

# Irony and Stereotype Spreaders Detection using BERT-large and AutoGulon

Yuning Zhang, Hui Ning\*

Harbin Engineering University (HEU), 145 Nantong Street, Nangang District, Harbin, China

## Abstract

With the continuous development of the Internet, the Internet has become the mainstream way for people to socialize, and there is more and more content on the Internet. However, with the development of social networks comes the emergence of many Irony and stereotyped remarks, making the need for an automatic detection system more urgent. This paper provides a solution to the "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" task proposed by PAN CLEF 2022, using BERT-large and AutoGulon to process and predict the data, and the final submitted score is 94.44%.

## Keywords

Irony detection, Twitter, BERT-large, AutoGulon

## 1. Introduction

Online social media plays a vital role in People's Daily life. With the development of the Internet and the improvement of corresponding functions, the proportion of online social media in People's Daily life will increase, and more people will use the Internet to socialize. People can communicate freely on Twitter, which has led to a series of Ironic, stereotypical comments, often directed at women or LGTB people. Due to fast transmission, anonymity and easy access to online media [1, 2, 3], such improper remarks are even more rampant. These inappropriate statements, spread by large numbers of people and spread quickly, are impractical to identify and approve manually. So it makes sense to identify these inappropriate comments automatically. This paper solves the task of "Profiling Irony and Stereotype Spreaders on Twitter" [4] published by Pan in 2022 [5], implements an algorithm to identify sarcastic and stereotype remarks, and is submitted on TIRA [6]. This task extracts text features through BERT-large text embedding and then uses AutoGulon to predict the model and obtain experimental results.

## 2. Related Works

Successive PAN at CLEF has published similar classification algorithm tasks. Some of them are PAN 2018: Multimodal Gender Identification In Twitter [7], PAN 2019: Bots and Gender Profiling in Twitter [8], PAN 2020: Profiling Fake News Spreaders on Twitter [9] and PAN 2021: Profiling Hate Speech Spreaders on Twitter [10]. In last year's task, Uzan et al. used classic machine

---


\*Corresponding author

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

✉ pigeon\_zyn@163.com (Y. Zhang); ninghui@hrbeu.edu.cn (H. Ning)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

learning methods like Support Vector Classifier, Multi-Layer Perceptron, Logistic Regression, Random Forest, Ada-Boost Classifier and K-Neighbors Classifier to more recent deep learning methods like BERT and Bidirectional LSTM [11]. In addition, many people are also trying to use different approaches to identify hate and ironic speech. Salminen et al. experimented with several classification algorithms (Logistic Regression, Naive Bayes, Support Vector Machines, XGBoost, and Neural Networks) and feature representations (Bag-of-Words, TF-IDF, Word2Vec, BERT, and their combination) [12]. Gonzalez et al. describe a model for irony detection based on the contextualization of pre-trained Twitter word embeddings utilizing the Transformer architecture [13].

### 3. Methodology

This paper presents an automated machine learning (AutoML) tool, AutoGluon, submitted to the task "Profiling Irony and Stereotype Spreaders on Twitter". This task can be viewed as a binary text categorization problem, categorizing Twitter users as "IROSTEREO spreaders" or "non- IROSTEREO spreaders" based on their tweets.

#### 3.1. Corpus

The task's corpus consists of 420 XML files corresponding to the author. Each file contains 200 tweets from an author. In addition to tweet content, XML includes tags.

#### 3.2. Pre-processing

Firstly, we preprocess the text to improve the accuracy of prediction. For example, we were removing 'URL' and 'USER', unifying the case of files, converting emojis to corresponding characters, and so on. The specific pretreatment work is shown in the following table.

**Table 1**  
The Method of Pre-processing

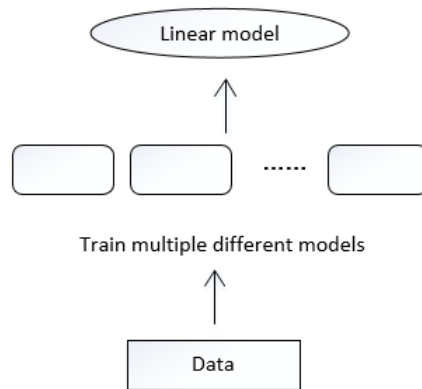
Item	Processing method
'URL' and 'USER'	Eliminate
Text case	Converts all text to lowercase
Emoji	Convert to the corresponding text
Sentence abbreviations	Convert sentence abbreviations to extended mode
Duplicate words	Delete and simplify

#### 3.3. Data prediction by AutoGluon

AutoGluon is a robust and accurate automated machine learning (AutoML) tool for structured data [14], developed by Amazon. Its purpose is to extract features from input as far as possible without human help, select suitable machine learning models and train them. There are several frameworks for Automl, most of which are based on hyperparametric search technology, which

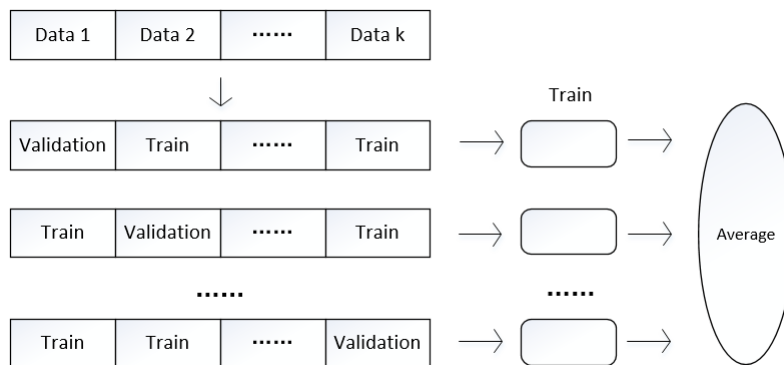
is to select a good model from dozens or hundreds of hyperparametric candidates in the hope of avoiding manual tuning. Autogluon, on the other hand, wants to avoid searching for hyperparameters so that multiple different models can be trained at the same time. Train multiple models without hyperparametric search and combine them to achieve better results than using hyperparametric search.

Autogluon uses three techniques to achieve this effect. The first is stacking, training multiple different models such as KNN, tree model or complex neural network on the same data set. The outputs of these models are then entered into a linear model to obtain the final output, which is the weighted sum of the outputs, where the weights are obtained by training.



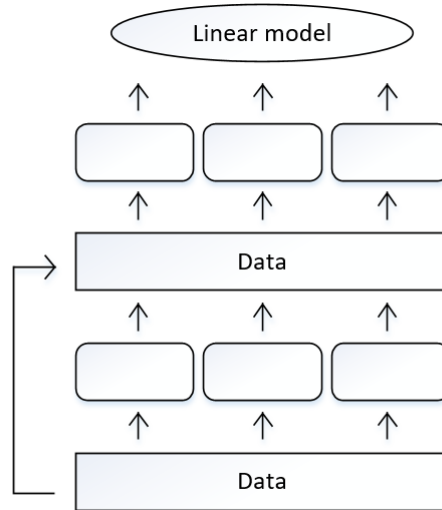
**Figure 1:** The schematic diagram of the stacking

The second is repeated k-fold bagging. Bagging is the training of multiple models of the same class, which may use multiple weights and data blocks with different initial values, and finally averages the output of these models to reduce the variance of the prediction. The k-fold bagging stem from the k-fold cross validation.



**Figure 2:** The schematic diagram of the k-fold bagging

The last is multi-layer stack ensembling. Multi-layer stack ensembling combines the outputs and data of multiple models for another stacking. That is, train multiple models on top of it, and then use a linear model to get an output. In order to prevent overfitting, multi-layer stack ensembling is often used in conjunction with repeated k-fold bagging.



**Figure 3:** The schematic diagram of the multi-layer stack ensembling

We extract embeddings from the last hidden layer of the BERT model. We then average these Twitter-based features down to the user level. Finally, these features are fed to AutoGluon tabular predictor for classification. Moreover, to avoid overfitting and underfitting, we use 5-fold cross-validation. For each classification feature, AutoGluon uses a separate embedding layer, and the dimension of the embedding is proportional to the number of layers observed for the feature [15]. The analysis process includes Neural networkANN, LightGBM boosted tree [16], CatBoost boosted tree [17], random forest (RF), extremely randomized tree (ExtRa Trees) and k-nearest neighbors (KNN). The embedded classification features and numerical features are connected in series into the three-layer feedforward network and directly connected with the output prediction.

## 4. Results

As shown in Table 2, We used AutoGulon to make five predictions, and there were some differences in the prediction results obtained using different machine learning algorithms. Then we selected the group with the best experimental results to submit. The table shows that the prediction results are relatively accurate, with the highest accuracy of train set being 95.238%. And the accuracy of test set is 94.444%.

**Table 2**  
Results

Number	Model	Average accuracy(%)
1	LightGBMXT RandomForestGini Neuralnettorch LightGBMXT LightGBMLarge	94.048
2	RandomForestEntr NeuralnetFastAI LightGBM LightGBM Neuralnettorch	95.238
3	Neuralnettorch RandomForestGini NeuralnetFastAI LightGBMXT ExtraTreesEntr	95.476
4	NeuralnetFastAI RandomForestGini LightGBM ExtraTreesGini LightGBMXT	94.524
5	LightGBMXT RandomForestGini ExtraTreesGini Neuralnettorch Neuralnettorch	95.238

## 5. Conclusions

In this paper, we describe our participation in the task "Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO)" organized by PAN @ CLEF 2022 and detail the process of completing the task. The whole experiment preprocessed the text and embedded it with BERT-large. Finally, AutoGluon was used to predict the model. We can see that the accuracy of the final experiment reached 94.44%. In the future, we will try more NLP and text categorization tasks, using different methods to achieve the best results.

## 6. Acknowledgments

Whenever we have problems, we can get timely help from the organizers. Many thanks to the organizers and reviewers for their guidance and support.

## References

- [1] M. Khan, A. Abbas, A. Rehman, R. Nawaz, Hateclassify: A service framework for hate speech identification on social media, *IEEE Internet Computing PP* (2020) 1–1.
- [2] A. T. Martini, M. Farrukh, H. Ge, Recognition of ironic sentences in twitter using attention-based lstm, *International Journal of Advanced Computer Science and Applications* (2018).
- [3] A. Natalie, Hate speech on social media networks: towards a regulatory framework?, *Information Communications Technology Law* (2018) 1–17.
- [4] O.-B. Reynier, C. Berta, R. Francisco, R. Paolo, F. Elisabetta, Profiling Irony and Stereotype Spreaders on Twitter (IROSTEREO) at PAN 2022, in: *CLEF 2022 Labs and Workshops, Notebook Papers*, CEUR-WS.org, 2022.
- [5] J. Bevendorff, B. Chulvi, E. Fersini, A. Heini, M. Kestemont, K. Kredens, M. Mayerl, R. Ortega-Bueno, P. Pezik, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, E. Zangerle, Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection, in: M. D. E. F. S. C. M. G. P. A. H. M. P. G. F. N. F. Alberto Barron-Cedeno, Giovanni Da San Martino (Ed.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022)*, volume 13390 of *Lecture Notes in Computer Science*, Springer, 2022.
- [6] M. Potthast, T. Gollub, M. Wiegmann, B. Stein, TIRA Integrated Research Architecture, in: N. Ferro, C. Peters (Eds.), *Information Retrieval Evaluation in a Changing World, The Information Retrieval Series*, Springer, Berlin Heidelberg New York, 2019. doi:10.1007/978-3-030-22948-1\_5.
- [7] E. Stamatatos, F. Rangel, M. Tschuggnall, B. Stein, M. Potthast, Overview of PAN 2018: 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, *Proceedings, Experimental IR Meets Multilinguality, Multimodality, and Interaction*, 2018.
- [8] W. Daelemans, M. Kestemont, E. Manjavacas, M. Potthast, F. Rangel, P. Rosso, G. Specht, E. Stamatatos, B. Stein, M. a. Tschuggnall, Overview of pan 2019: Bots and gender profiling, celebrity profiling, cross-domain authorship attribution and style change detection, Springer, Cham (2019).
- [9] J. Bevendorff, B. Ghanem, A. Giachanou, M. Kestemont, E. Zangerle, Overview of pan 2020: Authorship verification, celebrity profiling, profiling fake news spreaders on twitter, and style change detection, Springer, Cham (2020).
- [10] Overview of pan 2021: Authorship verification, profiling hate speech spreaders on twitter, and style change detection, in: *European Conference on Information Retrieval*, 2021.
- [11] M. Uzan, Y. Hacoheh-Kerner, Detecting hate speech spreaders on twitter using lstm and bert in english and spanish - notebook for pan at clef 2021 keywords, in: *CLEF 2021 – Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings (CEUR-WS.org)*, 2021.
- [12] J. Salminen, M. Hopf, S. A. Chowdhury, S. G. Jung, B. J. Jansen, Developing an online hate classifier for multiple social media platforms, *Human-centric Computing and Information Sciences* 10 (2020) 1.
- [13] J. González, L. F. Hurtado, F. Pla, Transformer based contextualization of pre-trained

word embeddings for irony detection in twitter - sciencedirect, Information Processing Management 57 (2020).

- [14] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, A. Smola, Autogluon-tabular: Robust and accurate automl for structured data (2020).
- [15] C. Guo, F. Berkhahn, Entity embeddings of categorical variables (2016).
- [16] M. Qi, Lightgbm: A highly efficient gradient boosting decision tree, in: Neural Information Processing Systems, 2017.
- [17] A. V. Dorogush, V. Ershov, A. Gulin, Catboost: gradient boosting with categorical features support (2018).