User profiling: voting scheme

Notebook for PAN at CLEF 2022

Maria Fernanda Artigas-Herold[1], Daniel Castro-Castro[2]

[1]Informatics Department, University of Oriente "Julio A. Mella", Santiago de Cuba, Cuba
[2]Information Retrieval Lab, Computer Science Department, University of A Coruña, Spain

Abstract
Social networks have been playing an important role in the life of human beings for the last years, they have become a way to express and share information widely. In them, many people create harmful and offensive content towards others, such as irony, sarcasm and the use of stereotypes to refer to certain groups in society. Because the information shared on the internet grows very fast, it is necessary to have systems that can automatically detect this unwanted behavior on networks. In this paper, we describe our approach to the Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) task promoted by PAN CLEF 2022, where we want to identify profiles of users who post ironic content on Twitter. Our proposal is to builds models based on n-grams of characters and words, as well as non-English words in combination with SVM and RF as classification algorithms, and obtains a majority vote of those with the best results for each representation. Our solution reached an accuracy of 91.67%.

Keywords
Author Profiling, N-grams, Voting Classifier, Irony, Stereotypes Spreaders

1. Introduction

Irony is a rhetorical figure that consists of saying the opposite of what is meant, using a tone, gesture or words that insinuate the interpretation that should be made. On the other hand, sarcasm is a way of mocking in which it is intended to imply the opposite or to express displeasure. Sarcasm contains indirect criticism, but most of the time it is exposed in an obvious way. Sarcasm and irony have always been present in our society, either to make fun of other people or a group of them with certain characteristics such as immigrants and other minorities, but since the growing use of the internet and social networks, it is used frequently to hurt and express hatred towards these groups in society. Stereotypes are often used, especially in discussions of controversial issues such as immigration, sexism, and misogyny. Therefore, there is a need to have tools that are capable of determining when a user is employing sarcasm or irony to affect other people or groups of them, since the large amount of information that is generated daily makes manual control of it impossible.

The prediction of traits such as gender, age, occupation and origin of a person are topics that have been widely studied in the field of Author Profiling (AP). Recent research has been dedicated to detecting social behaviors and psychological characteristics of users by applying the techniques used in AP task. In previous works, the aim has been to identify users who

CLEF 2022 – Conference and Labs of the Evaluation Forum, September 5–9, 2022, Bologna, Italy
mfaherold@gmail.com (M. F. Artigas-Herold); daniel.castro3@udc.es (D. Castro-Castro)

propagate harmful content with different approaches. Most of the research presented in the state-of-the-art of the last three PAN AP tasks use a traditional approach to Machine Learning (ML) algorithms such as Support Vector Machine (SVM) and Logistic Regression (LR), with which the best results are obtained [1,2,4,6,14,17,19,24,25,27,28,29].

During the last 3 years, PAN has launched several AP tasks dedicated to identifying users on Twitter who spread harmful content, as well as being able to identify those profiles that constitute chat bots, due to their participation in this activity. Precisely the task of the year 2019 was dedicated to differentiating bots from human profiles [34] and in 2020 it was aimed at identifying those users who shared fake news on the network [33].

At PAN 2021, last year, the AP task proposed was: Profiling Hate Speech Spreaders on Twitter 2021 [32], whose main objective was to determine, from a set of 200 tweets per author, whether or not a user profile was a hate speech spreader. In addition, it had a multilingual approach in which it was intended to recognize hate speech in the English and Spanish languages. The baselines offered were the use of character n-gram models together with LR, word n-grams with SVM, and other models of Artificial Neural Networks (ANN) or Deep Learning (DL) such as Universal Sentence Encoder (USE) together with Long Short Term Memory (LSTM), Term frequency – Inverse document frequency (TF-IDF) with LSTM, among others. For each language, 200 authors were chosen and the data corpus was balanced with respect to their classes.

This year [36], the AP task was focused on determining ironic profiles on Twitter, paying special emphasis on those authors who use irony to spread stereotypes, for example, towards women or the collective LGBT. The goal of the task is to classify authors as ironic or not based on the number of tweets with ironic content. Therefore, given the Twitter authors along with their posts, the main objective will be to profile those authors that can be considered ironic. For this, the baselines offered are the use of n-grams of characters or words, together with ML classification algorithms such as SVM, LR, LDSE [30], among other models that may be used. Unlike the previous year, this time only texts in English were used, with a selection of 420 authors for training, distributed into 210 ironic and 210 non-ironic profiles, which represents a balanced data corpus. It also has 200 tweets per user. We have to remark that in the corpus all the urls, links, hashtags and user mentions that appear in the content were masked with a unique token for each type.

The work in [1] proposes an approach based on detecting users who spread hate on Twitter using n-grams of characters and words together with SVM as the main classifier, in addition to comparing its results with other classifiers such as: LR, Naive Bayes, Random Forest (RF), Stochastic Gradient Descent (SGD) classifier and K-Nearest Neighbor (KNN). In this case, the authors suggest that the reason for not using a DL approach, despite the fact that its methods offer good results, is the limited availability of samples in the PAN 2021 task. In the work of [2], it's proposed to make a majority vote of 4 different representations of the text (word tfidf, char tfidf, vader, and roberta's word embeddings) along with two classifiers (SVM and RF), which are still very strong text classifiers, even when compared to recent deep neural networks. The work of [28] obtained the best results for the 2019 task using a SVM classifier with character and word n-grams features. They choose to evaluate char and word n-grams with different n-gram orders and also opted to represent each document using TF-IDF.

Other approaches use DL models such as Convolutional Neural Network (CNN) and LSTM networks, also include attention mechanisms [3,6,7,10,12,13,22,23]. Among the DL models, the bidirectional encoder representations from Transformers (BERT) have been widely used. The authors of [10] describe a system which was trained on a corpus of English Twitter posts with a goal to predict whether or not the author of the given posts spreads hate speech. The features were crafted using fine-tuned BERT contextualized embeddings summed over the last 12 hidden states corresponding to the classification token, concatenated with the three binary variables called indicators. Binary variables were indicating whether a hashtag, retweet or url were present in the author's tweet posts, respectively. Feature vectors were then fed into a LR classifier. The approach of [12] is based on making a comparison between the traditional approach of ML with SVM and DL using Bi-LSTM, with which, when performing the experiments, slightly lower results were obtained than with the traditional approach.

Research was also presented aimed at profiling the authors based on the analysis of emotions and psychological characteristics in the text [5,9,15,16,18]. The authors of [5] tackled the PAN 2021 Hate Speech identification task through Semantic Emotion-based models in both Spanish and English languages. They implement several approaches, one of them designed to output explainable results based on the user's emotional charge.

This paper describes our participation in the PAN at CLEF 2022 task of AP: Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) 2022 [35], which focuses on profiling ironic authors on Twitter. Our proposal employs n-gram of characters and word-based models along with traditional classification algorithms such as RF and SVM, further including a representation based only on n-grams of those elements that do not correspond to correct words in the English language. In addition, the chi_square and f_classif algorithms were used to reduce features in the text of each profile. The main idea is to use the best classification models obtained in the training and perform a majority vote to determine whether or not a user profile on Twitter can be considered ironic.

2. Our approach

2.1. Method

Considering that in this case also few examples are available per author, we decided to use machine learning models instead of a DL approach, due to the good results that have been obtained with these models in previous tasks. This is because DL is typically data intensive, and with so little data available, it could easily turn into an overfit. Our proposal is based on making three different representations of the textual content, to build n-gram models and gather the best ones, and be able to obtain a majority vote that allows us to classify the different profiles. It is intended to start from a character and word n-gram representation, and also incorporate one based on elements that do not constitute correct words of the English language, called out-of-vocabulary (OOV) words [31]. Although these OOVs are, in theory, within the representation based on words n-grams, what we aimed to achieve by representing them independently is to capture expressions or characters that can denote irony and evaluate how effective they are in solving the classification task. In the development of this work, we are interested in investigating if these representations based on OOV would be enough to differentiate the user profiles that spread ironic content from those that do not.

For this task, we have used a combination of three different representations for the training

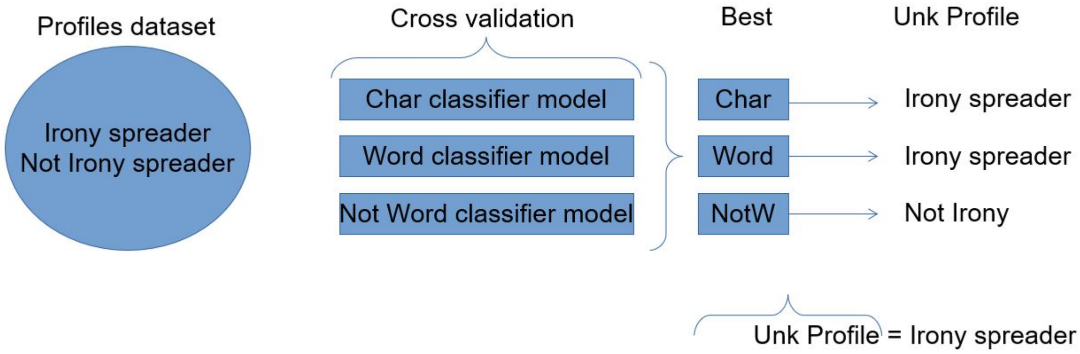

Figure 1: Diagram of the majority vote proposal.

dataset, as we can appreciate in Figure 1. Also, we have employed two traditional classifiers to build the ML models and two feature extraction algorithms to perform the different reductions that were made in the experiments. A 10-fold cross validation was developed, as can be observed in Figure 1. Finally, the models that obtained the best results for each representation were chosen to carry out a majority vote among all of them to determine the final classification of an unknown profile.

The machine learning models we chose were those of SVM and RF, as well as the algorithms used for feature extraction were chi_square[1] and f_classif[2], with which made several vocabulary reductions. The implementation used for all models and algorithms is the one available in the python library scikit-learn.

For our final proposal we use a combination of three strategies: represent the text of the tweets using n-grams of characters, n-grams of words and n-grams OOV words [31]. For all of them, different values of n were used in the n-grams together with several reductions of the vocabulary using the two mentioned algorithms. Finally, the model that obtains the best results in the 10-fold cross validation for each value of n is chosen and saved for later use in a majority vote.

2.2. Experiments

For the tokenization of the texts, we used the CountVectorizer of the python sklearn library that allowed us to represent the collection of documents from a vocabulary of known words. The hyper-parameters of the Vectorizer are the following: "parser", which determines the level at which feature extraction should be performed, either on n-grams of words or characters; "ngram_range" which determines the order of the language model to use; "lowercase" to convert all characters to lowercase before tokenization; "vocabulary" to build the objects from a given vocabulary if necessary.

In the first strategy, the 200 tweets of each author of the training set are joined as if they were a single text and are represented in a bag of words with the CountVectorizer used in the

[1]https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html
[2]https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html

Table 1
The best hyper-parameter models, Accuracy and F macro obtained by performing a 10-fold cross validation.

Model	Strategy	Reduction Algorimth / k	n-grams / w-grams	Accuracy	F Macro
SVM	1	f_classif / 500	n=2	0.9102	0.9100
SVM	1	chi_square / 2000	n=3	0.9682	0.9682
SVM	1	chi_square / 2000	n=4	0.9814	0.9814
SVM	1	chi_square / 2000	n=5	0.99867	0.9867
SVM	1	chi_square / 2000	n=6	0.9761	0.9760
SVM	1	chi_square / 2000	w=1	0.9761	0.9760
SVM	1	f_classif / 2000	w=2	0.9920	0.9920
SVM	1	chi_square / 2000	w=3	0.9894	0.9894
SVM	1	chi_square / 2000	w=4	0.9680	0.9679
SVM	1	f_classif / 2000	w=5	0.9555	0.9545
RF	2	f_classif / 200	w=1	0.8889	0.8873
RF	2	chi_square / 100	w=2	0.5132	0.3391

tokenization in n-grams of characters. The values of n used in the first method range from n=2 to n=6 and the value k for word reductions ranges from k=100 to k=2000. For each of the configurations of these parameters, the classifiers are trained and a 10-fold cross validation is performed with the Accuracy[3] and F macro[4] measures, from which the mean is calculated to obtain a final value of these two evaluation measures in every model. Once all Accuracy and F macro values have been calculated, the model with the best result for each n used is saved. Thus, in the end, the 6 best models are obtained for n-grams of characters in the used dataset.

In the second strategy, the profiles are built considering all the tweets of the same author as a single one and they are represented in a bag of words through the process of tokenization in n-grams of words. This, time only word unigrams and bigrams were used and the value of k for vocabulary reductions ranged from k=100 to k=2000. Next, the classifiers are trained and the same 10-fold cross validation process is carried out and the Accuracy and F macro are calculated to proceed to save the two best models.

In the third strategy, only those words that do not constitute correct words of the English language were chosen to represent the profiles, that is, OOV words: emoticons, word lengthenings, expressions, word shortenings, word mergings and proper nouns. These words were selected by using a python library for Aspell[5], a spell checker with dictionaries available in several languages. Before representing the profiles, a pre-processing work was carried out, where the 200 tweets were joined into one and the OOV words were chosen for each profile using the automatic english spell checker from Aspell library. Letters that were unnecessarily repeated in words that constituted expressions were eliminated, and two vocabularies were constructed, one for ironic and non-ironic classes. Once the vocabularies were built, all those common tokens for both were eliminated, thus leaving two sets of words without coincidences. To represent the profiles, the union of these two sets was given to the CountVectorizer as a

[3]https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html
[4]https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
[5]https://github.com/WojciechMula/aspell-python

vocabulary, which allowed each profile to be represented in bags of words based on the OOV words present in the text. In this strategy, only word unigrams and reductions from k=100 to k=2000 were used, and the same 10-fold cross validation was performed after training the classifiers. Once the Accuracy and F macro values were obtained, the model with the best results was saved.

Finally, experiments were performed with the test dataset provided by the organizers of the task in which all the previously saved models were used to get a majority vote from which we obtained a final classification for each author's profile.

If we take a look at Table 1, it can be seen that of the two classifiers used, the one that obtained the best results for almost all the orders of n-grams was the SVM in the experiments of the first strategy and in most cases with a reduction of 2000 features. This demonstrates that the profiles can be adequately represented and classified even with a large reduction in vocabulary from the text. The table shows the chosen models that achieved the best results for each order of the n-grams in the first two strategies described as performing a 10-fold cross validation. It can be seen that, in general, most of the models obtained for all values of n represent a high accuracy and F macro, above 90%. It is appreciable how the results of the two quality measures reach very similar values, since we are working on a binary classification problem and the classes are balanced.

In the case of the best models for the second strategy, the classifier that achieves the best results is RF. Due to the fact that during the experiments carried out with the 10-fold cross validation no differences were shown between the values of accuracy and F macro reached by the different orders of n, those with the greatest applied reduction were chosen. Also, we can appreciate the lowest results are obtained with the models built from the second strategy with the OOV, which indicates that a representation based only on these elements is not good enough to differentiate between profiles.

In the table, k means the number of features of the applied reduction, n in "n-grams" means n-grams of characters and w in "w-grams" means n-grams of words.

Finally, after carrying out the pertinent experiments, the results obtained with strategy 3 and the test corpus provided by the organizers for the task, which consisted of a majority vote of all the best models, were sent. The best Accuracy value obtained was 91.67%, which does not meet expectations but is still a good result.

3. Conclusions

In this paper we described our participation in the task Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) organized by PAN @ CLEF 2022. We use three different text representations along with two strong classifiers available in the literature, SVM and RF. Our majority vote system, which combines the use of the best n-grams models obtained in the first two strategies developed, achieves accuracy as high as 91.67% in the test corpus given by the task organizers, although better values were obtained in the training stage. From the experiments described we can conclude that the representation of the profiles using only those elements that consist in OOV words is not enough to classify them into ironic profiles or not. The entire content of the text should always be taken into account, since people use both words and OOV expressions, to denote sarcasm and irony. This is demonstrated by the results of the experiments during training where the values achieved by the OOV words were remarkably lower than for representations of words and characters n-grams of the entire text content.

In future work, it would be interesting to take those profiles labeled as Irony Spreader and eliminate from them those tweets that do not represent irony within the content of the profile, to use them in the classifier training process in order to check if better results are obtained.

4. Acknowledgments

This work was supported by projects PLEC2021-007662 (MCIN/AEI/10.13039/501100011033, Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, Plan de Recuperación, Transformación y Resiliencia, Unión Europea-Next Generation EU) and RTI2018-093336-B-C22 (Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación). The second author also thanks the financial support supplied by the Consellería de Cultura [**?**], Educación e Universidade (GPC ED431B 2022/33).

5. References

[1] Jorge Alcañiz and José Andrés. Profiling Hate Spreaders using word N-grams—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[2] Claudio Moisés Valiense de Andrade and Marcos André Gonçalves. Profiling Hate Speech Spreaders on Twitter: Exploiting Textual Analysis of Tweets and Combinations of Multiple Textual Representations. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[3] Christopher Bagdon. Profiling Spreaders of Hate Speech with N-grams and RoBERTa—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[4] Fazlourrahman Balouchzahi, Hosahalli Lakshmaiah Shashirekha, and Grigori Sidorov. HSSD: Hate Speech Spreader Detection using N-grams and Voting Classifier—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[5] Hiram Cabrera, Sabino Miranda-Jiménez, and Eric S. Tellez. INFOTEC-LaBD at PAN@CLEF21: Profiling Hate Speech Spreaders on Twitter through Emotion-based Representations—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[6] Ángel Andújar Carracedo and Raquel Jiménez Mondéjar. Profiling Hate Speech Spreaders on Twitter—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[7] Roberto Labadie Tamayo, Daniel Castro Castro, and Reynier Ortega Bueno. Deep Modeling of Latent Representations for Twitter Profiles on Hate Speech Spreaders Identification Task—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[8] Tanise Ceron and Camilla Casula. Exploiting Contextualized Word Representations to Profile Haters on Twitter—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[9] Riccardo Cervero. Use of Lexical and Psycho-Emotional Information to Detect Hate Speech Spreaders on Twitter—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021.

[10] David Dukić and Ana Sović Kržić. Detection of Hate Speech Spreaders with BERT—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[11] Darko Tosev and Sonja Gievska. Multi-level stacked ensemble learning for identifying hate speech spreaders on Twitter—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[12] Inna Vogel and Meghana Meghana. Profiling Hate Speech Spreaders on Twitter: SVM vs. Bi-LSTM—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[13] Oleg Bakhteev, Aleksandr Ogaltsov, and Petr Ostroukhov. Fake News Spreader Detection Using Neural Tweet Aggregation—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[14] Jakab Buda and Flora Bolonyai. An Ensemble Model Using N-grams and Statistical Featuresto Identify Fake News Spreaders on Twitter—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[15] Xinhuan Duan, Elham Naghizade, Damiano Spina, and Xiuzhen Zhang. RMIT at PAN-CLEF 2020: Profiling Fake News Spreaders on Twitter—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[16] María S. Espinosa, Roberto Centeno, and Álvaro Rodrigo. Analyzing User Profiles for Detection of Fake News Spreaders on Twitter—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[17] Daniel Yacob Espinosa, Helena Gómez-Adorno, and Grigori Sidorov. Profiling Fake News Spreaders using Characters and Words N-grams—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[18] Elisabetta Fersini, Justin Armanini, and Michael D'Intorni. Profiling Fake News Spreaders: Stylometry, Personality, Emotions and Embeddings—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[19] Ahmad Hashemi, Mohammad Reza Zarei, Mohammad Reza Moosavi, and Mohammad Taheri. Fake News Spreader Identification in Twitter using Ensemble Modeling—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[20] Clemens Hörtenhuemer and Eva Zangerle. A Multi-Aspect Classification Ensemble

Approach for Profiling Fake News Spreaders on Twitter—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[21] Catherine Ikae and Jacques Savoy. UniNE at PAN-CLEF 2020: Profiling Fake News Spreaders on Twitter—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[22] Soumayan Bandhu Majumder and Dipankar Das. Detecting Fake News Spreaders on Twitter Using Universal Sentence Encoder—Notebook for PAN at CLEF 2020. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[23] Tiziano Fagni and Maurizio Tesconi. Profiling Twitter Users Using Autogenerated Features Invariant to Data Distribution. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[24] Régis Goubin, Dorian Lefeuvre, Alaa Alhamzeh, Jelena Mitrović, Előd Egyed-Zsigmond, and Leopold Ghemmogne Fossi. Bots and Gender Profiling using a Multi-layer Architecture. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[25] Fredrik Johansson. Supervised Classification of Twitter Accounts Based on Textual Content of Tweets. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[26] Dijana Kosmajac and Vlado Keselj. Twitter User Profiling: Bot and Gender Identification. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[27] Asad Mahmood and Padmini Srinivasan. Twitter Bots and Gender Detection using Tf-idf. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[28] Juan Pizarro. Using N-grams to detect Bots on Twitter. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[29] Mahendrakar Srinivasarao and Siddharth Manu. Bots and Gender Profiling using Character and Word N-Grams. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[30] F. Rangel, P. Rosso, M. Franco-Salvador, A Low Dimensionality Representation for Language Variety Identification, in: In 17th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing'16. Springer-Verlag, LNCS(9624), 2018, pp. 156–169.

[31] Maity, Suman & Sarda, Chaitanya & Chaudhary, Anshit & Patil, Abhijeet & Kumar, Shraman & Mondal, Akash & Mukherjee, Animesh. (2016). WASSUP? LOL : Characterizing Out-of-Vocabulary Words in Twitter. Cornell University, arxiv.org

[32] Francisco Rangel, Paolo Rosso, Gretel Liz De La Peña Sarracén, Elisabetta Fersini, and BERTa Chulvi. Profiling Hate Speech Spreaders on Twitter Task at PAN 2021. In Guglielmo Faggioli, Nicola Ferro, Alexis Joly, Maria Maistro, Florina Piroi, editors, CLEF 2021 Labs and Workshops, Notebook Papers, 2021. CEUR-WS.org.

[33] Francisco Rangel, Anastasia Giachanou, Bilal Ghanem, and Paolo Rosso. Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. In Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, September 2020. CEUR-WS.org.

[34] Francisco Rangel and Paolo Rosso. Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling. In Linda Cappellato, Nicola Ferro, David E. Losada, and Henning Müller, editors, CLEF 2019 Labs and Workshops, Notebook Papers, September 2019. CEUR-WS.org.

[35] Ortega-Bueno R., Chulvi B., Rangel F., Rosso P. and Fersini E. Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) at PAN 2022. In: CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org

[36] Janek Bevendorff and Berta Chulvi and Elisabetta Fersini and Annina Heini and Mike Kestemont and Krzysztof Kredens and Maximilian Mayerl and Reyner Ortega-Bueno and Piotr Pezik and Martin Potthast and Francisco Rangel and Paolo Rosso and Efstathios Stamatatos and Benno Stein and Matti Wiegmann and Magdalena Wolska and Eva Zangerle: Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection. In: Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022)