciertos. Estos resultados resaltan la complejidad emocional en las interacciones en los entornos digitales y subrayan la importancia de que los estudiantes que tienden a tener emociones intensas requieran de asistencia psicológica y social.

Resultados de la relación entre emociones y contenido mediante la herramienta

Se observó una correlación significativa entre las emociones detectadas y el contenido de los comentarios analizados. Aquellos que expresaban tristeza tendían a utilizar palabras asociadas con la pérdida, el dolor, la soledad o la decepción. Por otro lado, los comentarios que transmitían alegría mencionaban términos vinculados a la felicidad, la satisfacción, el éxito o la celebración. Asimismo, los comentarios cargados de enojo se caracterizaban por palabras relacionadas con la ira, la frustración, la injusticia o la queja. Por último, los comentarios que expresaban miedo se identificaron a través de términos asociados con la incertidumbre, el peligro, la amenaza o la ansiedad. La herramienta muestra el mensaje "Sufre depresión" si la suma de las emociones Miedo intenso, Sorpresa intensa, Tristeza intensa, Dolor intenso,

Vergüenza intensa, Enojo intenso, Decepción intensa es mayor a 2. El valor de 2 se obtiene del promedio de la suma de esas emociones de 10 textos que mostraban claramente palabras relacionadas con la depresión. Calculando la suma de cada uno de esos textos, se obtuvo que el valor máximo de la suma de las emociones anteriormente mencionadas es de 2.0157 y el mínimo de 1.8566, finalmente obteniendo un promedio de 1.9995 que equivale a 2.

Tabla 14

Resultados del análisis de emociones extraídos con la herramienta

Autores	Comentarios	Emoción Detectada
Estudiante 1	A dar conjuntas	Intensidad Emocional: Decepción intensa - No sufre depresión
Estudiante 2	No mames jajaja Es como decir vas a estudiar o a tomar?	Intensidad Emocional: Neutral leve - No sufre depresión
Estudiante 3	Cagar	Intensidad Emocional: Enojo intenso - No sufre depresión
Estudiante 4	Recuerdo que cuando visité a mi abuela en nuestro pequeño pueblo, me dejó caer una gran sabiduría. Ella dijo: "Nieto, algún día algunas personas perderán el tiempo leyendo tu comentario". Hoy es ese día.	Intensidad Emocional: Decepción intensa - No sufre depresión
Estudiante 5	Hay un poco de foto en tus píxeles jaja	Intensidad Emocional: Neutral leve - No sufre depresión
Estudiante 6	Una chica de cabello rojo también estaba linda	Intensidad Emocional: Decepción intensa - No sufre depresión
Estudiante 7	Y por qué no le preguntaste tu mismo ..?	Intensidad Emocional: Decepción intensa - No sufre depresión
Estudiante 8	Todo river mall	Intensidad Emocional: Neutral intenso - No sufre depresión - No refleja emoción clara
Estudiante 9	Vingala	Intensidad Emocional: Neutral intenso - No sufre depresión - No refleja emoción clara

Autores	Comentarios	Emoción Detectada
Estudiante 10	Por el San Luis	Intensidad Emocional: Neutral intenso - No sufre depresión - No refleja emoción clara
Estudiante 11	Mi casa	Intensidad Emocional: Neutral intenso - No sufre depresión - No refleja emoción clara
Estudiante 12	Confirma Cristhian Gutierrez en tus épocas si hubieses invertido en ese negocio, ahora en Ferrari rodarías mijo	Intensidad Emocional: Enojo intenso - No sufre depresión
Estudiante 13	Esos ingenieros mucho exigen, poco enseñan xd	Intensidad Emocional: Decepción Intensa - No sufre depresión
Estudiante 14	Razhiel Galvez jajaja están hablando del Lugobelto	Intensidad Emocional: Alegría intensa - No sufre depresión
Estudiante 15	Henry Hernan pero no ponga en anónimo ññ ya sabemos que es uste'	Intensidad Emocional: Decepción intensa - Sufre depresión
Estudiante 16	Francisco Barba Flores	Intensidad Emocional: Neutral intenso - No sufre depresión - No refleja emoción clara
Estudiante 17	Tan mal paga la universidad que los docentes tienen que cOim4r 2?	Intensidad Emocional: Decepción intensa - No sufre depresión

Nota. En la tabla se puede observar la información extraída con la herramienta al igual que la emoción analizada a través del modelo entrenado.

Capítulo V

Conclusiones y Recomendaciones

Conclusiones

Tras el desarrollo y evaluación de la propuesta presentada en este trabajo de titulación, se pueden extraer conclusiones relevantes en relación con los objetivos planteados. En primer lugar, la evaluación del modelo de aprendizaje Vertex AI reveló que, para un correcto funcionamiento, es esencial contar con un dataset considerable, asegurando al menos 10 datos por etiqueta para el entrenamiento equitativo. Se destaca la importancia del tiempo de entrenamiento, un proceso que generalmente demanda varias horas, y la necesidad de utilizar datos externos para pruebas desde la interfaz de Google Cloud.

Los resultados obtenidos del uso de la herramienta desarrollada son alentadores, alcanzando aproximadamente el 80% de la precisión en el entrenamiento se logró obtener una diversidad emocional en los comentarios analizados donde se subraya la complejidad de las interacciones en entornos digitales, especialmente destacando la decepción como la emoción más común.

El análisis de casos específicos con el modelo de prueba Vertex AI muestra la complejidad en la relación entre expresión emocional y contenido, evidenciando la variabilidad en la interpretación de emociones en entornos digitales.

A pesar de los resultados alentadores y el cumplimiento de los objetivos establecidos en este trabajo de titulación, es importante señalar que el desarrollo de la herramienta presenta oportunidades de mejora significativas. La interfaz actual, aunque funcional, es simple y puede beneficiarse de mejoras para aumentar la usabilidad y experiencia del usuario.

Recomendaciones

Con toda la información obtenida del proceso de desarrollo y evaluación se han establecido las siguientes recomendaciones:

- En relación con el modelo de aprendizaje Vertex AI, se observa que, a pesar de los buenos resultados, la precisión podría mejorarse. Actualmente, se ha alcanzado aproximadamente el 80% de precisión, pero se recomienda trabajar en la optimización de los conjuntos de datos para lograr un aumento en la precisión hasta alcanzar el 90% o más. Esto podría lograrse mediante la expansión y diversificación del dataset de entrenamiento, asegurando una representación equitativa de todas las etiquetas y escenarios posibles.
- Se recomienda realizar una revisión centrada en el usuario para identificar áreas de mejora. La implementación de un diseño más intuitivo, con características de personalización y adaptabilidad, podría elevar la usabilidad y la experiencia general del usuario. Asegurarse de que la interfaz sea accesible para una variedad de usuarios, considerando incluso aquellos con necesidades especiales, puede ampliar la utilidad de la herramienta.

Trabajos Futuros

El presente trabajo de investigación y el proyecto propuesto han abierto la puerta a diversas posibilidades de investigación y desarrollo en el campo del análisis de emociones en entornos virtuales de aprendizaje. A pesar de los avances logrados y los resultados obtenidos, existen áreas que no fueron exploradas en profundidad debido a las limitaciones de alcance establecidas en este proyecto. Por tanto, se identifican líneas de investigación prometedoras que

podrían ser abordadas en futuros trabajos, ampliando así el conocimiento y las aplicaciones prácticas en este campo emergente. A continuación, se presentan algunos temas que representan oportunidades para investigaciones futuras, con el potencial de enriquecer y expandir el impacto de este proyecto inicial.

- Mejora de la precisión del modelo de aprendizaje donde se continúa trabajando en la optimización de los conjuntos de datos para aumentar la precisión del modelo hasta alcanzar el 90% o más. Esto podría implicar la expansión y diversificación del dataset de entrenamiento, asegurando una representación equitativa de todas las etiquetas y escenarios posibles.
- Implementación de retroalimentación automática para desarrollar un sistema de retroalimentación automática basado en las emociones detectadas, que pueda proporcionar apoyo y orientación personalizada a los estudiantes en tiempo real.
- Revisión centrada en el usuario donde se realiza una revisión centrada en el usuario para identificar áreas de mejora en la interfaz de la herramienta. Implementar un diseño más intuitivo, con características de personalización y adaptabilidad, para elevar la usabilidad y la experiencia general del usuario. Asegurarse de que la interfaz sea accesible para una variedad de usuarios, considerando incluso aquellos con necesidades especiales, puede ampliar la utilidad de la herramienta.
- Validación y escalabilidad del modelo: Realizar pruebas adicionales para validar la efectividad del modelo en diferentes contextos educativos y con una mayor cantidad de usuarios, evaluando su escalabilidad y adaptabilidad a diferentes entornos.

Bibliografía

Amazon. (2024). *¿Qué es Python?* <https://aws.amazon.com/es/what-is/python/>.

Apify. (2024). *Apify Blog*. <https://blog.apify.com>.

AWS. (2024). *¿Qué es extracción, transformación y carga (ETL)?* <https://aws.amazon.com/es/what-is/etl/>.

Chamorro-Atalaya, O., Arce-Santillan, D., Morales-Romero, G., León-Velarde, C., Ramos-Salaza, P., Auqui-Ramos, E., & Levano-Stella, M. (2022). Sentiment analysis through twitter as a mechanism for assessing university satisfaction. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(1), 430. <https://doi.org/10.11591/ijeecs.v28.i1.pp430-440>

Ganti, V., & Das Sarma, A. (2013). *Data cleaning : a practical perspective*.

IBM. (2024). *Análisis exploratorio de datos*. <https://www.ibm.com/mx-es/topics/exploratory-data-analysis>.

Kelmansky, & Diana M. (2009). *ESTADÍSTICA PARA TODOS Estrategias de pensamiento y herramientas para la solución de problemas*.

Microsoft. (2024). *Power BI*. <https://www.microsoft.com/es-es/power-platform/products/power-bi>.

Ochoa Quispe, F. Z., & Barragán Condori, M. (2022). El uso en exceso de las redes sociales en medio de la pandemia. *ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades*, 9(1), 85–92. <https://doi.org/10.30545/academo.2022.ene-jun.8>

Pérez-Padilla, E. A., Cervantes-Ramírez, V. M., Hijuelos-García, N. A., Pineda-Cortes, J. C., & Salgado-Burgos, H. (2017). Prevalencia, causas y tratamiento de la depresión mayor. *REVISTA BIOMÉDICA*, 28(2). <https://doi.org/10.32776/revbiomed.v28i2.557>

Selenium. (2024). *Selenium*. <https://www.selenium.dev>.

Valentine Enedah. (2024). *Una breve introducción a FLET: creación de aplicaciones de Flutter con Python*. <https://hackernoon.com/es/una-breve-introduccion-a-la-construccion-de-aplicaciones-de-flutter-con-python-flet>.

Apéndices