**Sistema experto que identifica sintomatología ansiosa que afecta al rendimiento académico de los estudiantes del Instituto de Ciencias Básicas e Ingeniería de la UAEH, durante el ciclo escolar Enero-Junio 2024**

Ulises Sebastian García García

Universidad Autónoma del Estado de Hidalgo

Instituto de Ciencias Básicos e Ingeniería

Licenciatura en Ciencias Computacionales

Fundamentos de Metodología de la Investigación

M.M.P.H. Edrein Marcela Aguilar Ramírez

29 de abril de 2024

Resumen

La salud mental de los estudiantes universitarios es una preocupación creciente, particularmente en lo que respecta a la ansiedad, que ha aumentado significativamente entre los adultos jóvenes. En el Instituto de Ciencias Básicas e Ingeniería (ICBI) de la Universidad Autónoma del Estado de Hidalgo (UAEH), la ansiedad es un problema frecuente que puede afectar seriamente el rendimiento académico. Este proyecto tiene como objetivo desarrollar un sistema experto para diagnosticar la sintomatología ansiosa en los estudiantes del ICBI. Este sistema permitirá la identificación temprana y precisa de los síntomas de ansiedad, facilitando intervenciones oportunas y efectivas. La implementación de este sistema experto se espera que contribuya de manera significativa al bienestar y rendimiento académico de los estudiantes, promoviendo un entorno saludable y favorable para su desarrollo integral.

*Palabras clave:* Salud mental, ansiedad, estudiantes universitarios, rendimiento académico, sistema experto, diagnóstico, intervención temprana, UAEH, ICBI.

Introducción

Hoy en día, la salud mental de la población se ha convertido en un tema de creciente preocupación, además de que la ansiedad ha aumentado de manera significativa en adultos jóvenes en los últimos años, y los estudiantes del Instituto de Ciencias Básicas e Ingeniería de la Universidad Autónoma del Estado de Hidalgo no han sido la excepción.

La ansiedad es uno los malestares que más impacto tiene en la salud mental de los universitarios. Se caracteriza por preocupaciones excesivas, nerviosismo, tensión y síntomas físicos como taquicardias, sudoración y dificultad para mantener el foco de atención.. Cuando estos síntomas alcanzan niveles significativos, pueden repercutir de manera considerable en la capacidad de los estudiantes para cumplir con sus responsabilidades académicas y tener un buen desempeño académico.

A pesar de la relevancia del problema, hasta el momento es difícil abordarlo puesto que aún no existe una verdadera cultura que fomente el cuidado de la salud mental, y por lo tanto en su mayoría las personas que acuden a terapia psicológica son las únicas que pueden sobrellevar el problema. Además, hasta el momento no se ha implementado un sistema de diagnóstico específico para identificar y abordar la sintomatología ansiosa entre los estudiantes del Instituto de Ciencias Básicas e Ingeniería. La falta de herramientas especializadas dificulta la detección temprana y sobre todo, la intervención oportuna, lo que puede resultar en un deterioro progresivo del desempeño académico, aunado a un impacto negativo en la salud mental y emocional de los estudiantes.

Dadas las circunstancias, el presente proyecto de investigación tiene como objetivo desarrollar un sistema experto para el diagnóstico de la sintomatología ansiosa en los estudiantes del Instituto de Ciencias Básicas e Ingeniería de la UAEH. Este sistema estará diseñado para identificar de manera precisa y eficiente los síntomas de ansiedad que afectan al rendimiento académico de estudiantes universitarios, proporcionando una herramienta para la detección temprana y la intervención adecuada.

A través del desarrollo de este sistema, se espera obtener una contribución significativa en el ámbito de la salud mental estudiantil, proporcionando una solución innovadora y efectiva para abordar la problemática identificada. El propósito fundamental es mejorar el bienestar y el rendimiento académico de los universitarios, promoviendo un entorno saludable y propicio para el desarrollo integral de los estudiantes del ICBI.

Índice

[Resumen 2](#_Toc196571323)

[Introducción 3](#_Toc196571324)

[Índice 5](#_Toc196571325)

[Índice de tablas 7](#_Toc196571326)

[Índice de figuras 8](#_Toc196571327)

[Capitulo I. Planteamiento del problema 9](#_Toc196571328)

[Antecedentes 9](#_Toc196571329)

[Justificación 12](#_Toc196571330)

[Objetivos 13](#_Toc196571331)

[Objetivo general 13](#_Toc196571332)

[Objetivos específicos 13](#_Toc196571333)

[Capítulo II. Marco teórico, conceptual, y/o referencial 14](#_Toc196571334)

[Marco Conceptual 14](#_Toc196571335)

[Introducción al Procesamiento del Lenguaje Natural y las Emociones 14](#_Toc196571336)

[NLP 14](#_Toc196571337)

[IA y las Emociones 14](#_Toc196571338)

[Utilidad del Análisis de Sentimientos 14](#_Toc196571339)

[Beneficios 14](#_Toc196571340)

[Bases Teóricas de las Emociones en el Lenguaje 14](#_Toc196571341)

[La Emoción en el Lenguaje Natural 14](#_Toc196571342)

[Diferencia entre Emoción y Sentimiento 14](#_Toc196571343)

[Clasificación de las Emociones 14](#_Toc196571344)

[Modelo de las Emociones Básicas de Ekman 14](#_Toc196571345)

[Clasificación de Emociones en Sentimientos: Positivo, Negativo, Ambiguo y Neutral 14](#_Toc196571346)

[Expresiones Multietiqueta de Emociones 14](#_Toc196571347)

[Métodos Computacionales para la Detección de Emociones 14](#_Toc196571348)

[Técnicas de Minería de Datos para Clasificación de Emociones 14](#_Toc196571349)

[Representación del Lenguaje Natural: Bag of Words y TF-IDF 14](#_Toc196571350)

[Modelos Supervisados de Clasificación de Texto 14](#_Toc196571351)

[Métricas para Evaluación de Clasificación 15](#_Toc196571352)

[Regresión Logística. 15](#_Toc196571353)

[Lenguaje de Programación Python 15](#_Toc196571354)

[Dataset y Fuentes de Datos 15](#_Toc196571355)

[El Dataset GoEmotions 16](#_Toc196571356)

[La Emoción en el Lenguaje Natural 16](#_Toc196571357)

[Diferencia entre Emoción y Sentimiento 17](#_Toc196571358)

[Clasificación de las Emociones 17](#_Toc196571359)

[Modelo de las Emociones Básicas de Ekman 18](#_Toc196571360)

[Utilidad del análisis de sentimientos 18](#_Toc196571361)

[IA y las emociones 18](#_Toc196571362)

[NLP 18](#_Toc196571363)

[Beneficios 18](#_Toc196571364)

[Antecedentes (IMDB, yelp, GoEmotions es el actual) 18](#_Toc196571365)

[Clasificación de Emociones en Sentimientos: Positivo, Negativo, Ambiguo y Neutral 18](#_Toc196571366)

[Expresiones Multietiqueta de Emociones 18](#_Toc196571367)

[Clasificación Multiclase y Multietiqueta 18](#_Toc196571368)

[Técnicas de Minería de Datos Aplicadas a Emociones 18](#_Toc196571369)

[Representación del Lenguaje Natural: Bag of Words y TF-IDF 18](#_Toc196571370)

[Modelos Supervisados de Clasificación de Texto 18](#_Toc196571371)

[Métricas para Evaluación de Clasificación 18](#_Toc196571372)

[Experimentos y resultados 19](#_Toc196571373)

[Tecnología para la detección y seguimiento de la ansiedad 19](#_Toc196571374)

[Cronograma 22](#_Toc196571375)

[Referencias 23](#_Toc196571376)

[Notas al pie 25](#_Toc196571377)

[Tablas 26](#_Toc196571378)

[Figuras 27](#_Toc196571379)

[Anexos 28](#_Toc196571380)

Índice de tablas

Índice de figuras

Capitulo I. Planteamiento del problema

Antecedentes

Según Ecured (s.f), la ansiedad es una emoción cotidiana que podemos sentir en situaciones de miedo, de angustia, de peligro, cuando algo nos causa inquietud o incertidumbre. En ocasiones, puede llegar a manifestarse con algún síntoma físico como lo puede ser la sudoración, taquicardias, adormecimiento de extremidades, náuseas, despersonalización, entre otros.

Diversos factores influyen en el desarrollo de la sintomatología ansiosa en los jóvenes universitarios, Manley (2022) plantea que a través de un estudio de más de 6000 participantes universitarios, se llegó a la conclusión de que más del 60% de las causas del estrés o ansiedad dentro de la población de estudio fueron por factores relacionados con la salud física, la salud mental y el desempeño académico, por ejemplo:

* Promedios bajos.
* Baja o nula actividad física.
* Síntomas depresivos o relacionados a la tristeza.
* Descuido de la higiene personal y salud física.

Además, según las encuestas que aplicó a más de 200 estudiantes se puede llegar a varias conclusiones las cuales se citan a continuación:

Una de las preocupaciones más grandes del estudiantado es su capacidad intelectual, seguido por su estabilidad emocional y la dificultad para crear vínculos con otras personas.

Más del 60% considera que su rendimiento académico se puede ver influenciado por algún problema de salud mental, por lo que hay una estrecha correlación entre la sintomatología ansiosa y el bajo desempeño escolar.

Ahora bien, hay una fuerte correspondencia entre la ansiedad y las Tecnologías de la Información y Comunicaciones (TIC), Zavala (2017) afirma que más de una cuarta parte de la población mexicana cuenta con acceso a algún dispositivo relacionado a la TIC, ya sea un televisor, un teléfono celular o una computadora, desde la implementación de estos dispositivos la sociedad y los comportamientos de la misma se han visto influenciados por los medios. En concreto define a la sociedad como la “sociedad del conocimiento” por lo que adquiere gracias a estas tecnologías. Al usar constantemente estos dispositivos, diversos aspectos de la vida se pueden ver afectados, algunos de ellos pueden incluir hipersensibilidad electromagnética, sedentarismo, problemas en articulaciones y dolores musculares, pero también estrés, y en algunos casos, se genera dependencia y por lo tanto, termina desencadenando en ansiedad.

Si las causas de la ansiedad son recurrentes pueden terminar desencadenando en un trastorno de ansiedad generalizada, por ello es importante detectar los factores causantes de la ansiedad, pues parte de ellos se basan en la experiencia individual, según Díaz y Castro (2020):

El nivel del rendimiento académico puede ser producto de eventos o situaciones que generan ansiedad en los estudiantes universitarios, poniendo en constatación la propia realidad de los sujetos, sin embargo, esta realidad se encuentra personalizada, es decir, es subjetiva; pues para algunos sujetos el desajuste se plantea respecto a lo novedoso, para otros se halla siempre presente en las actividades que se realizan, dudando y sufriendo de manera continua; para otros el desajuste y la aparición ansiosa sólo tienen lugar ante la presencia de determinados elementos en su contexto, en su realidad. (p. 8).

Hoy en día existen múltiples herramientas enfocadas en el diagnóstico de sintomatología ansiosa, como lo es el sistema experto basado en el test psicológico “Inventario de Ansiedad de Beck” (Vilcauri y Rodríguez, 2019), e incluso se ha llegado a implementar la detección de ansiedad por medio del uso de sensores vestibles (Cuevas, 2020).

Justificación

El motivo principal de la presente investigación sobre la identificación de sintomatología ansiosa que afecta al rendimiento académico de los estudiantes del Instituto de Ciencias Básicas e Ingeniería de la UAEH se deriva de estudios recientes que remarcan el crecimiento de la ansiedad en la población adulta en México, y que por lo tanto es un problema de salud pública. Muchas personas que presentan esta sintomatología no son conscientes de ello, ya sea porque no tienen los recursos para ir al psicólogo, por falta de tiempo, falta de interés o simplemente lo pasan por desapercibido, es por ello por lo que la investigación necesaria para el desarrollo de un sistema de diagnosis puede ayudar a que los estudiantes del ICBI se percaten de ello y puedan empezar a trabajar este padecimiento, ya sea por sus propios medios, o asistiendo formalmente a consulta para tener un diagnóstico oficial, pues el propósito del sistema es ser una guía para el diagnóstico de su ansiedad y la correlación que puede haber con su desempeño escolar, es importante aclarar, que el sistema no es una solución directa a sus problemas, solo encamina al usuario para que pueda atender la problemática.

Objetivos

Objetivo general

Plantear las bases para un sistema experto que identifique sintomatología ansiosa que pueda afectar al rendimiento académico de los estudiantes del Instituto de Ciencias Básicas e Ingeniería de la UAEH, durante el ciclo escolar Enero-Junio 2024.

Objetivos específicos

1. Identificar la sintomatología ansiosa que afecta en el rendimiento académico de los estudiantes universitarios.
2. Conocer los síntomas ansiosos que presentan los estudiantes universitarios de bajo rendimiento académico.
3. Analizar los síntomas ansiosos que presentan los estudiantes universitarios de alto desempeño.
4. Estudiar la diferencia entre síntomas presentados por alumnos de alto y rendimiento académico.
5. Determinar la relación entre los síntomas presentados por los estudiantes objeto de estudio y el rendimiento académico de los mismos.
6. Establecer el conocimiento necesario para el sistema.
7. Desarrollar un sistema experto para el diagnóstico de sintomatología ansiosa con base en el conocimiento obtenido.

Capítulo II. Marco teórico, conceptual, y/o referencial

Introducción al Procesamiento del Lenguaje Natural y las Emociones

NLP

De acuerdo con Alija (2020)., el Procesamiento del Lenguaje Natural (NLP) se basa en la necesidad de que las máquinas entiendan el lenguaje humano, ya sea de forma textual u oral, este campo híbrido entre la informática y la lingüística, hace uso de herramientas tales como la Inteligencia Artificial para guiar esta interpretación.

IA y las Emociones

En los últimos años, el uso y aplicación de la inteligencia artificial se ha expandido en múltiples áreas de investigación, y el campo de las emociones ha sido uno de los más destacados. Aunque las computadoras son incapaces de percibir las emociones humanas de forma natural, debido a que su lógica de procesamiento se basa en representaciones binarias (ceros y unos), se han desarrollado técnicas y herramientas para aproximarse a esta tarea de reconoocimiento. Entre ellas, destaca la **minería de opiniones**, la cual “es una subdisciplina de la lingüística computacional que se centra en extraer la opinión de las personas de la web” (Nigro, 2020). Esta técnica permite que mediante técnicas computacionales, no solo permite determinar qué dice el autor de un comentario, sino también el tono/actitud en que este está siendo expresado.

Utilidad del Análisis de Sentimientos

Dentro de este marco, el análisis de sentimientos se consolida como una aplicación directa de la minería de opiniones, orientada a interpretar la carga afectiva de los mensajes expresados en lenguaje natural. Gracias a esta subdisciplina, es posible clasificar de forma automática la polaridad de grandes volúmenes de información de acuerdo con su carga emocional (positiva o negativa), permitiendo aportes en diversas áreas del conocimiento.

Por ejemplo, Nigro (2020) menciona que la minería de opiniones puede aplicarse al campo de la gobernanza, permitiendo mediante el uso de ciertos programas y herramientas, que un votante pueda determinar qué partido político se asemeja más a sus opiniones personales. No obstante, este análisis automatizado no solo se limita al campo político, sino que también permite generar información de utilidad para campos como la mercadotecnia, la sociología, monitoreo de tendencias, etc.

Bases Teóricas de las Emociones en el Lenguaje

La Emoción en el Lenguaje Natural

El lenguaje es uno de los medios principales a través del cual podemos expresar nuestras emociones. Según Soriano (2016), las emociones no se representan de manera directa, sino que se hace a través de recursos simbólicos como lo es la metáfora, entre otras expresiones ligüísticas que reflejan conceptos subyacentes. Estas metáforas permiten definir estados emocionales abstractos a través de dominios concretos, por ejemplo, relacionar el fuego con la ira o la tristeza con una caída. En este tipo de expresiones, el lenguaje no solo ayuda a comunicar lo que sentimos, sino también lo que pensamos sobre lo que sentimos.

Este componente simbólico es imprescindible para el análisis computacional del lenguaje natural, ya que permite detectar la presencia de emociones aún cuando no se mencionan de forma implícita dentro del mensaje. Por dicha razón, el presente proyecto no solo considera la detección de términos explícitos, sino también el reconocimiento de estructuras que conllevan a la conceptualización de emociones implícitas.

Diferencia entre Emoción y Sentimiento

Damasio (2003), diferencia la emoción y el sentimiento indicando que la primera se trata de una respuesta biológica automática y son expresadas mediante el cuerpo, mientras que el segundo es la respuesta mental hacia una emoción. Esto permite determinar que una emoción precede al sentimiento, ya que estas se construyen de reacciones corporales simples ante estímulos externo, sin la necesidad de ser catalogadas como un sentimiento.

De igual manera, Damasio menciona que un sentimiento emerge cuando el cerebro es capaz de interpretar dichos cambios corporales, es decir, muchas veces el cuerpo reacciona de manera insconsciente, sin embargo, cuando el organismo percibe dichos cambios, la experiencia consciente se define como el sentimiento.

Clasificación de las Emociones

A lo largo del tiempo, diversos autores han propuesto formas clasificar las emociones, ya sea desde un enfoque categórico o bien dimensional, para el desarrollo del proyecto se consideran principalmente dos modelos teóricos que guardan similitudes con la estructura adoptada en el dataset GoEmotions y su artículo de referencia, los cuales se de forma posterior.

Robert Plutchik (1980) creó la rueda de las emociones, la cual determina la interrelación entre las emociones del ser humano, misma que consiste en ocho emociones básicas organizadas en pares opuestos (alegría-tristeza, ira-miedo, confianza-desagrado, anticipación-sorpresa), las cuales pueden combinarse dando a lugar emociones aún más complejas.

Por otra parte, Russell (1989) propuso un modelo dimensional de las emociones a través de un plano de cuatro cuadrantes dividido por dos ejes: valencia (negativa/positiva) y agitación (alta/baja). Este modelo permite que al llevar a cabo tareas computacionales, las emociones puedan mapearse de acuerdo con su carga afectiva y nivel de excitación.

Modelo de las Emociones Básicas de Ekman

Ekman (1992) planteó que las emociones son el resultado de un proceso evolutivo y adaptativo, en el cual cada emoción posee características únicas que las diferencian de otras experiencias humanas. A partir de dicha hipótesis, clasificó seis emociones básicas: ira, asco, miedo, felicidad, tristeza y sorpresa. De igual manera, determinó que toda otra experiencia que no cumpliera con dichas características, no sería considerada una emoción, sino estados de ánimo.

Clasificación de Emociones en Sentimientos: Positivo, Negativo, Ambiguo y Neutral

¿??

Métodos Computacionales para la Detección de Emociones

Técnicas de Minería de Datos para Clasificación de Emociones (vectorización)

Representación del Lenguaje Natural: Bag of Words y TF-IDF

Selección de características: Chi2

Modelos Supervisados de Clasificación de Texto

Regresión Logística. Para

Máquinas de Vectores de Soporte (SVM). Para

Árboles de Decisión. Para

Métricas para Evaluación de Clasificación

Accuracy. Para

F1-Score. Para

Matriz de Confusión. Para

Datasets y Recursos para el análisis de emociones

Dataset y fuentes de datos

Para llevar a cabo el análisis de emociones a partir de frases, fue necesario el uso de un conjunto estructurado de datos que permitiese la aplicación de técnicas de minería de datos y machine learning, esto justifica el uso de un ‘dataset’, que, de acuerdo con la definición de la Fundación Española para la Ciencia y la Tecnología, FECYT (2012):

Se define como una colección de datos reunidos durante la ejecución de un proyecto de investigación. Los datasets son objetos digitales compuestos y heterogéneos. Es decir, pueden comprender diferentes elementos o tipos de datos: documentos de texto, hojas de cálculo, ficheros de operaciones matemáticas, gráficos, imágenes, etc.

Otros datasets (IMDB, Yelp)

El dataset GoEmotions

Para el desarrollo del proyecto se utilizó el dataset “GoEmotions”, desarrollado por Demsky et al. (2020), el cual está conformado por 58 mil comentarios en inglés extraídos de subreddits populares de Reddit, los cuales fueron etiquetados manualmente con 27 categorías de emociones. A diferencia de otros datasets, GoEmotions contempla 12 emociones positivas, 11 negativas, 4 ambiguas y 1 neutral, lo cual lo convierte en una herramienta apta para tareas en las que se requiere comprensión conversacional, es decir, distinguir diferencias sutiles dentro de expresiones.

Lenguaje de Programación Python

Para el desarrollo del presente proyecto, se utilizó el lenguaje de programación Python,

debido a su versatilidad, eficiencia y amplia variedad y disponibilidad de librerías enfocadas en minería de datos y procesamiento del lenguaje natural. Tal y como lo define Amazon Web Services (s. f.), “Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning (ML).” Es por ello por lo se convirtió en la elección más apropiada para el desarrollo de este análisis.

Experimentos y resultados

Features

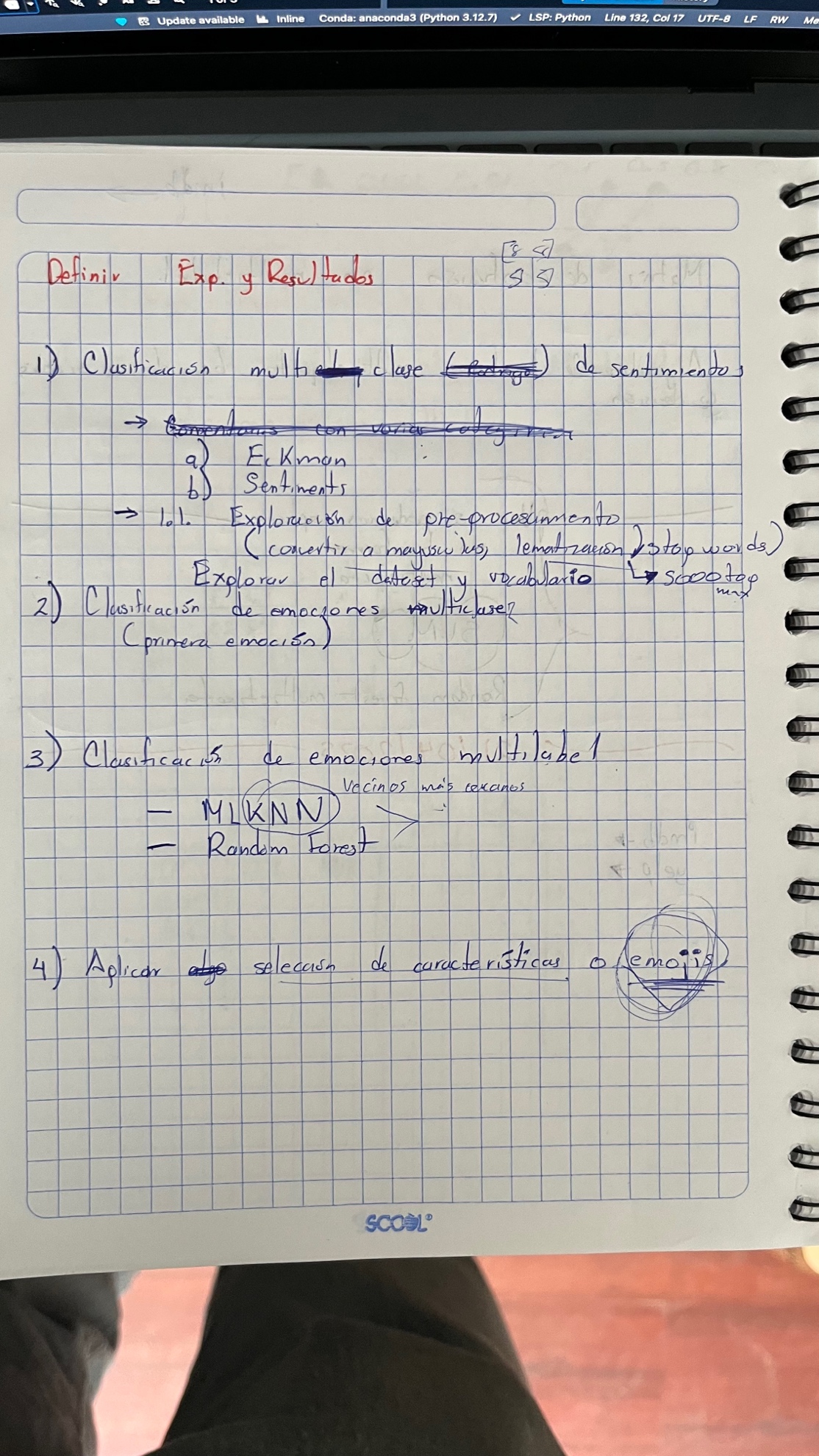
Clasificación multiclase de sentimientos

BasedOnEkman

SentimentsLevel

Clasificación de emociones multiclase

Clasificación de emociones multietiqueta



Cronograma

Gráfico

Descripción generada automáticamente con confianza mediaEl cronograma detallado a continuación abarca 48 semanas y se divide en las siguientes fases: identificación y análisis de síntomas ansiosos, recopilación y análisis de datos mediante encuestas, establecimiento del conocimiento necesario para el sistema, y el diseño, desarrollo, pruebas y ajustes del sistema experto. Este plan asegura que todas las actividades se completen de manera eficiente y oportuna, cumpliendo con los objetivos del proyecto.

**Figura 1** *Cronograma de actividades*

*Nota:* La figura muestra las actividades designadas para el desarrollo del sistema experto

Referencias

¿Qué es Python? - Explicación del lenguaje Python - AWS. (s. f.). Amazon Web Services, Inc. https://aws.amazon.com/es/what-is/python/

Alon, D., & Ko, J. (2021, 21 octubre). GoEmotions: A Dataset for Fine-Grained Emotion Classification. Google Research. https://research.google/blog/goemotions-a-dataset-for-fine-grained-emotion-classification/

Damasio, A. R. (2003). Looking for Spinoza: joy, sorrow, and the feeling brain. Harcourt.

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., Ravi, S. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. arXiv preprint arXiv:2005.00547.

Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas [en línea] Madrid: Fundación Española para la Ciencia y la Tecnología, FECYT (2012)

Hatton, L., & Warr, G. (2025). The architecture of the genome integrates scale independence with inverse symmetry. Academia Molecular Biology and Genomics, 2(2). https://doi.org/10.20935/AcadMolBioGen7650

Soriano, C. (2016). El lenguaje de las emociones. En M. C. Horno Chéliz, I. Ibarretxe Antuñano & J.L. Mendívil Giró (Eds.). Panorama actual de la ciencia del lenguaje (pp. 243-259). Zaragoza: Prensas de la Universidad de Zaragoza.

Alija, A. (2020). Tecnologías emergentes y datos abiertos: Procesamiento de lenguaje natural. Gobierno de España, Ministerio de Asuntos Económicos y Transformación Digital. https://datos.gob.es/sites/default/files/datosgobes/tecnologias\_emergentes\_y\_opendata\_pln\_awareness.pdf

Nigro, H. (2020). MINERIA DE OPINIÓN, ACCESO ABIERTO Y GOBERNANZA. In Revista Ingeniería, Matemáticas y Ciencias de la Información (Vol. 7, Issue 14, pp. 33–37). Corporacion Universitaria Republicana. https://doi.org/10.21017/rimci.2020.v7.n14.a82

Ekman, P. (1992). An Argument for basic Emotions. Cognition and Emotion. San Francisco: Lawerence Eribaum Associates Limited.

Notas al pie

Tablas

Figuras

Anexos