Procesamiento del Lenguaje Natural, Revista nº 72, marzo de 2024, pp. 87-98 recibido 14-12-2023 revisado 15-01-2024 aceptado 29-01-2024

Generation of social network user profiles and their relationship with suicidal behaviour

Generación de perfiles de usuarios de redes sociales y su relación con el comportamiento suicida

Jorge Fernandez-Hernandez,[1] Lourdes Araujo,[1,2] Juan Martinez-Romo[1,2]
[1]NLP & IR Group - Universidad Nacional de Educación a Distancia (UNED)
[2]Instituto Mixto UNED-ISCIII (IMIENS)
{jfernandez, lurdes, juaner}@lsi.uned.es

Abstract: Suicide is one of the leading causes of death worldwide, so characterising individuals with such tendencies can help prevent suicide attempts. In this study, a corpus, called *SuicidAttempt*, of Telegram messaging app users, both with and without explicit mentions of suicide attempts, has been compiled in Spanish. For each user, different demographic features were semi-automatically annotated by different systems, some supervised and some unsupervised. Finally, the collected features and linguistic features extracted from users' messages were analysed to characterise different groups based on their relationship with suicidal behaviour. The results indicate that by detecting these demographic and psycholinguistic features, it is possible to characterise specific at-risk groups and gain detailed insight into the profiles of those who engage in such acts.
Keywords: Suicidal behaviour, profiling, corpus creation, social networks

Resumen: Actualmente el suicidio es una de las principales causas de muerte en el mundo, por lo que poder caracterizar a personas con esta tendencia puede ayudar a prevenir posibles intentos de suicidio. En este trabajo se ha recopilado un corpus, llamado *SuicidAttempt* en español compuesto por usuarios con o sin menciones explícitas de intentos de suicidio, usando la aplicación de mensajería Telegram. Para cada uno de los usuarios se han anotado distintos rasgos demográficos de manera semi-automática mediante el empleo de distintos sistemas, en unos casos supervisados y en otros no supervisados. Por último se han analizado estos rasgos recogidos, junto con otros lingüísticos extraídos de los mensajes de los usuarios, para intentar caracterizar distintos grupos en base a su relación con el comportamiento suicida. Los resultados sugieren que la detección de estos rasgos demográficos y psicolingüísticos permiten caracterizar determinados grupos de riesgo y conocer en profundidad los perfiles que realizan dichos actos.
Palabras clave: Comportamiento suicida, identificación de perfiles, creación de corpus, redes sociales

1 Introduction

Suicide is currently one of the leading causes of death worldwide, with approximately 700,000 deaths annually, and the fourth death cause among the young people, as reported by the World Health Organization[1]. In Spain, the number of suicide-related deaths is estimated to be thrice that of those resulting from traffic accidents[2]. Therefore, it is crucial to comprehend the characteristic patterns of individuals with suicidal tendencies and classify them to identify particular social groups with an increased susceptibility

[1]https://www.who.int/news/item/17-06-2021-one-in-100-deaths-is-by-suicide

[2]https://www.mjusticia.gob.es/es/ElMinisterio/GabineteComunicacion/Paginas/211221-NP-Estudio-Epidemiologia-y-Toxicologia-de-las-muertes-por-suicidio.aspx

ISSN 1135-5948 DOI 10.26342/2024-72-6

to suicide.

Characterising population groups with different sensitivities to mental health issues is a crucial step in identifying the most vulnerable populations and designing appropriate support initiatives. Social media platforms offer a wealth of information for this research, as they allow the identification of linguistic and demographic patterns within user-generated content.

This work contributes to the development of a mental health observatory that furnishes health professionals with the most recent data on various population groups, thereby enabling them to accurately interpret individual cases and propose general intervention strategies. The principal objective of this study is to enhance the identification of population profiles associated with suicidal behaviour, using information gathered from social media platforms.

Specifically, a collection of messages regarding suicide attempts on the instant messaging app Telegram has been compiled. The collection has been classified manually, separating positive and negative cases, and has also been annotated with demographic features, such as age, gender, origin or employment status. Semi-automated detection systems, outlined in this article, were utilised for the annotation of these features. These systems are designed specifically for each feature, depending on the availability of external training data and the difficulty of detecting each feature. These systems not only represent a support to manual annotation, but they also serve as the foundation for an automatic system to collect messages and profile the population at risk.

Based on the annotated collection, a study was conducted on the correlation between demographic and linguistic features, and suicidal tendencies. Despite the limited size of the corpus, the findings suggest two key facts. Firstly, there are specific demographic groups that exhibit a considerably higher incidence of suicidal behaviour. In addition, the usefulness of the use of certain linguistic features in at-risk groups has been highlighted.

The article is structured as follows: Section 2 presents the state of the art on the use of artificial intelligence for suicide-related topics, as well as in the profiling of authors. Section 3 presents the collected corpus and the different features annotated in it. Section 4 presents the different methods that have been developed with the aim of extracting the demographic features, and which have been used as an aid to the annotation of the corpus. Section 5 analyses the different features considered in this work that constitute a profile of each user and their relationship with suicide cases. Finally, the conclusions and future work will be presented in Section 6.

2 State of the art

The growing popularity of artificial intelligence has led to its use in more and more fields, including psychology and the identification of different mental disorders. One of the most comprehensive and well-known studies was conducted by Schwartz et al. (2013), which analysed 700 million words and 75,000 volunteers to associate certain words, phrases or speech patterns with different personality profiles. Another recent example is the *CLPsych2019* task (Zirikly et al., 2019), which aims to classify suicide risk into four levels based on Reddit posts written in English. In Du (2023), linguistic features are utilized with classical machine learning methods to predict the most representative psychological state of a text (anxiety, depression, suicide ideation, or "normal"). In (Fernandes et al., 2018), a rule-based system is employed to detect instances of suicidal ideation in English texts, alongside a hybrid approach that utilizes both rules and machine learning techniques to identify suicide attempts. It is also worth mentioning the competition eRisk, which has covered the early detection on the Internet of a wide variety of mental disorders since their first edition in 2017 (Losada, Crestani, and Parapar, 2017). For example, in their last edition the disorders were depression, gambling and eating disorders (Parapar et al., 2023).

The association between distinct demographic features and suicidal behaviour has been a topic of research. Rancāns et al. (2016) conducted a study of the Latvian population and found that middle-aged men living alone and with a low level of education were more likely to exhibit suicidal tendencies, while women with only a low level of education exhibited the highest risk factor. In Akkaya-Kalayci et al. (2018), the study focuses on features associated with personal relationships among young people in Turkey.

The findings suggest that for women, intra-family issues tend to be linked to suicidal behaviour, while for men, relationship problems tend to have a stronger association.

Among the features to be extracted in this work, gender seems to be the one that has been studied the most. Its identification has largely employed classical machine learning techniques, like support vector machines (SVM) or decision trees, rather than deep learning. Among these conventional algorithms, the most successful have been the SVMs (Pizarro, 2019) (Yang et al., 2021), although ensembles have also yielded promising results (Piot-Perez-Abadin, Martin-Rodilla, and Parapar, 2021). Regarding deep learning algorithms, Heidari, Jones, and Uzuner (2020) train separate neural networks for each gender using the Bi-LSTM architecture. Unsupervised learning techniques, such as clustering, can be useful not only for gender identification, but also for analysing the different groups obtained (Bamman, Eisenstein, and Schnoebelen, 2014). On social networks like Twitter, each user has an associated profile picture, which can be used to create classifiers on two levels: on one hand, the images are analysed, while on the other, it focuses on text, with their output being combined. (Wang et al., 2019).

The determination of nationality or provenance has had limited research, with a greater emphasis on handwritten texts rather than text (Al Maadeed and Hassaine, 2014) (Choudhury et al., 2022). Consequently, it is more akin to image analysis than text analysis.

Employment status and profession have typically been addressed as a problem of entity recognition and POS-tagging, as in the case of the MEDDOPROF task (Lima-López et al., 2021). The use of transformers such as *XLM-R* (Lange, Adel, and Strotgen, 2021), the more familiar *BERT* (Mesa-Murgado et al., 2021) or a mixture of the latter with *FLAIR* (Balouchzahi, Sidorov, and Shashirekha, 2021), is the most common method in this scenario.

There is a limited number of studies regarding age, with the majority focusing on *PAN* tasks between 2013 and 2016 (Rangel et al., 2013) (Rangel et al., 2014) (Rangel et al., 2015) (Rosso et al., 2016). These studies predominantly use classical machine learning algorithms such as SVMs or ensembles.

3 Corpus

The corpus was created gathering messages from two different Telegram groups, both focused on mental health problems. One of them was more focussed on suicide, and the other one with focus in anxiety and depression, both groups having positives and negatives users for suicide attempts. These groups are not restricted to a certain nationality, and have users from different Spanish-speaking countries. Despite the fact that the main language is Spanish, some users from non-Spanish-speaking countries can be found, although all of them write in Spanish. So the corpus *SuicidAttempt* comprises 141,894 messages authored by 589 unique users, each user having a mean of 290 messages, between late 2021 and mid-2023 in groups associated with mental disorders.

The users in the corpus can be classified as either positive, where an explicit suicide attempt is mentioned, or negative, where such mention is absent. For classify a user, we first search in his messages for terms related to suicide, with a later manual review to validate. A user who solely mentions suicidal ideation is considered negative. Some examples of sentences that could be considered as explicit mentions are: “I attempted suicide one year ago” or “I consumed thirty pills, but I woke up in the hospital”. Some examples of sentences that could be considered as only ideation are: “I want to die” or “Could anyone give me a quick way to die”.

The 589 users are divided as follows: 156 are positive and 433 are negative users.

Each user in the corpus has received semi-automatic annotation in terms of their gender, origin and employment status. It should be noted that in some cases, users may not provide all information relevant to the traits in question, and therefore no annotation can be made.

The evaluation of the agreement among annotators was measured by Fleiss kappa value (Fleiss, 1971) obtaining 0.78 (“substantial agreement”) in the case of the attempt suicidal annotation, and 0.86 (“almost perfect agreement”) in the annotation of the traits. In simple terms, the kappa coefficient corresponds to the ratio of observed concordances over the total of observations, having excluded all random concordances. The kappa coefficient takes values between -1 and +1.

Three different categories have been defined for the trait "Gender": "**Male**", "**Female**" and "**No binary**", appearing unannotated in case the user in question could not be classified in any of the above three categories during the manual review.

In the corpus, the trait "Origin" was split into two different categories - "Place of birth" and "Place of residence" - because for some users they were different. However, for most of the users both locations will be the same, and will only differ if a specific reference is detected during the manual review.

Something similar occurs with the employment status, which in the corpus is split into two different traits: "Employment status" and "Profession". "Employment status" could be five different categories: "**Work**" for users with job, "**Unemployed**" for those who are unemployed, "**Student**" for those who an explicit mention was found (i.e "I study computer science") or implicitly (i.e "I just got out of class"), "**Homemaker**" for those who have explicitly mentioned their role of homemaker and "**Voluntary**" if an explicit mention exists. If a user does not fit into any of the above categories, then this trait will not be annotated in the corpus.

The other trait is "Profession", which indicates the specific employment activity or place of work. It is possible to find the combination of "**Work**" as an employment status with no profession, e.g. due to a mention of "I just got off work", but without further details about the profession.

Finally, the "Age" trait, instead of being taken as a number, has been divided into the following age ranges: "**<18**", "**18-24**", "**25-34**", "**35-49**" and "**>50**". These ranges are based on those proposed in Rangel et al. (2014), although in our case there were no users over 65, so the highest range is over 50.

An example of the corpus data can be seen in the Table 1.

4 Techniques used for corpus annotation

The annotation process was supported by a series of systems that carried out initial automated tagging, followed by manual revision. The choice of system for each trait relied on the available resources.

In the case of gender, employment status and profession, the systems were supervised; while in the case of place of birth and residence, the approach has been unsupervised. Age was recorded manually, as no dataset with a sufficient number of cases was found to create a system.

4.1 Gender

In the case of gender, there are publicly available datasets that have allowed us to annotate this trait using a machine learning system. A system using the *transformers* technology was also tested but found to be less effective than systems based on classical classification algorithms and thus discarded.

For the base gender detection method, data from the PAN's 2018 and 2019 author profiling task was used (Rangel et al., 2018; Rangel and Rosso, 2019). In both tasks, a dataset of 100 tweets is provided for each user along with their gender information. Users identified as bots were discarded for the 2019 data set. Additionally, messages that were just retweets were also eliminated. The base models were trained on a total of 4,479 users, which were divided as 2,238 female users and 2,241 male users.

For each user, their messages were concatenated using *<FIN>* as tag. Certain special text sequences may appear in the tweets, which have been edited and replaced by different tags:

- The links by *<URL>*
- User mentions (@*username*) by *<USR>*
- Hashtags by *<HTG>*
- The emojis have been removed, although the number of them used by each user has been counted beforehand, to be employed as a feature by the classification algorithms.

The features employed by the classification algorithm can be divided in three different groups:

- **LIWC**: Features obtained from *LIWC 2015* (*Linguistic Inquiry and Word Count*)[3] employing *Spanish Dictionary 2007*. This software identifies 90 dimensions, each one determining the degree that the users employ words that connote positive or negative emotions, self-references, pronouns, etc.

[3]https://www.liwc.app/

Gender	P. Birth	Residence	E. Status	Profession	Age	Suicide
Male	Spain	Spain	Work	Lawyer	35-49	Positive
Female	Spain	Spain	Work	Health	25-34	Positive
Female	Argentina	Argentina	Student		<18	Negative
Male	Colombia	Colombia	Work	English Teacher	25-34	Negative

Table 1: Example of annotation from 4 users in the corpus.

- **TF-IDF**: Features obtained with *Tf-Idf* technique (Term Frequency – Inverse Document Frequency). This method was employed to analyse words, using unigrams and bigrams as terms, as well as characters, using trigrams, tetragrams and pentagrams as terms. In both cases, terms that appeared in over 70% of the documents were excluded. Finally, the Singular Value Decomposition (SVD) technique was applied to reduce the number of features.
- **Number of emojis**: This feature defines the total number of emojis used by a user in their messages.

To obtain the system for semi-automatic annotation of the corpus, firstly the most effective feature mentioned earlier have been found. The SVM method has been employed, as it provided the best results in a preliminary study. The evaluation was carried using a cross-validation with 10 folds. Table 2 shows how the best results are obtained by combining the 3 groups of features (*Tf-Idf* Measure (TFIDF), linguistic features of the *LIWC* (LIWC) and the number of emojis (N_EMO)).

	P	R	F1
LIWC	71.79	71.73	71.71*
N_EMO	65.49	62.61	60.71*
TFIDF	80.93	80.84	80.83*
LIWC+TFIDF	80.97	80.91	80.90*
LIWC+TFIDF+N_EMO	**82.02**	**81.96**	**81.95**

Table 2: Precision (P), Recall (R) and F1-Score (F1) employing SVM and cross-validation with different combination of features for gender identification. A statistical significance test has been carried between the best combination of features and the other options, being significant the difference in all cases (marked with *).

After selecting the optimal combination of features, the next step was to choose the best model. The algorithms tested were: SVM, Decision Trees, Naive-Bayes, Gradient Boosting, Random Forest and AdaBoost, all of them being implementations of the *Sklearn*[4] library, each one trained with their default hyperparameters. As can be seen in Table 3, the best results are achieved with SVM, so this was the model trained on the PAN18 and PAN19 data using the training set. Furthermore, we aimed to assess the model's performance on the test set of the PAN 2019 task, in addition to *SuicidAttempt* corpus, obtaining a precision of 78.55, a recall of 78.39 and a F1-Score of 78.37.

	P	R	F1
SVM	**82.02**	**81.96**	**81.95**
Decision Tree	67.98	67.96	67.96
Naive Bayes	66.79	66.44	66.25
GradientBoosting	79.19	79.15	79.14
RandomForest	77.48	77.45	77.45
AdaBoost	74.48	74.44	74.43

Table 3: Precision (P), Recall (R) and F1-Score (F1) for multiple classification algorithms employing cross-validation for gender identification with Tf-Idf, LIWC and the number of emojis.

The gender of the Telegram users was annotated employing the previous model,and manually revised, achieving the next results (see Table 4): 75.27 of precision, 72.21 recall and 72.56 of F1-Score. The results in this case are slightly lower than those obtained with PAN data. The lower results in the SuicidAttempt corpus may be due to differences with the PAN texts. In any case, the results are high enough to be useful for assisting with the annotation task and can serve as a baseline for future research on gender identification systems.

4.2 Place of birth and residence

The origin of the user is another of the traits considered in this work. In this case, the semi-automatic annotation has not been done based on a supervised learning system, due

[4]`https://scikit-learn.org`

to the unavailability of a dataset that included all the nationalities considered. The methodology employed made use of a dictionary that included all the Spanish-speaking nations (including the Philippines and Equatorial Guinea), Brazil and Portugal as possible origins. For each country, a list of related expressions has been generated, considering the name of the country, the capital, the nationality, and principal cities. For each user, the occurrence of terms associated with a country is counted, and the place with the highest frequency is noted as the origin.

During the manual review, certain users have been found to mention their birth in one country while living in another. For this reason, two different traits were considered: the place of birth and the place of residence.

To derive metrics and perform an analysis on this initial algorithm, we considered an annotation as a hit if it corresponded to the place of birth or the place of residence. This system achieves the next results (see Table 4): a precision of 88.66; 92.46 as recall and 90.33 of F1-Score. These results are sufficient to provide us with a reasonably accurate annotation of the user's nationality.

4.3 Employment status and profession

For the semi-automatic annotation, we used the data from the task *MEDDOPROF* (Lima-López et al., 2021), specifically subtask 2. In this subtask, the objective is to tag professions and identify if they refer to the patient, a sanitary, a familiar or another category.

The data from *MEDDOPROF* have been used, together with the code and process proposed in Lange, Adel, and Strotgen (2021), to train three *transformers* using *xlm-roberta-large* as architecture. One is trained from scratch, while the other two were fine-tuned on the pre-existing models discussed in the article. Once the annotation is done for each of the 3 models, with an ensemble, the results of the three models are combined with a majority vote strategy. Of the referenced entities, only those that have been identified as "PACIENTE" (patient) have been considered. These annotations serve two purposes. Firstly, they serve as input for the employment status identification system. Secondly, they aid in speeding up the annotation of the profession trait.

The annotation of employment status is obtained through a rule-based system using the annotations from the previous ensemble as input. The multiple tagged parts of text, for each user, are reviewed to identify expressions associated with each of the considered employment status, except for "Work". For example, looking for "no" (negation) and "trabajo" (work) in the same sentence to classify a user as "unemployed". If the user has tagged text, but does not meet any of the rules being considered, then the user is classified as "work". If the user has no employment-related mentions, the employment status is left blank. The results of the system can be viewed in the Table 4.

	P	**R**	**F1**
Gender	75.27	72.21	72.56
Origin	88.66	92.46	90.33
E.Status	65.34	60.61	60.87

Table 4: Results obtained by the different systems developed to annotate the traits.

5 *Analysis between the traits and suicidal behaviour*

The study's ultimate aim is to characterise individuals based on their connection to suicidal behaviour. This is done using the demographic traits and the *LIWC* features. It is also important to remember that the corpus is gathered from groups centred around mental disorders, such as anxiety and depression. Thus, these connections are useful for population groups with similar situations, but they may not necessarily be generalizable.

Of the traits noted, gender, age, employment status and place of residence were considered. The place of birth has not been taken into account, as in most cases it is the same as the place of residence and the actual residence has been considered more important to characterize the user. Profession has not been taken into account because it is a category with open labels. In this case, it remains pending, for future work, to standardise this trait in order to be able to analyse it statistically. Users with unidentified traits have been excluded from consideration for each trait. Similarly, labels that appear in less than 5% of users have not been taken into account.

In the case of gender, due to the restrictions mentioned above, only users labelled as

"Male" and "Female" were considered. The results for this trait are presented in Figure 1. The findings reveal a slightly higher frequency of positives among women compared to men, although the difference is small. However, more data is required to determine whether this difference is maintained or increased before concluding that women are a high-risk group.

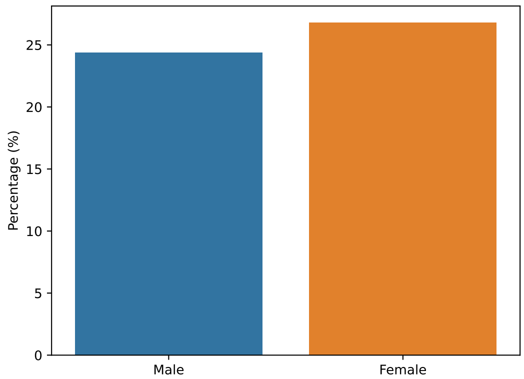

Figure 1: Percentage of positive users for each gender considered.

The next trait studied was age. After applying the above-mentioned restrictions, the ranges *"<18"*, *"18-24"*, *"25-34"*, *"35-49"* were considered. In this case, one can clearly find an at-risk group, the under-18s, with approximately 45% of these users reporting suicidal behaviour, as shown in Figure 2. The following most frequent user group is the 18-24 age range, where almost 40% of users are reported to have attempted suicide. If we divide the age into two groups, under 25 and over 25, and we consider only the positive users, 55% of them were under 25, as shown in Figure 3. Despite the reduced size of the dataset, it seems that age could be a distinguishing factor to consider when examining suicidal behaviour.

Given the high density of users aged below 25 who have attempted suicide, the study examined their employment status to establish any potential correlations. Figure 4 indicates that students have the highest density of positive users, followed by the unemployed.

As it was expected that these groups might be conditioned by age, mainly students, it was decided to study the relationship between both traits. As can be seen in Figure 5, among the users labelled as students, about 90% of them are under 25 years old, which has been observed to be a prevalent group among the positives. Even more interesting is the case of unemployed users, the second group with the highest frequency of positives with values very close to those of students, which in this case are perfectly distributed between those under and over 25 years old. This suggests that the connection between unemployment and suicidal behaviour is more direct and not as age-related as in the case of students.

The last demographic trait considered was place of residence, with Spain, Mexico, Argentina, Colombia, Venezuela, and Peru fulfilling the criteria mentioned earlier. Among all of them, it can be clearly seen how Argentina stands out from the rest of the countries, whereas the remaining countries ex-

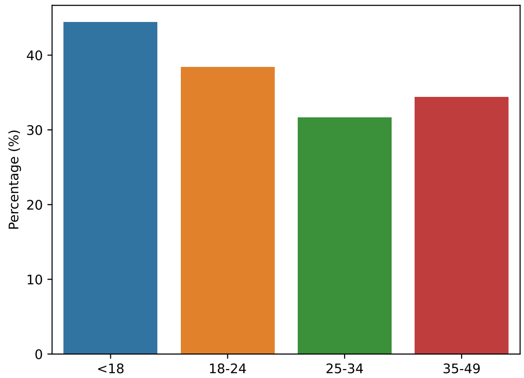

Figure 2: Percentage of positive users in each age group considered.

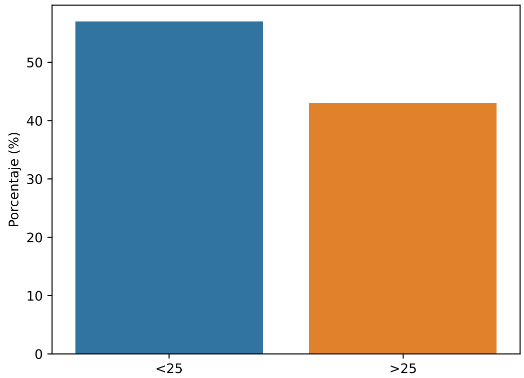

Figure 3: Percentage of individuals under and over the age of 25, from the positive group.

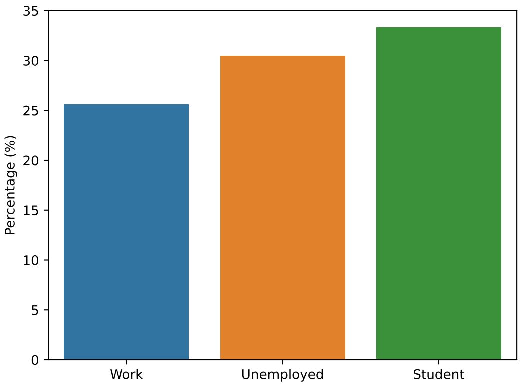

Figure 4: Percentage of positive users for each of the employment status considered.

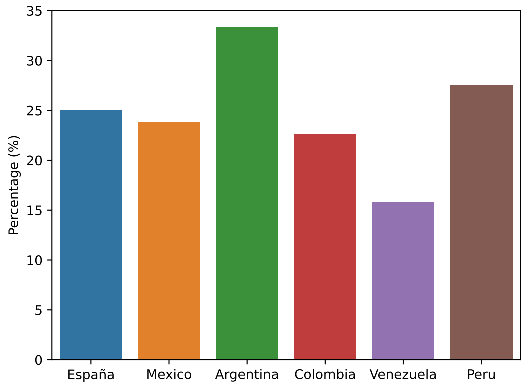

Figure 6: Percentage of suicidal behaviour in the different countries of residence considered.

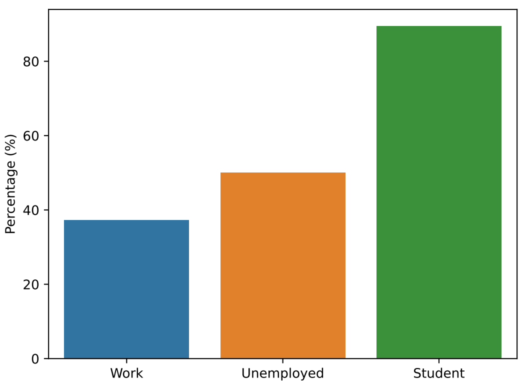

Figure 5: Percentage of individuals under 25, broken down by employment status.

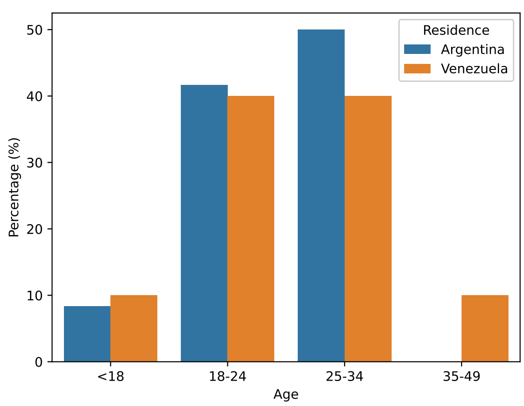

Figure 7: Percentage of age ranges among all users living in Argentina and Venezuela.

hibited comparable proportions, aside from Venezuela, which had a lower frequency, as shown in Figure 6.

To understand the higher suicide attempt rates in Argentina compared to lower rates in Venezuela, we examined age distribution in these countries. In this case, as can be seen in Figure 7, the percentages of age ranges are similar in both nations, with half of all users being under the age of 25. These results seem to indicate that the percentages of positives in these countries would not be conditioned by age.

Other features to consider are the characteristics extracted from the linguistic analyser *LIWC*. These will enable us to characterise users exhibiting positive and negative behaviour based on their linguistic characteristics. By analysing each feature and comparing differences between positive and negative users, we can identify the most significant characteristics (see Figure 8). The feature with the greatest average divergence is *Muerte* (death), reflecting a user's usage frequency of words that have been labelled in the *LIWC* on death-related topics, its usage is higher among positive users compared to negative ones. The subsequent five categories displaying the highest average difference (*Triste* (sad), *Salud* (health), *Enfado* (anger), *Maldec* (cursing) y *verbYO*) mainly relate to negative emotional states or attitudes. The *verbYO* feature, is interesting as it indicates how frequently the first-person singular verb forms are used. Thus, these results seem to indicate that those with suicidal behaviour talk more about themselves. Another interesting feature is the use of negations (feature *Negacio*), which seem to be used on average more by those with suicide attempts.

On the other side, Figure 8, seems to suggest that the negative group uses more words related to cognitive processes according to *LIWC*, in this case the categories *MecCog* and *Certeza* (certainty), as well as an increased use of non-standard punctuation symbols (*OtherP*).

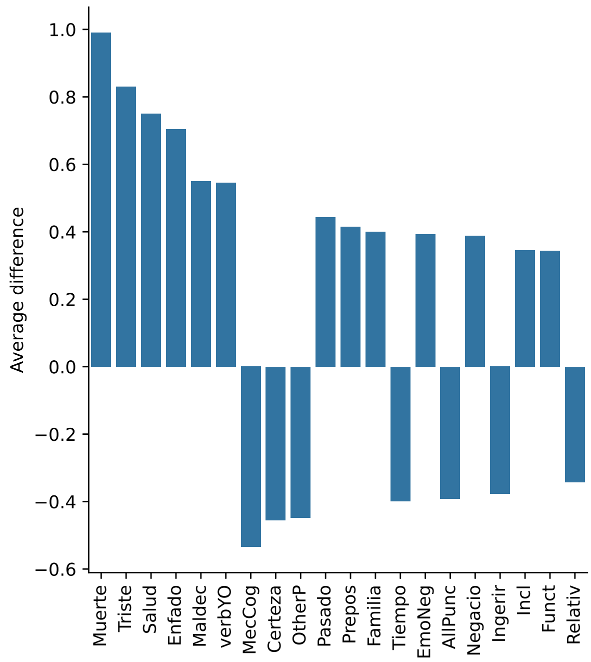

Figure 8: Selection of the 20 LIWC features with the highest average difference. A positive difference suggests higher usage of positive language.

6 Conclusions and future work

In this work, progress has been made in creating a corpus to characterise suicidal behaviour in Spanish. Information on users' gender, place of birth, residence, employment status, profession, and age were recorded. The study also explored different base systems for identifying demographic traits. Lastly, this work attempted to characterise and identify specific risk groups and distinguish linguistic characteristics.

From this analysis it has been observed that the most relevant demographic feature for the study is age, with a higher prevalence among younger users, with around 15% more users under the age of 18 having attempted suicide than in the 25-34 age range. WHO [1] also find prevalence of suicide attempts between the young people.

In terms of gender, a greater proportion of female users have attempted suicide compared to male users, although the variation is not considered significant. Therefore, acquiring further data to determine if this trend remains consistent or if the prevalence of suicidal behaviour among either gender intensifies, would be beneficial.

Our analysis suggests that there is a higher prevalence of suicidal behaviour among students and the unemployed. However, a noticeable correlation with age is evident in the case of the students. Conversely, such a relationship is not observed in the unemployed, a group that is usually associated with economic problems, that has been identified as a risk situation by some organizations, such as the OMS or the WHO. Therefore, it may be beneficial to further explore this demographic group.

Regarding origin, these analyses seem to indicate a higher incidence among people residing in Argentina, but the relationship to age remains unclear. Further investigation of this group is needed to determine if this pattern holds with additional data.

Our conclusions about the linguistic features are similar as the obtained in other studies (Lopez-Castroman et al., 2020). This means that positive users tend to use words that could be classified as negative emotions or feelings, such as sadness or anger, while also frequently using the first person. Negative users, on the other hand, tend to focus on topics that can be framed within different cognitive processes.

Our aim is to extend and advance the current work achieved. Specifically, continue expanding the corpus with more users, and consider other social networks such as Reddit. Furthermore, we will explore additional traits, for example, social or economic issues or addictions.

It would be worthwhile to carry on with the development of automatic extraction systems for various traits, with special attention to age, which could not be determined automatically because there was not a large enough dataset to facilitate this process. The annotations obtained during the development of this work, could be used to develop more sophisticated systems. For example, the counts obtained from each country could be used as features for a machine learning system. Additionally, the best features observed in this work could be combined with the embeddings from a transformer in the case of gender.

Regarding employment status and profession, more work could be done. For example, the annotations of the system, trained with the MEDDOPROF task data, could be used to automatically infer employment status. For the profession, it seems necessary to define categories of professions with similar characteristics, for example, combining doctors and nurses in a category that could be "Health professional". In this way, we will be able to analyse possible relationships between professional groups from different fields and suicidal behaviour.

We also have planned the release of the corpus in the future, under certain commitments. Before the release, we have to study the legal requirements and how to deal with the anonymization of sensitive data such as names.

Finally, it would be interesting to have automatic systems to detect if a user has suicidal behaviour through text, since as seen in Section 5, it is possible to characterise users in this way.

7 Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Innovation within the DOTT-HEALTH Project (MCI/AEI/FEDER, UE) under Grant PID2019-106942RB-C32, OBSERMENH Project (MCIN/AEI/10.13039 and NextGenerationEU"/PRTR) under Grant TED2021-130398B-C21, and EDHER-MED under grant PID2022-136522OB-C21 as well as project RAICES (IMIENS 2022). The authors would like to thank Hewei Shan for his contribution to the gathering of social media messages for the construction of the corpus.

References

Akkaya-Kalayci, T., N. D. Kapusta, D. Winkler, O. D. Kothgassner, C. Popow, and Z. Özlü-Erkilic. 2018. Triggers for attempted suicide in istanbul youth, with special reference to their sociodemographic background. *International journal of psychiatry in clinical practice*, 22(2):95–100.

Al Maadeed, S. and A. Hassaine. 2014. Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing*, 2014(1):1–10.

Balouchzahi, F., G. Sidorov, and H. L. Shashirekha. 2021. Adop fert-automatic detection of occupations and profession in medical texts using flair and bert. In *IberLEF@ SEPLN*, pages 747–757.

Bamman, D., J. Eisenstein, and T. Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160.

Choudhury, A. P., P. Shivakumara, U. Pal, and C.-L. Liu. 2022. Eau-net: A new edge-attention based u-net for nationality identification. In *International Conference on Frontiers in Handwriting Recognition*, pages 137–152. Springer.

Du, X. 2023. Lexical features and psychological states: A quantitative linguistic approach. *Journal of Quantitative Linguistics*, pages 1–23.

Fernandes, A. C., R. Dutta, S. Velupillai, J. Sanyal, R. Stewart, and D. Chandran. 2018. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Scientific reports*, 8(1):7426.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Heidari, M., J. H. Jones, and O. Uzuner. 2020. Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 480–487. IEEE.

Lange, L., H. Adel, and J. Strotgen. 2021. Boosting transformers for job expression extraction and classification in a low-resource setting.

Lima-López, S., E. Farré-Maduell, A. Miranda-Escalada, V. Brivá-Iglesias, and M. Krallinger. 2021. Nlp applied to occupational health: Meddoprof shared task at iberlef 2021 on automatic recognition, classification and normalization of professions and occupations from medical texts. *Procesamiento del Lenguaje Natural*, 67:243–256.

Lopez-Castroman, J., B. Moulahi, J. Azé, S. Bringay, J. Deninotti, S. Guillaume, and E. Baca-Garcia. 2020. Mining social networks to improve suicide prevention: A scoping review. *Journal of neuroscience research*, 98(4):616–625.

Losada, D., F. Crestani, and J. Parapar. 2017. erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. pages 346–360, 08.

Mesa-Murgado, J., P. López-Úbeda, M. C. Díaz-Galiano, M. T. M. Valdivia, and L. A. U. López. 2021. BERT representations to identify professions and employment statuses in health data. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021)*, volume 2943 of *CEUR Workshop Proceedings*, pages 758–769.

Parapar, J., P. Martín-Rodilla, D. E. Losada, and F. Crestani. 2023. erisk 2023: Depression, pathological gambling, and eating disorder challenges. In J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo, editors, *Advances in Information Retrieval*, pages 585–592, Cham. Springer Nature Switzerland.

Piot-Perez-Abadin, P., P. Martin-Rodilla, and J. Parapar. 2021. Gender classification models and feature impact for social media author profiling. In *International Conference on Evaluation of Novel Approaches to Software Engineering*, pages 265–287. Springer.

Pizarro, J. 2019. Using n-grams to detect bots on twitter. In *CLEF (Working Notes)*.

Rancāns, E., T. Pulmanis, M. Taube, L. Spriņģe, B. Velika, I. Pudule, and D. Grīnberga. 2016. Prevalence and sociodemographic characteristics of self-reported suicidal behaviours in latvia in 2010: a population-based study. *Nordic journal of psychiatry*, 70(3):195–201.

Rangel, F., F. Celli, P. Rosso, M. Potthast, B. Stein, W. Daelemans, et al. 2015. Overview of the 3rd author profiling task at pan 2015. In *CLEF2015 Working Notes. Working Notes of CLEF 2015-Conference and Labs of the Evaluation forum.* Notebook Papers.

Rangel, F. and P. Rosso. 2019. Overview of the 7th author profiling task at pan 2019: bots and gender profiling in twitter. *Working notes papers of the CLEF 2019 evaluation labs*, 2380:1–7.

Rangel, F., P. Rosso, I. Chugur, M. Potthast, M. Trenkmann, B. Stein, B. Verhoeven, and W. Daelemans. 2014. Overview of the 2nd author profiling task at pan 2014. In *CLEF 2014 Evaluation Labs and Workshop Working Notes Papers*, pages 1–30.

Rangel, F., P. Rosso, M. Koppel, E. Stamatatos, and G. Inches. 2013. Overview of the author profiling task at pan 2013. In *CLEF conference on multilingual and multimodal information access evaluation*, pages 352–365. CELCT.

Rangel, F., P. Rosso, M. Montes-y Gómez, M. Potthast, and B. Stein. 2018. Overview of the 6th author profiling task at pan 2018: multimodal gender identification in twitter. *Working notes papers of the CLEF*, 192.

Rosso, P., F. Rangel, M. Potthast, E. Stamatatos, M. Tschuggnall, and B. Stein. 2016. Overview of pan'16: new challenges for authorship analysis: cross-genre profiling, clustering, diarization, and obfuscation. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 332–350. Springer.

Schwartz, H. A., J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Wang, Z., S. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, and D. Jurgens. 2019. Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference*, pages 2056–2067.

Yang, Y.-C., M. A. Al-Garadi, J. S. Love, J. Perrone, and A. Sarker. 2021. Automatic gender detection in twitter profiles for health-related cohort studies. *JAMIA open*, 4(2):ooab042.

Zirikly, A., P. Resnik, O. Uzuner, and K. Hollingshead. 2019. Clpsych 2019 shared task: Predicting the degree of suicide risk in reddit posts. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology*, pages 24–33.