Development of a Natural Language Processing-Based System for Characterizing Eating Disorders

Miguel Rujas[1], Beatriz Merino-Barbancho [1], Peña Arroyo[1] and Giuseppe Fico [1]

[1] *Life Supporting Technologies, Universidad Politécnica de Madrid, Madrid, Spain*

Abstract
Eating disorders represent a significant global health concern, characterized by high prevalence and profound repercussions for affected individuals. However, the timely identification of these disorders is impeded by the current reliance on time-consuming questionnaires as diagnostic tools. This article aims to outline the development of a Natural Language Processing (NLP)-based system designed to address this challenge. The methodology employed comprises three primary phases: a text preprocessing pipeline, topic extraction using BERTopic and the creation of binary classification and linear regression models based on fine-tuned BERT. These models enable the prediction of message origin from individuals with eating disorders and the estimation of the probability thereof. This comprehensive system facilitates the characterization and early detection of eating disorders within text messages.
Keywords
Natural language processing, Eating disorders, Transformers, BERT, Text messages.

1. Introduction

Eating disorders, such as anorexia nervosa, bulimia nervosa and binge eating disorder, are serious mental illnesses that affect people all over the world. These disorders are characterized by abnormal eating behaviours, distorted body image and psychological distress (Haddadi & Ali Besharat, 2010). Early detection and characterization of eating disorders are crucial for effective intervention and treatment (Dingemans et al., 2002). In recent years, Natural Language Processing (NLP) has emerged as a powerful tool for analyzing textual data in various domains, including mental health (Arowosegbe & Oyelade, 2023). NLP techniques allow the automatic extraction of meaningful information from large amounts of text, providing information about people's mental states, emotions and behaviours. By leveraging NLP, researchers can develop systems that help identify and characterize mental illnesses, including eating disorders (Nadkarni et al., 2011).

Several studies have explored the application of NLP techniques in mental health analysis, including the characterization of eating disorders. Cliffe and colleagues (Cliffe et al., 2021) developed a system using NLP to identify linguistic markers associated with risk factors associated with suicide extracted from electronic health record (EHR) data. Their work demonstrated the potential of NLP to automatically identify individuals at risk for suicide-related behaviour. Another relevant study by Feldman (2023) (Feldman, 2023) focused on the development of a deep learning-based system to analyze online photos from Twitter and extract information about individuals' body image concerns and disordered eating patterns. In addition, other lines of research employed thematic modeling techniques to classify eating disorder-related diagnoses using EHR data (Bellows et al., 2014). While these studies highlight the potential of NLP for understanding eating disorders, there is still a need for more specialized systems that focus specifically on characterizing eating disorders and draw on other data sources such as text messages in instant messaging applications or forums. Therefore, our work aims to expand that area by developing a comprehensive NLP-based system that captures various facets of eating disorders, including eating behaviours, attitudes and psychological well-being.

IberLEF 2023, September 2023, Jaén, Spain
EMAIL: mrujas@lst.tfo.upm.es (A. 1); bmerino@lst.tfo.upm.es (A. 2); pena.arroyo@lst.tfo.upm.es (A. 3) gfico@lst.tfo.upm.es (A. 4)
ORCID: 0009-0005-8198-3356 (A. 1); 0000-0001-5070-4178 (A. 2); 0000-0002-4292-9841 (A. 3) 0000-0003-1551-4613 (A.4)

The aim of this article is to describe the development of an NLP-based system specifically designed to characterize eating disorders. Employing advanced text analysis techniques, our system aims to extract relevant features and patterns from textual data coming from Telegram related to eating behaviours, attitudes towards body image and psychological well-being.

2. Materials and Methods

2.1. MentalRiskES 2023

This study is framed within the scope of task 1 of MentalRiskES2023 (Mármol-Romero et al., 2023), a shared task in the evaluation campaign of Natura Language Processing systems known as IberLEF. The primary aim of this task was to construct a binary classification system and a linear regression system that could effectively determine the presence or absence of eating disorders in textual messages.

2.2. Database description

The provided database comprises a total of 6320 instances and encompasses five distinct variables as depicted in Table 1.

Table 1
Database variables

Name	Type	Description
id	int64	Identifier of the user who sent the message
Message	object	Text message sent by the user
Date	object	The date on which the message was sent.
task1a	int64	A label indicating the absence (0) or presence (1) of an eating disorder in the user sending the message.
task1b	float64	A label indicating the probability that the user has an eating disorder, with 0 being the lowest value and 1 the highest value.

The distribution of the labels ('*task1a'* and '*task1b*') is depicted in Figure 1, which illustrates that a higher percentage of instances correspond to individuals who do not have the disorder or possess a very low probability of having the disorder. This indicates an inherent imbalance within the database.

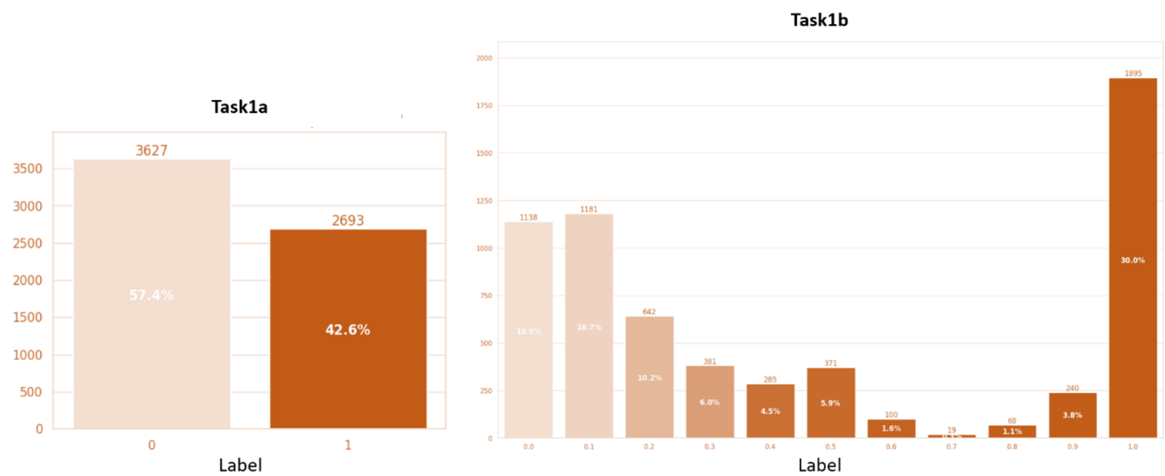

Figure 1: Distribution of label values

2.3. Development environment

The development activities were conducted in a Jupyter Notebook hosted on a server with specific hardware resources, as described in Table 2.

Table 2
Development environment resources

Type of Resource	Resource Model	Resource Quantity
CPU	Intel(R) Core(TM) i9-9820X CPU @ 3.30GHz	8
GPU	TITAN RTX Graphics Card	1
RAM Memory		32 GB

2.4. BERT

BERT (Bidirectional Encoder Representations for Transformers) (Devlin et al., 2018) is a transformer-based model specifically developed for natural language processing (NLP) tasks. Its primary capability lies in generating deep bidirectional representations from unlabeled text, enabling it to grasp the syntax and semantic meaning of words effectively.

The significance of BERT stems from its ability to comprehend contextual information bi-directionally. This unique characteristic allows the model to achieve state-of-the-art performance across various NLP tasks, including sentiment analysis and entity recognition. By fine-tuning BERT on labeled datasets, it becomes possible to obtain highly performant models for these specific tasks.

Figure 2 provides a visual representation of the methodology of the study, which has been described in detail in the following subsections.

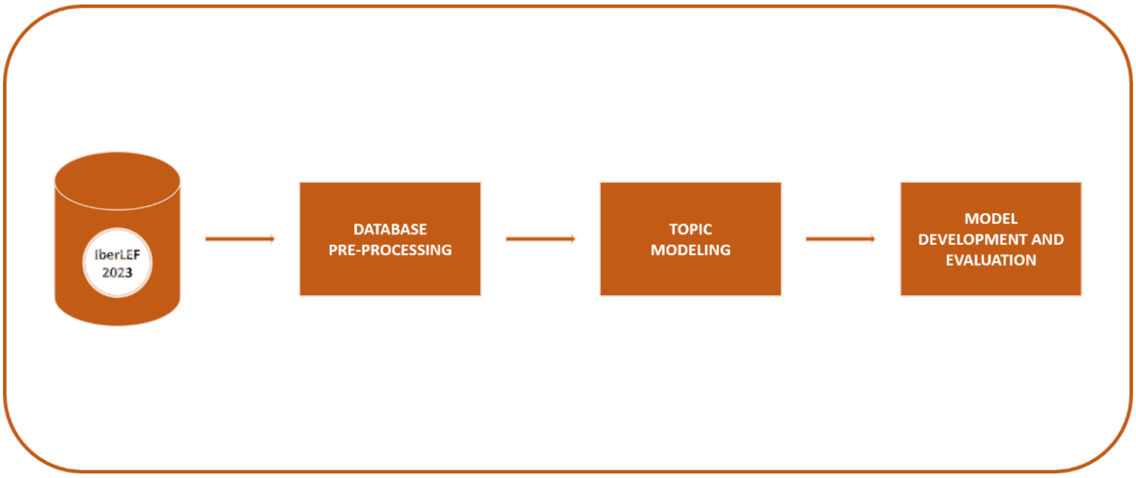

Figure 2: Methodology followed

2.5. Database Pre-processing

The pre-processing of the database involved the utilization of a customized natural language processing pipeline specifically tailored for this particular case. The pipeline was designed to enhance the quality and suitability of the data for subsequent analysis and modeling. It consisted of six distinct steps, outlined in the following sections.

2.5.1. Pre-processing of emoticons

The pre-processing of emoticons within the text messages of the database involved addressing two distinct scenarios: emoticons presented graphically and those described in Spanish text. In response to this situation, a method was developed to standardize the representation of emoticons and leverage the valuable information they can provide, as demonstrated in previous use cases (C. Liu et al., 2021). This method comprised two steps:

1. **Transformation of Emoticons to Text:** Emoticons represented graphically were converted to their standardized text representations in Spanish using the "*Emoji*" library (*Emoji · PyPI*, n.d.). For example, a face-tasting food emoticon would be transformed into "*cara_saboreando_comida*".
2. **Standardization of Textual Representations:** Since the standardized representation differed from the emoticons found in the database messages, it was decided to adapt the latter. To achieve this, a dictionary was created with the standardized representations as keys and the corresponding representations found in the database, without underscores, as values. The text messages were then searched for occurrences of the dictionary values, and each time a match was found, the respective pattern was replaced with its standardized representation (dictionary key). This replacement process commenced with the longest values to account for potential variations in emoticon names, such as different colors. Starting with longer patterns ensured the full representation of the emoticon was identified, thereby avoiding any loss of information. For instance, the thumbs-up emoticon could have various variations, each with different standardized representations such as "*pulgar_hacia_arriba*" or "*pulgar_hacia_arriba_tono_de_piel_claro_medio*". By prioritizing longer patterns, the complete representation of the emoticon could be accurately identified, preserving all relevant information.

By implementing these steps, the emoticons present in the text messages were standardized, enabling consistent handling and preserving their informative value for subsequent analysis within the pipeline.

2.5.2. Standardization of Expressions

Given that the database contains messages from an instant messaging service, a significant number of colloquial expressions and abbreviations with similar meanings were observed. To facilitate their interpretation by various models, these expressions were standardized by categorizing them into five distinct groups, as described below:

- **Expressions of Laughter:** This category encompasses expressions that indicate laughter. To process these expressions, words containing more than one '*j*' were identified and replaced with the word '*risa'*. Additionally, words with multiple '*j*'s were manually reviewed to ensure their classification as laughter expressions.
- **Uncertainty Expressions:** This category includes expressions that convey uncertainty. Words containing consecutive '*m*'s were identified and replaced with the word '*incertidumbre'*. Furthermore, words with multiple consecutive '*m*'s were manually verified to confirm their status as uncertainty expressions.
- **Sighs:** This category comprises expressions representing sighs. Words containing the letter '*f*' were detected, and those identified as sigh expressions were manually replaced with the word '*suspiro'*.
- **Expressions of Admiration:** This category encompasses expressions indicating admiration. Words containing the letter '*w*' were identified, and expressions belonging to this category were manually identified and substituted with the word '*admiración*'.
- **Abbreviations:** This category includes commonly used abbreviations in instant messaging. To process these abbreviations, a dictionary was created by gathering the most prevalent abbreviations used in various instant messaging applications. Each abbreviation was assigned its corresponding meaning. Subsequently, the text messages were scanned for the presence of

these abbreviations, and when detected, they were replaced with their corresponding expanded form using the dictionary.

By implementing these standardization techniques, expressions within the text messages were unified into specific categories, enhancing the interpretability and consistency of the data for subsequent modeling and analysis.

2.5.3. Expression Filtering

To ensure the cleanliness and clarity of the text messages, a filtering process was applied to remove words that did not qualify as alphanumeric or emoticons. This involved the elimination of punctuation marks and expressions containing special symbols, with the exception of emoticons. Specifically:

- **Punctuation Marks:** All punctuation marks, including periods, commas, exclamation marks, question marks, and other non-alphanumeric characters, were excluded from the text messages.
- **Special Symbols:** Expressions containing any form of a special symbol, such as the underscore ("_"), were filtered out, except for emoticons. Emoticons were preserved as they hold specific meanings in the context of the messages.

By implementing this expression filtering approach, words that did not contribute to the analysis and did not meet the criteria of being alphanumeric or emoticons were removed from the text messages. This process aimed to refine the data and focus on the essential content, facilitating subsequent analysis and modeling tasks.

2.5.4. Word Correction

During the analysis of the text messages, it was identified that several users made minor spelling errors, such as inadvertently typing a neighboring letter on the keyboard or using an English word instead of its Spanish equivalent. To rectify these errors, each word was individually processed through the following steps:

1. **Spelling Correction:** The "*Pyspellchecker*" library (*Pyspellchecker · PyPI*, n.d.) was employed to check if each word could be corrected to its intended spelling. If a word was found to have a potential correction, it was replaced with the correct spelling.
2. **Translation to Spanish:** After the spelling correction, an attempt was made to translate the word into Spanish using the "*Translate*" library. (*Translate · PyPI*, n.d.) This step aimed to ensure consistency in the language used throughout the text messages.

By implementing these word correction techniques, spelling errors within the text messages were addressed, and words were standardized to improve the accuracy and coherence of the data. The combination of spelling correction and translation to Spanish enhanced the overall quality and consistency of the text for subsequent analysis and modeling processes.

2.5.5. Lemmatization

Lemmatization is a crucial process that involves reducing words to their base form or lemma, aiming to normalize words and treat all their variations as a unified entity. This technique holds significant importance in the analysis of textual data, as demonstrated in various studies within the field of natural language processing (Balakrishnan & Ethel, 2014). Lemmatization is often compared to another technique called stemming, which reduces words to their root form by removing suffixes.

In this particular case, the lemmatization process was implemented using the "*Spacy*" library (*Spacy · PyPI*, n.d.). This library provides robust capabilities for identifying and transforming words into their respective base forms, allowing for better consistency and comparability in subsequent analyses. By lemmatizing the words within the text messages, variations of the same word were unified, enabling a more comprehensive understanding and analysis of the text data.

It is worth noting that lemmatization differs from stemming in that it considers the morphological structure and context of words, resulting in more accurate and linguistically meaningful lemmas. Stemming, on the other hand, relies on simple suffix removal and may produce less precise results in terms of maintaining word meaning.

Overall, the lemmatization process performed with the "*Spacy*" library played a crucial role in normalizing the words within the text messages, facilitating more effective analysis and interpretation of the data.

2.5.6. Elimination of Stopwords

Stopwords are a collection of commonly occurring words in texts that typically carry little semantic meaning. In many natural language processing studies, it is customary to exclude these words from the analysis to eliminate noise and focus on more informative content (Blanchard, 2007; Fox, 1989). The list of stopwords utilized in this study was obtained from the "*NLTK*" library (*Nltk · PyPI*, n.d.).

By removing stopwords, the goal is to filter out words that frequently appear across the text messages but contribute little to the overall understanding of the content. These stopwords often include pronouns, conjunctions, prepositions, and other common functional words. Eliminating these words can streamline the analysis process and enhance the extraction of meaningful information.

The provided list of stopwords, obtained from the "NLTK" library, ensures comprehensive coverage of common stopwords in the Spanish language. By discarding these words, the resulting dataset is refined, focusing on more significant and contextually meaningful terms. This step contributes to improving the accuracy and efficiency of subsequent analysis and modeling tasks by reducing noise and irrelevant information.

2.6. Topic Modeling

To gain further insights into the characteristics of different users in the Telegram group and extract information about the topics in which individuals with eating disorders actively engage, the technique of topic modeling was employed. Topic modeling is a natural language processing approach that aims to identify latent variables within large text datasets (Blei, 2012). While it is commonly utilized in various domains such as biomedical research (L. Liu et al., 2016), environmental studies (Girdhar et al., 2012), and, in this case, social network analysis (Liangjie & Davison, 2010), its primary objective is to cluster text data based on thematic similarity.

Several methods exist for conducting topic modeling on text data, including "Latent Dirichlet Allocation" (LDA), "Top2Vec," and "BERTopic." Among these options, BERTopic was selected for this study due to its notable stability, versatility across different domains, and its current state-of-the-art performance (Egger & Yu, 2022).

By applying the BERTopic method, the study aimed to uncover underlying topics within the text data, allowing for the clustering and categorization of discussions related to eating disorders. This analysis provides valuable insights into the predominant themes and discussions in which individuals with eating disorders are actively participating. The state-of-the-art nature of BERTopic ensures optimal performance and accurate representation of the topics present in the dataset.

Utilizing topic modeling techniques, specifically, BERTopic, contributes to a more comprehensive understanding of the textual content, enabling researchers to identify key topics and trends within the Telegram group.

2.6.1. BERTopic

BERTopic (Grootendorst, 2022) is a topic modeling technique that operates through three distinct stages, which are elaborated upon in the following points:

1. **Transforming documents into vector representations:** In the initial stage, the textual documents are transformed into numerical vector representations, facilitating their processing by various algorithms. For this purpose, Sentence-BERT (SBERT) (Reimers & Gurevych, 2019), a modified version of the BERT model discussed earlier in section **2.4 BERT**, was chosen. This selection is based on the recommendation by the author Grootendorst (Grootendorst, 2022), highlighting SBERT's capability to generate vector representations with high semantic value.
2. **Clustering the vector representations into themes:** In the second stage, the high-dimensional vector representations produced by the SBERT model undergo dimensionality reduction to enhance the clustering results. This process aims to reduce computational complexity and eliminate potential noise. For dimensionality reduction, the technique called UMAP (Uniform Manifold Approximation and Projection) (Mcinnes et al., 2020) is employed. UMAP preserves both local and global relationships in the new lower-dimensional vector space without imposing any constraints on the size of the vector representations.

   Following dimensionality reduction, a clustering process is conducted using the lower dimensional vector representations to group texts with similar characteristics. This clustering process provides insights into the underlying structure of the dataset by identifying texts that share thematic similarities. The clustering algorithm employed in this study is HDBSCAN (McInnes et al., 2017), known for its ability to detect complex relationships, prevent the grouping of unrelated documents, and automatically determine the optimal number of clusters or topics.
3. **Topic representation:** In the final stage of the process, the word distributions within the identified clusters or topics are modeled to determine the weight of each word. This is achieved through a modification of the TF-IDF (Term Frequency-Inverse Document Frequency) technique, known as c-TF-IDF (cluster-TF-IDF). c-TF-IDF enables the calculation of word importance per cluster rather than per individual document.

   Lastly, a step to reduce outliers, i.e., documents that have not been assigned to a topic, is performed. This situation may arise due to the minimum similarity requirement set by the HDBSCAN clustering algorithm. To address this issue, a probability-based criterion is applied, wherein each document is assigned to the topic it is most likely to belong to, mitigating the occurrence of unassigned documents.

2.7. Model Development and Evaluation

In the study, two types of models were developed: binary classification models and linear regression models. Although there are differences in developing these models, they followed similar approaches. The general process followed, along with the specific differences for each type of model, is described below.

2.7.1. Model

Both binary classification models and linear regression models used a modified version of BERT called BETO(Cañete, 2020), which was pre-trained on a large corpus of unlabeled Spanish data. The model was adapted for each task as follows:

- **Binary Classification Model:** The "*num_labels*" field was modified to indicate the number of classes to predict, which was set to 2 (presence or absence of the disease). The model was fine-tuned using 80% of the database as the training set and the remaining 20% as the test set. Hyperparameter tuning was also performed, including batch size (16), learning rate (5e-5, 3e-5, and 2e-5), and the number of epochs (2, 3, and 4).
- **Linear Regression Model:** The "*num_labels*" field was set to 1 since only one value between 0 and 1 was predicted. The model was also fine-tuned using 80% of the database for training, and hyperparameters such as batch size, learning rate, and number of epochs were adjusted.

2.7.2. Scenarios

The models were developed for three different scenarios, each with different input data:

- **Scenario 1:** The model input consisted of pre-processed text messages.
- **Scenario 2:** The model input included pre-processed text messages concatenated with the date of the message and the topic.
- **Scenario 3:** The model input included pre-processed text messages concatenated with the time of the message and the topic.

Temporal data was added in scenarios 2 and 3 based on previous studies indicating certain patterns of activity for individuals with mental disorders on social networks(Leis et al., 2019)

2.7.3. Evaluation Metrics for Binary Classification Task

The performance of the binary classification models was evaluated using the following metrics, considering the imbalance of the dataset:

- **Area Under the ROC Curve (AUROC):** Measures the discriminative ability of the model to distinguish between classes.
- **Weighted Accuracy:** Measures the proportion of correct predictions, considering the class distribution.
- **Weighted Recall:** Measures the proportion of correctly identified individuals with eating disorders, considering the class distribution.
- **Weighted F1 Score:** Harmonic mean of precision and recall, providing an average performance metric considering the class distribution.

2.7.4. Evaluation Metrics for Linear Regression Task

The performance and quality of fit of the linear regression models were evaluated using the following metrics:

- **Mean Square Error (MSE):** Measures the average squared errors between predicted and actual values, assessing the fit of the model to the data.
- **Mean Absolute Error (MAE):** Measures the average absolute errors between predicted and actual values, assessing the fit of the model to the data.
- **Coefficient of Determination (R^2):** Measures the fit of the model to the variability exhibited by the data.

These metrics were used to assess the performance and quality of the models in their respective tasks, providing insights into their effectiveness in detecting eating disorders and predicting numerical values related to the disorder.

3. Results

3.1. Binary Classification

Table 3 presents the outcomes obtained from the models exhibiting the most favorable evaluation metrics for the binary classification task within each scenario.

Table 3
Evaluation metrics results for Binary Classification

Scenario	Learning rate	Epochs	AUROC	Weighted Accuracy	Weighted Recall	Weighted F1 Score
1	3e-5	4	0.7918	0.72	0.72	0.71
2	3e-5	4	0.9860	0.95	0.95	0.95
3	5e-5	2	0.8338	0.78	0.78	0.78

The typical emissions generated by these models during prediction operations amount to 1.51 e-5 kgCO2e, while the average energy consumption amounts to 7.94e-5 J.

3.2. Linear Regression

Table 4 presents the outcomes obtained from the models exhibiting the most favorable evaluation metrics for the linear regression task within each scenario.

Table 4
Evaluation metrics results for Linear Regression

Scenario	Learning rate	Epochs	MSE	MAE	R^2
1	5e-5	2	0.1228	0.2749	0.2763
2	3e-5	3	0.0249	0.0996	0.8532
3	5e-5	2	0.1030	0.2411	0.3934

The typical emissions generated by these models during prediction operations amount to 1.51 e-5 kgCO2e, while the average energy consumption amounts to 7.94e-5 J.

4. Discussion

4.1. Binary Classification

Upon examining the evaluation metrics presented in Table 3, it is evident that the model corresponding to scenario 2 achieves exceptional performance with evaluation metrics exceeding a threshold of 0.95. Regrettably, this model had to be disregarded due to the existence of a bias in the underlying database. Specifically, a substantial portion of instances pertaining to the same day exhibited an imbalance in label distribution, with most instances having only one label.

Consequently, the model trained using the scenario 3 dataset was deemed as the most suitable choice for the task at hand. This particular model outperformed the one associated with scenario 1 in terms of performance. Furthermore, a meticulous analysis of the label distribution at different time intervals throughout the day revealed a consistent alignment with the empirical findings reported by Leis et al. (Leis et al., 2019). Thus, confirming the reliability and validity of our selected model.

An examination of the model's errors revealed that most of the incorrect predictions failed due to the absence of meaningful information in the messages after undergoing the pre-processing pipeline. This was a result of removing a significant number of component words, which were considered filler words. Consequently, setting a threshold for message quality would be necessary for the system to detect disruptions in the messages effectively.

In a real implementation of this system, a potential improvement would be to make predictions based on a set of messages from the same user, rather than just one message. This enhancement would further enhance the system's robustness.

In terms of energy efficiency, the developed models generate minimal emissions and consume low amounts of energy, underscoring their suitability for the development of tasks with minimal pollution impact.

4.2. Linear Regression

The results summarized in Table 4 highlight the performance of the linear regression models. Notably, the model derived from scenario 2 demonstrates the most favourable performance, exhibiting commendable evaluation metrics, similar to the findings in the binary classification task. However, due to the presence of data anomalies that introduce bias and compromise its reliability, this model had to be excluded from further consideration. It is worth mentioning that the dataset exhibits a similar bias as observed in the binary classification task.

Consequently, the linear regression model employed for the task was built using the dataset from scenario 3. This alternative model showcases superior performance compared to the one derived from scenario 1. Moreover, a detailed examination of the distribution of predicted values and residuals across different predictor variables confirms the effectiveness of the model, aligning well with established theoretical expectations.

In light of the comprehensive evaluation and analysis, the linear regression model trained on the scenario 3 dataset emerges as the most suitable choice, providing accurate predictions and demonstrating adherence to the expected patterns outlined in the existing literature.

In relation to this task, the error analysis yielded a similar outcome as observed in the case of binary classification. It was observed that certain messages contained minimal information after undergoing the pre-processing pipeline, necessitating the establishment of a message quality threshold. Furthermore, it was discovered that the majority of errors in probability estimation were of minor magnitude.

With regard to energy efficiency, the developed models exhibit a negligible emission footprint and consume minimal energy. This emphasizes their appropriateness for conducting tasks with minimal impact on pollutants.

5. Conclusion

In this study, a comprehensive methodology was employed to develop a system for characterizing eating disorders in text messages. The methodology encompassed a pre-processing pipeline specifically designed for text messages, the BERTopic topic modeling technique, and fine-tuning of a BERT model, which was pre-interleaved with unlabeled Spanish text.

The final outcome of the study resulted in the development of two models: a binary classification model and a linear regression model. These models were trained using pre-processed text messages, along with additional features such as the time the message was sent and its subject. The binary classification model is capable of predicting whether a message has been sent by an individual with an eating disorder or not, while the linear regression model provides the probability of this classification.

The utilization of the proposed methodology and models provides valuable insights and predictive capabilities in the domain of eating disorder characterization through text messages. This can contribute to early detection, intervention, and support for individuals at risk.

6. Acknowledgements

Thanks to all our colleagues at the Universidad Politécnica de Madrid, from the Life Supporting Technologies research group, who support us in all our endeavors.

7. References

Arowosegbe, A., & Oyelade, T. (2023). Application of Natural Language Processing (NLP) in Detecting and Preventing Suicide Ideation: A Systematic Review. *International Journal of Environmental Research and Public Health 2023, Vol. 20, Page 1514*, *20*(2), 1514. https://doi.org/10.3390/IJERPH20021514

Balakrishnan, V., & Ethel, L.-Y. (2014). Stemming and Lemmatization: A Comparison of Retrieval Performances. *Lecture Notes on Software Engineering*, *2*(3), 262–267. https://doi.org/10.7763/LNSE.2014.V2.134

Bellows, B. K., LaFleur, J., Kamauu, A. W. C., Ginter, T., Forbush, T. B., Agbor, S., Supina, D., Hodgkins, P., & DuVall, S. L. (2014). Automated identification of patients with a diagnosis of binge eating disorder from narrative electronic health records. *Journal of the American Medical Informatics Association : JAMIA*, *21*(e1). https://doi.org/10.1136/AMIAJNL-2013-001859

Blanchard, A. (2007). Understanding and customizing stopword lists for enhanced patent mapping. *World Patent Information*, *29*(4), 308–316. https://doi.org/10.1016/J.WPI.2007.02.002

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Cañete, J. and C. G. and F. R. and H. J.-H. and K. H. and P. J. (2020). Spanish Pre-Trained BERT Model and Evaluation Data. *PML4DC at ICLR 2020*.

Cliffe, C., Seyedsalehi, A., Vardavoulia, K., Bittar, A., Velupillai, S., Shetty, H., Schmidt, U., & Dutta, R. (2021). Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders: a retrospective cohort study. *BMJ Open*, *11*(12). https://doi.org/10.1136/BMJOPEN-2021-053808

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Dingemans, A. E., Bruna, M. J., & van Furth, E. F. (2002). Binge eating disorder: a review. *International Journal of Obesity and Related Metabolic Disorders : Journal of the International Association for the Study of Obesity*, *26*(3), 299–307. https://doi.org/10.1038/SJ.IJO.0801949

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*, 65. https://doi.org/10.3389/FSOC.2022.886498/BIBTEX

emoji · PyPI. (n.d.). Retrieved May 27, 2023, from https://pypi.org/project/emoji/

Feldman, J. (2023). Classifying Pro-Eating Disorder Images on Twitter Using Deep Learning. *Intersect: The Stanford Journal of Science, Technology, and Society*, *16*(2). https://ojs.stanford.edu/ojs/index.php/intersect/article/view/2441

Fox, C. (1989). A stop list for general text. *ACM SIGIR Forum*, *24*(1–2), 19–21. https://doi.org/10.1145/378881.378888

Girdhar, Y., Giguère, P., & Dudek, G. (2012, June). *Autonomous Adaptive Underwater Exploration using Online Topic Modelling*.

Grootendorst, M. R. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv*, *abs/2203.05794*.

Haddadi, P., & Ali Besharat, M. (2010). Resilience, vulnerability and mental health. *Procedia - Social and Behavioral Sciences*, *5*, 639–642. https://doi.org/10.1016/J.SBSPRO.2010.07.157

Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2019). Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis. *Journal of Medical Internet Research*, *21*(6). https://doi.org/10.2196/14199

Liangjie, H., & Davison, B. D. (2010). *Empirical Study of Topic Modeling in Twitter*. http://www.bit.ly

Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, *2*, 246–252. https://doi.org/10.1016/j.jnlssr.2021.10.003

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, *5*(1). https://doi.org/10.1186/S40064-016-3252-8

Mármol-Romero, A. M., Moreno-Muñoz, A., Plaza-del-Arco, F. M., Molina-González, M. D., Martín-Valdivia, M. T., Ureña-López, L. A., & Montejo-Ráez, A. (2023). Overview of MentalRiskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish. *Procesamiento Del Lenguaje Natural*, *71*.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*. https://doi.org/10.21105/joss.00205

Mcinnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association : JAMIA*, *18*(5), 544–551. https://doi.org/10.1136/AMIAJNL-2011-000464

nltk · PyPI. (n.d.). Retrieved May 27, 2023, from https://pypi.org/project/nltk/#data

pyspellchecker · PyPI. (n.d.). Retrieved May 27, 2023, from https://pypi.org/project/pyspellchecker/

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

spacy · PyPI. (n.d.). Retrieved May 27, 2023, from https://pypi.org/project/spacy/

translate · PyPI. (n.d.). Retrieved May 27, 2023, from https://pypi.org/project/translate/