

Estudio de posts sobre salud mental en Reddit

PCD Proyecto 1

Introducción

Reddit es una red social en la que comunidades de personas con intereses comunes comparten, entre otras cosas, sus experiencias e historias del día a día. La cantidad de personas con problemas de salud mental es cada vez mayor, y esto se hace notar también en esta red social. Existen múltiples comunidades nombradas en función del trastorno mental que sufren sus miembros, como r/depression, sobre la depresión o r/bpd en la que interactúan personas con trastorno límite de la personalidad. Debido a la gran variedad de posts que se pueden encontrar en estas comunidades, decidí hacer un análisis de los distintos posts con el uso de métodos propios del procesamiento de lenguaje natural y el empleo de modelos de aprendizaje automático y profundo.

Abstract

En los últimos años, las plataformas de redes sociales se han convertido en una fuente clave de información y conocimiento sobre condiciones de salud mental. Este estudio presenta un análisis de datos de Reddit para investigar patrones en las discusiones sobre trastornos mentales. Se procesó un conjunto de datos generado por usuarios, aplicando técnicas de procesamiento de lenguaje natural (NLP) para la limpieza de datos, análisis de sentimientos y extracción de características. Se utilizó una combinación de redes neuronales y clasificadores tradicionales de aprendizaje automático para clasificar y evaluar el sentimiento en torno a diversos temas de salud mental. Además, se realizó un análisis de temas (topic analysis) para identificar los tópicos predominantes en las conversaciones, empleando técnicas de modelado de temas y reducción de dimensionalidad. El rendimiento de los modelos se evaluó mediante métricas como la precisión de clasificación, matrices de confusión y visualizaciones de valores de Shapley, proporcionando una visión detallada de la toma de decisiones del modelo. Los resultados sugieren que los modelos desarrollados capturan de manera efectiva el sentimiento e identifican tendencias temáticas en las discusiones sobre salud mental, ofreciendo implicaciones valiosas para el monitoreo y apoyo de la salud mental a través de plataformas digitales.

Keywords: Análisis de sentimiento, Procesamiento de lenguaje natural (NLP), Salud mental, Análisis de temas, Redes sociales, Modelado de temas, Redes neuronales, Aprendizaje automático, Reddit, Visualización de datos.

Métodos aplicados y experimentos

Dataset

El dataset se encuentra en [kaggle](#) con un total de 621002 filas que tienen los siguientes atributos:

- **Title:** El título del post (string).
- **Selftext:** El contenido textual del post (string).
- **Created_utc:** Fecha de creación del post (int, formato utc).
- **Over_18:** Indica si un post es para mayores de 18 o no (contenido NSFW) (bool).
- **Subreddit:** Subreddit al que pertenece el post.

Por tanto se realizará inicialmente un **análisis el contenido del dataset**, seguido de una posible reducción de los datos y un pequeño procesamiento para los siguientes pasos. Seguidamente se utilizará un modelo de lenguaje para obtener los embeddings de los textos completos (título + contenido) con el fin de obtener una representación numérica de los textos para su posterior análisis. A continuación se realizará un **análisis de sentimientos** para comprobar posibles relaciones entre los contenidos de los distintos subreddits y la enfermedad mental que representan y comparar estas comunidades para ver si existe relación entre ellas. Una vez hecho el análisis de sentimiento el siguiente paso será comprobar si hay relaciones entre los contenidos de los posts y si están clasificados como para **mayores de 18**, pues, debería existir una clara relación y, por tanto, se podría entrenar un modelo de deep Learning para su detección. Después se procesarán en mayor medida los datos para aplicar el algoritmo **LDA** para detectar temas recurrentes en los posts como conjunto total y ver si existen conexiones temáticas entre subreddits. Una vez hecho el análisis quedaría entrenar un modelo para **clasificar los posts por subreddit**, pudiendo este ser de ayuda para pacientes que pueden utilizar esta inteligencia artificial para obtener un autodiagnóstico que pueda orientarles y saber qué comunicar a un profesional. Para finalizar, uno de los subreddits, **r/mentalillness** no corresponde a ninguna enfermedad mental como tal, por tanto puede ser interesante realizar algún tipo de clustering para determinar si existen comunidades dentro de este subreddit y qué tipo de contenido tratan, además de utilizar el modelo de detección de subreddits para ver qué enfermedades mentales son prevalentes en cada uno de estos subconjuntos.

Análisis de contenido

Una vez cargado el dataset lo primero que hice fue ver las líneas de tiempo de los distintos subreddits para ver la distribución de posts a lo largo de los años. Subreddits más antiguos como BPD y mentalillness llevan activos desde 2012, mientras que otros como bipolar existen desde 2021, por tanto, con el fin de reducir la cantidad de datos, y que estos sean más similares he tomado los posts publicados desde el momento del primer post del subreddit más “joven”, terminando esta criba con alrededor de unos 400000 datos, casi 300000 menos.

En la descripción del dataset explicaba que había algunos post con la marca [removed], estos no contienen nada por lo que los elimino y borro algún que otro duplicado. Además proceso los datos creando un nuevo campo con la unión de título y contenido en el que solo mantengo caracteres alfanuméricos y elimino los emojis.

Seguidamente para realizar los embeddings utilicé un modelo de huggingface [mxbai-embed-large-v1](#), que genera embeddings en un espacio 512 dimensiones y que detecta emojis, lo cual puede ser útil para posterior uso.

Finalmente para el análisis de sentimiento que realizaré después utilizo el modelo VADER (Valence Aware Dictionary and sEntiment Reasoner) creado por el MIT mediante el uso del módulo VaderSentiment, una herramienta de análisis de sentimientos diseñada para evaluar el tono de textos en inglés, especialmente en redes sociales. Se basa en un léxico que clasifica palabras y frases en función de su polaridad (positiva, negativa o neutral) y la intensidad emocional. Es particularmente efectiva para analizar comentarios cortos y expresiones informales, ya que toma en cuenta aspectos como la puntuación, el uso de mayúsculas y emojis, por tanto es extremadamente útil para este caso. VaderSentiment produce varios outputs al analizar un texto, entre los cuales se encuentran los scores de polaridad, que indican la proporción de sentimientos positivos, negativos y neutrales en el contenido. El score positivo, que varía entre 0 y 1, refleja la fuerza del sentimiento positivo; el score negativo hace lo mismo para el sentimiento negativo; y el score neutral muestra el porcentaje de contenido neutral. Además, genera un Compound Score, que resume la polaridad general en un solo valor que oscila entre -1 (extremadamente negativo) y +1 (extremadamente positivo), con valores cercanos a 0 indicando una opinión neutral.

Dentro del dataset la distribución de subreddits es la siguiente:

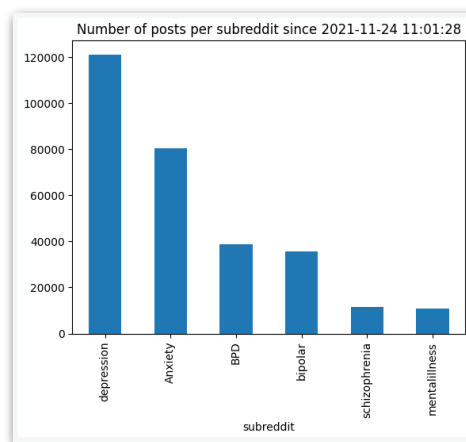
Una vez cargado el dataset lo primero que hice fue ver las líneas de tiempo de los distintos subreddits para ver la distribución de posts a lo largo de los años. Subreddits más antiguos como BPD y mentalillness llevan activos desde 2012, mientras que otros como bipolar existen desde 2021, por tanto, con el fin de reducir la cantidad de datos, y que estos sean más similares he tomado los posts publicados desde el momento del primer post del subreddit más “joven”, terminando esta criba con alrededor de unos 400000 datos, casi 300000 menos.

En la descripción del dataset explicaba que había algunos post con la marca [removed], estos no contienen nada por lo que los elimino y borro algún que otro duplicado. Además proceso los datos creando un nuevo campo con la unión de título y contenido en el que solo mantengo caracteres alfanuméricos y elimino los emojis.

Seguidamente para realizar los embeddings utilicé un modelo de huggingface [mxbai-embed-large-v1](#), que genera embeddings en un espacio 512 dimensiones y que detecta emojis, lo cual puede ser útil para posterior uso.

Finalmente para el análisis de sentimiento que realizaré después utilizo el modelo VADER (Valence Aware Dictionary and sEntiment Reasoner) creado por el MIT mediante el uso del módulo VaderSentiment, una herramienta de análisis de sentimientos diseñada para evaluar el tono de textos en inglés, especialmente en redes sociales. Se basa en un léxico que clasifica palabras y frases en función de su polaridad (positiva, negativa o neutral) y la intensidad emocional. Es particularmente efectiva para analizar comentarios cortos y expresiones informales, ya que toma en cuenta aspectos como la puntuación, el uso de mayúsculas y emojis, por tanto es extremadamente útil para este caso. VaderSentiment produce varios outputs al analizar un texto, entre los cuales se encuentran los scores de polaridad, que indican la proporción de sentimientos positivos, negativos y neutrales en el contenido. El score positivo, que varía entre 0 y 1, refleja la fuerza del sentimiento positivo; el score negativo hace lo mismo para el sentimiento negativo; y el score neutral muestra el porcentaje de

contenido neutral. Además, genera un Compound Score, que resume la polaridad general en un solo valor que oscila entre -1 (extremadamente negativo) y +1 (extremadamente positivo), con valores cercanos a 0 indicando una opinión neutral.

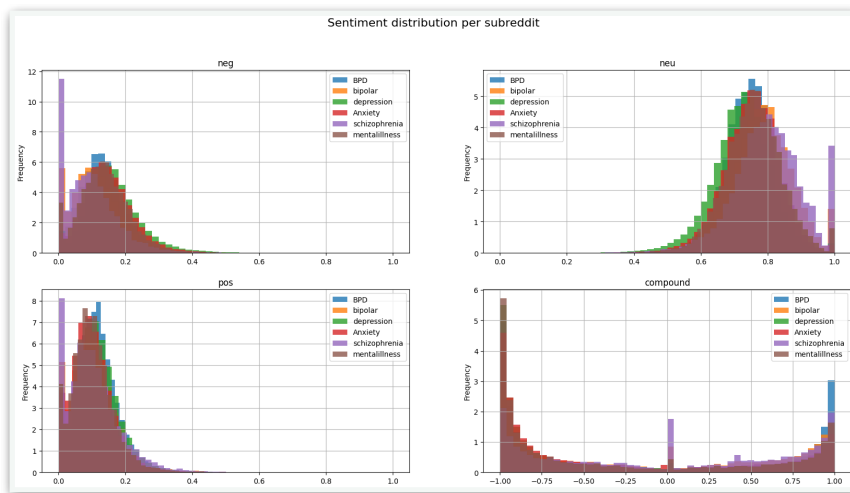


El dataset parece estar extremadamente desbalanceado, reflejando la realidad, pues según las estadísticas, la depresión y la ansiedad son cada vez más predominantes, sobre todo por encima de trastornos como la bipolaridad o el trastorno límite de personalidad. Además parece ser que los usuarios parecen preferir comunicar sus experiencias con las enfermedades mentales en su subreddit dedicado, pues mentalillnes parece ser la comunidad con menos posts.

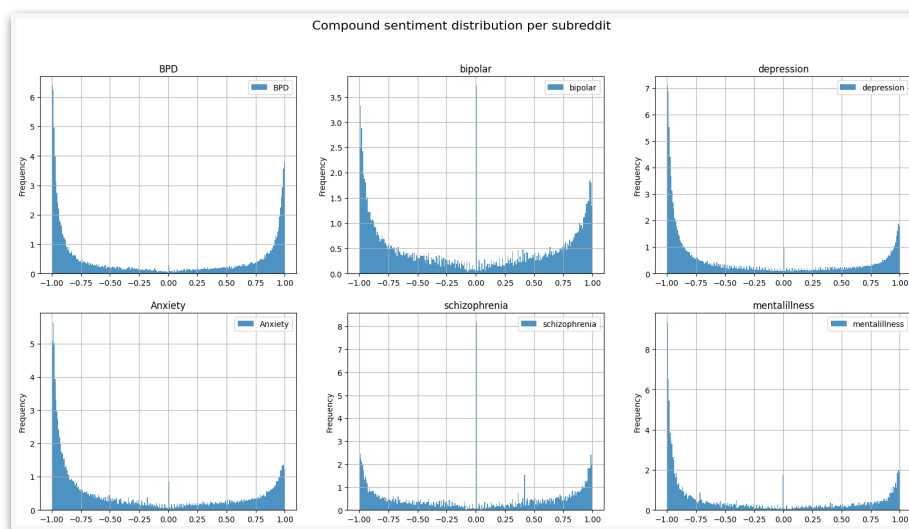
Además, tras un análisis de los distintos wordclouds de los subreddits, y de todos los posts como conjunto, parecen predominar las palabras como feel, think, want, etc. Algo lógico pues estas comunidades están compuestas de personas que comunican sus experiencias diarias, sus sentimientos acerca de su vida con la enfermedad.

Análisis de sentimiento

Utilizando las puntuaciones de sentimientos generadas por VaderSentiment, calculé medias y medianas de los distintos valores por subreddits, encontrando distribuciones extremadamente similares. Los valores de más interés parecen encontrarse en compound, pues nos indican una tendencia negativa en los posts de los distintos subreddits como se puede ver en los diagramas de densidad.



Como se puede ver, la puntuación neutral es por lo general alta, aunque compound muestra una alta polarización en cuanto a los sentimientos, schizophrenia parece ser el más neutral, aunque se puede ver una predominancia negativa.

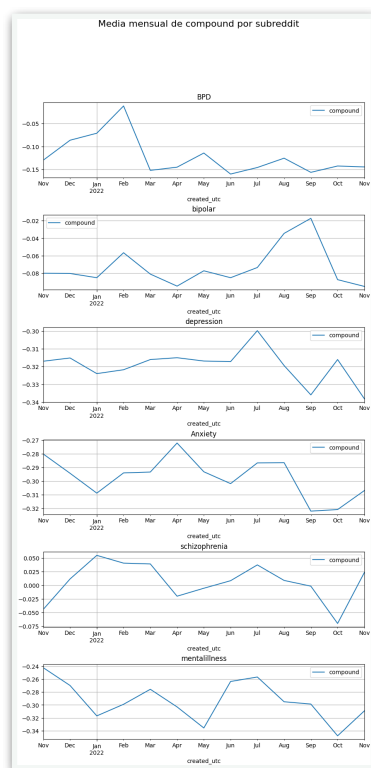


Viendo la distribución de compound en los distintos subreddits, se pueden describir:

- **BPD (Borderline Personality Disorder):** La distribución tiene forma de "U", con muchos comentarios polarizados, tanto en los extremos negativos (-1.00) como positivos (1.00). Los comentarios neutros (cerca de 0) son menos frecuentes.
- **Bipolar:** También muestra una distribución en forma de "U", aunque con un pico significativo alrededor de 0. Esto sugiere que los comentarios en este subreddit son más polarizados, pero con una tendencia leve hacia la neutralidad.
- **Depression:** Este subreddit tiene una fuerte inclinación hacia los sentimientos negativos, con un pico considerable alrededor de -1.00. Aunque hay una pequeña cantidad de comentarios positivos, predominan los comentarios negativos.
- **Anxiety:** Al igual que los otros subreddits, muestra una distribución en forma de "U", con una alta frecuencia de comentarios en los extremos negativos y positivos. Sin embargo, los comentarios cercanos al 0 son más frecuentes que en algunos de los otros subreddits.
- **Schizophrenia:** Tiene una distribución bastante polarizada, con picos significativos tanto en el extremo negativo como positivo. Parece que las emociones expresadas en este subreddit son altamente intensas.

- **Mentalillness:** Este subreddit muestra una fuerte polarización con picos en ambos extremos, pero los comentarios negativos parecen ser más predominantes.

Finalmente, veamos las líneas de tiempo de compound en los distintos subreddits.



- Como se había reflejado en los histogramas, BPD, Bipolar y schizophrenia son los subreddits más polarizados, con una mayor cantidad de comentarios en los extremos negativos y positivos, llevamos a medias más cercanas a 0.
- Depression y mentalillness tienen una tendencia más negativa, con una mayor cantidad de comentarios en el extremo negativo a lo largo del tiempo, con picos de negatividad a final de año en el caso de r/depression, y una periodicidad de 4 meses aproximadamente en r/mentalillness.
- Anxiety tiene picos de negatividad a final y principios de año, y una tendencia positiva en los meses de verano, eso puede ser debido a que la mayoría de los usuarios de este subreddit son estudiantes y sufren más ansiedad en épocas de exámenes.

Análisis de posts +18

Para realizar un análisis de los posts +18 primero comprobé si existía alguna correlación con los sentimientos, aunque los valores resultaron muy cercanos a 0, por lo que no se puede determinar que los sentimientos estén relacionados linealmente con si un post es para mayores de edad o no.

Dado que al principio se obtuvieron los embeddings de los textos, intenté realizar un modelo que clasificara los posts +18, que pudiera hacer de moderador de estos subreddits. Hay que tener en cuenta que la proporción entre posts +18 y que no lo son es extremadamente desbalanceada, por tanto a la hora de evaluar la función de loss del modelo, utilicé pesos mayores para la clase

minoritaria (`over_18 == True`) pues aunque el accuracy fuera bueno, esto no implica una buena clasificación. Por tanto mi objetivo fue maximizar el recall siguiendo la lógica de que es mejor clasificar un post +18 cuando no lo es que no clasificar incorrectamente uno que sí lo sea.

Una vez entrenado el modelo, los resultados no fueron demasiado favorables. Emplear data augmentation a embeddings realizados por un modelo de lenguaje es una tarea peligrosa puesto que los nuevos datos no tienen por qué ser fieles a la realidad y consecuentemente solo se pudo depender de los pesos de clases.

Estos malos resultados también se pueden justificar con la subjetividad de los moderadores a la hora de clasificar un post, pero, por la duda, entrené el modelo en un único subreddit, el mayoritario `r/depression`, y, dado que los resultados son igual de malos concluyo que se necesitaría más información para poder clasificar correctamente los posts. Además existen miles de anécdotas de moderadores de reddit que han utilizado sus opiniones personales para moderar, lo que puede llevar a una clasificación incorrecta de los posts.

Análisis de Topics

Una vez hecho el análisis anterior, puede arrojar buenos resultados el uso de LDA para detectar las diferentes temáticas presentes en los posts.

LDA, o Latent Dirichlet Allocation, es un modelo de análisis de temas que se utiliza en el procesamiento de lenguaje natural y en la minería de textos. Su principal objetivo es identificar y extraer temas ocultos en un conjunto de documentos. LDA asume que cada documento se puede representar como una mezcla de varios temas, y cada tema se compone de palabras con cierta probabilidad.

En esencia, LDA funciona asignando cada palabra de un documento a un tema específico, lo que permite descubrir patrones de coocurrencia en el texto. Al final del proceso, proporciona una lista de temas y las palabras más representativas de cada uno, así como la distribución de esos temas en los diferentes documentos. Esto es útil para tareas como la categorización de documentos, la recomendación de contenido y el resumen automático, ya que ayuda a comprender mejor el contenido de grandes volúmenes de texto, como es el caso.

Para preparar los posts para el modelo, tokenicé y lematicé los textos con el fin de facilitar el procesamiento de LDA para encontrar los topics y una vez entrenado el modelo, lo evalué con el CoherenceModel ofrecido por la librería de nlp gensim. El CoherenceModel es una herramienta que se utiliza para medir la coherencia de los temas generados por modelos de tópicos, como LDA (Latent Dirichlet Allocation). La coherencia se refiere a qué tan bien se relacionan las palabras dentro de un tema y cuán intuitivo es el tema para los humanos.

El modelo calcula diferentes métricas de coherencia, como la Coherencia UMass y la Coherencia C_V, que ayudan a evaluar la calidad de los tópicos. Estas métricas se basan en la coocurrencia de palabras en documentos y en cómo esas palabras se agrupan dentro de los temas. Un puntaje de coherencia más alto indica que las palabras en un tema son más relacionadas entre sí, lo que sugiere que el tema es más claro y comprensible.

Según métricas de estudios, un valor entre 0.4 y 0.65 es aceptable, un valor mayor que el límite indica datos falsos y menos indica un modelo mal parametrizado. En mi caso obtuve un valor de 0.456, que no he logrado mejorar después de numerosos experimentos paramétricos. Es un valor aceptable, y por tanto se obtuvieron los siguientes temas.

El **primer tema** se refiere a una narrativa interna y autopercepciones, común en personas que luchan con su identidad o control de sus pensamientos, posiblemente asociado a la depresión, ansiedad o síntomas de confusión. Palabras como “not”, “can”, “scared” y “addicted” sugieren sensaciones de impotencia y miedo.

El **segundo tema** gira en torno al trauma infantil y familiar. Términos como “trauma”, “abuse”, “mother”, “father” y “childhood” indican una discusión sobre el abuso o experiencias traumáticas en la infancia que afectan la salud mental a lo largo del tiempo.

El **tercer tema** se centra en relaciones sociales, tanto románticas como amistades. Palabras como “friend”, “talk”, “love” y “relationship” destacan el impacto de las relaciones en la salud mental, lo que puede relacionarse con trastornos de la personalidad, ansiedad social o depresión.

El **cuarto tema** está enfocado en el tratamiento médico. Términos como “therapist”, “medication”, “doctor” y “anxiety” sugieren una discusión sobre la búsqueda de ayuda profesional, medicamentos y la gestión de trastornos como la ansiedad, el pánico o el trastorno obsesivo-compulsivo.

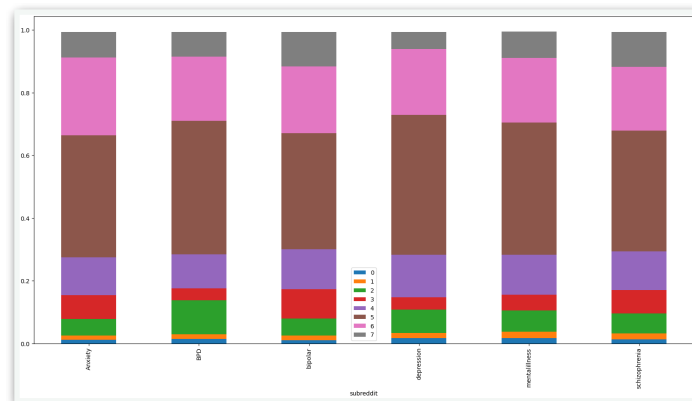
El quinto tema habla sobre la vida diaria y el impacto de la enfermedad mental en áreas como el trabajo, la escuela o la familia. Palabras como “live”, “mental”, “work” e “illness” sugieren preocupaciones sobre cómo lidiar con la enfermedad mental en el entorno cotidiano.

El **sexto tema** refleja emociones generales y la lucha interna. Términos como “feel”, “like”, “think” y “want” indican que las personas hablan sobre sus sentimientos, pensamientos y dificultades para encontrar significado o motivación, lo cual es típico en quienes luchan con depresión o ansiedad.

El **séptimo tema** describe rutinas diarias y acciones, posiblemente relacionadas con cómo la salud mental afecta la vida cotidiana. Palabras como “go”, “get”, “day” y “sleep” señalan problemas con la energía o el sueño, comunes en trastornos como la depresión o el trastorno bipolar.

Por último, el **octavo tema** agrupa diagnósticos y síntomas específicos. Palabras como “anxiety”, “depression”, “disorder” y “bipolar” se centran en la discusión de diagnósticos, síntomas y trastornos específicos, como la ansiedad, depresión, bipolaridad y trastorno límite de la personalidad, con un enfoque en la experiencia personal y las preguntas sobre tratamiento.

Una vez obtenidos los topics, vemos la distribución por subreddit.



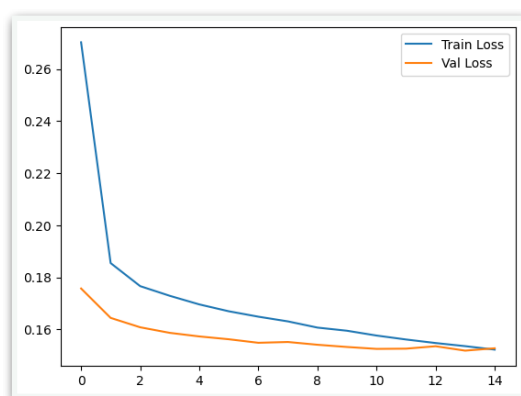
Los datos no arrojan conclusiones respecto a topics que puedan variar en función del foro, siguen siendo los mismos temas generales para todos los subreddits. Lo que sí se puede apreciar es una predominancia del topic 5, seguido del 6 y luego siendo similar para 7 y 4. Esto puede indicar que la mayoría de los posts (como es lógico) son de personas que describen su día a día con su enfermedad y la lucha interna que sufren.

Debido a la predominancia del topic 5, decidí ver el wordcloud de los posts dentro de este tema, el caso es que, debido a que es el más predominante, el wordcloud sigue una distribución muy similar al realizado al principio con todos los posts.

Modelo clasificador de subreddits

Dada la dificultad de clasificar los posts en función de si son +18 o no, probé si se podían clasificar los posts en función de si pertenecen a un subreddit u otro, es decir, si el modelo es capaz de detectar las diferencias en los textos de los distintos subreddits, y si es capaz de clasificar correctamente los posts. Además para esta clasificación eliminé el subreddit general, ya que puede dificultar el entrenamiento del modelo. Finalmente con el modelo entrenado y evaluado, puede ser interesante ver si es capaz de clasificar los posts de r/mentalillness y ver si existen incongruencias en la clasificación. Debido a la poca influencia que tienen los topics o los sentimientos por sus valores similares y distribuciones, no los incluí en el modelo, pues no aportan información relevante.

El modelo recibe de input el embedding de 512 dimensiones del post y tras varias capas ocultas fully-connected da un output categórico que representa el subreddit. La función de pérdidas fue categorical crossentropy debido a que es una clasificación de multiclase.



El modelo arrojó resultados bastantes aceptables, con un accuracy del 85% y un recall, precisión y f1-score cercanos al 0.8-0.85. Dado el desbalanceo de los datos y la interconexión de las comunidades, los resultados indican que ciertamente hay relaciones ocultas en el espacio de los embeddings que determinan la pertenencia a un subreddit u a otro.

Análisis de r/mentalillness

Para finalizar el análisis, me centré en analizar el subreddit general de enfermedades mentales, r/mentalillness, el proceso fue el siguiente:

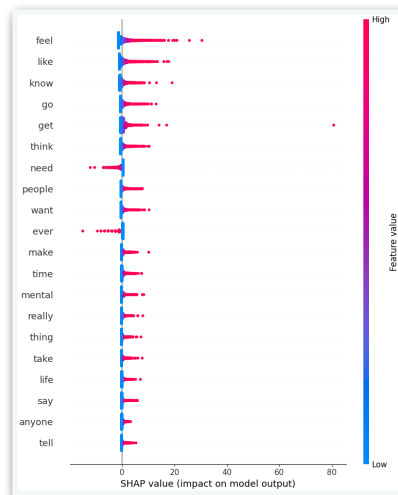
- Reducción de dimensionalidad mediante UMAP, para mantener la estructura de los datos y poder visualizarlos.
- Clustering mediante HDSCAN, para ver si se pueden encontrar grupos de posts en función de sus embeddings.
- Clasificación de pertenencia a un cluster con un modelo simple como un árbol de decisión o una regresión logística.
- Estudio de los shap values para ver las palabra que definen cada cluster, viendo si se puede detectar comunidades de determinados temas.
- Comparación con las predicciones del modelo de clasificación anterior.

Según casos en los que he trabajado, el mejor valor de UMAP para textos suele ser entre 20 y 30 dimensiones, después HDBSCAN, un algoritmo de agrupación por densidad en el espacio, a contrario que K-means que se basa en centroides, se optimizará con optuna[2], una librería de experimentos paramétricos, para obtener el máximo silhouette score, que mide qué tan bien están separados los clústeres y qué tan compactos son dentro de un análisis de agrupamiento, con valores más altos indicando mejor calidad de agrupación. El silhouette score va de -1 a 1, siendo 0 mal clustering y valores mayores a 0.5 como idóneos.

Tras experimentar obtuve un valor de 0.75, extremadamente bueno, obteniendo como resultado 2 clústers, uno mayoritario con 10610 posts (cluster 1) y otro con 91 (cluster 0). HDBSCAN no encontró ruido por lo que no hizo falta eliminar ningún dato.

Seguidamente se usó una regresión logística que clasificó los agrupamientos con un accuracy del 99% en base a una vectorización one-hot con recuento de palabras de los documentos.

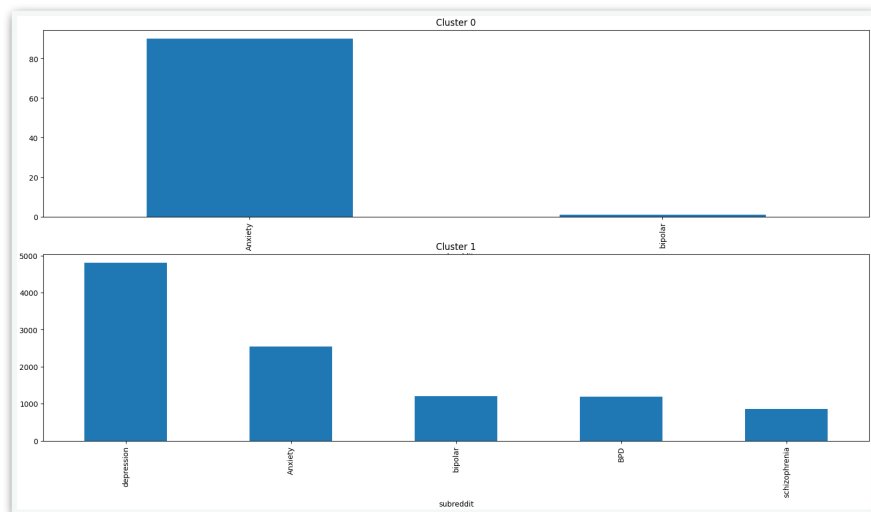
Viendo los wordcloud de ambos agrupamientos se pudo ver que el cluster minoritario 0 habla sobre special, need, autism, lo que puede indicar que es un cluster de posts de personas neurodivergentes, mientras que el cluster 1, el mayoritario, habla sobre pensar, sentir, querer, lo que puede indicar que es un cluster de posts de personas que hablan sobre sus sentimientos y pensamientos, no necesariamnete etiquetando un tema en concreto.



Finalmente utilizando `shap[1]`, librería para explicabilidad de modelos de aprendizaje automático tales como la regresión logística, determinamos qué palabras tienen influencia en cuanto a qué post pertenece a un cluster o a otro:

Como se podía ver en el wordcloud, las palabras más importantes para la clasificación son las más frecuentes de cada cluster, en este caso, las palabras más comunes de un cluster influyen positivamente a la pertenencia a ese cluster, y negativamente al otro.

Finalmente aplicando el modelo de subreddits a los posts obtenemos la siguiente distribución de subreddits.



Como se puede apreciar en la distribución de subreddits por cluster, el cluster mayoritario parece seguir una distribución similar a la de los subreddits, algo claramente influenciado por el entrenamiento del modelo, pero que tiene cierto sentido pues los subreddits de este tipo de temas suelen tener posts similares.

Finalmente en el cluster 0 minoritario, con palabras clave como autism, special o need, parece ser predominante el tema de la ansiedad, que puede estar ligada a las vivencias diarias de personas neurodivergentes.

Conclusión

Una vez terminado el análisis, podemos concluir los siguientes puntos:

- El sentimiento de los posts en los subreddits de enfermedades mentales es muy similar, con una distribución en forma de "U" que indica una alta polarización en los comentarios, tanto en los extremos negativos como positivos. Aunque por lo general, los posts tienden a ser más negativos que positivos, esto puede deberse a la naturaleza de los temas tratados en estos subreddits, y que los subreddits más populares tratan sobre la depresión y la ansiedad, que son enfermedades mentales que suelen asociarse con sentimientos negativos.
- La clasificación de los posts en función de si son +18 o no es una tarea complicada, sino imposible, ya que no existe una relación clara entre los sentimientos expresados en los textos y su contenido y la clasificación realizada por los moderadores de los subreddits, que puede ser, y seguramente es, subjetiva.
- Los subreddits de enfermedades mentales son muy similares en cuanto a los temas que tratan, siendo los más comunes los relacionados con las vivencias del día a día de las personas que sufren estas enfermedades, sus pensamientos y sentimientos, y la búsqueda de ayuda profesional y tratamiento y, minoritariamente, traumas de la infancia e historias de abuso. Aunque no se han encontrado diferencias claras en los temas tratados en los distintos subreddits, se ha podido identificar un cluster de posts de personas aparentemente neurodivergentes en r/mentalillness (el subreddit general del tema), que hablan sobre sus necesidades especiales y su experiencia de vida.
- A pesar de las similitudes en los temas tratados en los subreddits, el modelo de clasificación de subreddits ha sido capaz de clasificar correctamente los posts de los distintos subreddits con un accuracy aceptable, lo que indica que existen diferencias en los textos de los distintos subreddits a pesar de sus similitudes en cuanto a los temas tratados, concluyendo una distribución similar a la general en r/mentalillness en el cluster mayoritario.

Referencias

[1] *A Unified Approach to Interpreting Model Predictions*, Scott Lundberg and Su-In Lee, 2017

[2] *Optuna: A Next-generation Hyperparameter Optimization Framework*, Takuya Akiba and Shotaro Sano and Toshihiko Yanase and Takeru Ohta and Masanori Koyama, 2019