Interns@LT-EDI : Detecting Signs of Depression from Social Media Text

L. Koushik[1], Anand Kumar M[2], Hariharan R L[3]
Department of Information Technology,
National Institute of Technology Karnataka, Surathkal
koushik.201it131@nitk.edu.in[1],
m_anandkumar@nitk.edu.in[2],
hariharanrl.197it003@nitk.edu.in[3]

Abstract

In this paper we show our approach to solve the Shared Task on Detecting Signs of Depression from Social Media Text at LT-EDI@RANLP 2023. The given task is to classify the Reddit posts present in the dataset provided, into 3 levels of depression: 'not depression', 'moderate' and 'severe'. We have attempted classifying the posts using two models. We have explored multiple models for this task. Three of which will be included in this paper. The first model uses sentiment labels automatically extracted using TextBlob with TF-IDF for feature extraction and support vector machines (SVMs) for classification. For the second model, we leverage a convolutional neural network architecture for feature extraction and classification. Lastly, the third model incorporates a Bi-LSTM architecture with GloVe embeddings for feature extraction and classification. All the above models also used SMOTE for oversampling the dataset. Through our experimentation, we aim to evaluate the effectiveness of these models in accurately identifying signs of depression in social media text.

1 Introduction

Depression is one of the most severe mental-health diseases right now and it is important to detect the signs early on and take actions to stop troubled individuals from taking their own life and provide them with the help they need. Most methods we rely on today are medical procedures in clinics or actual human interaction which is highly ineffective leading to the high rates of suicide and a significant percentage of individuals affected by depression. According to the World Health Organization(WHO) almost 280 million people in the world suffer from depression and on an average about 703,000 people take their lives around the world as on 2022. However with the huge online presence in various social media platforms and individuals giving daily updates through social media posts it would make detecting depression very easy if we had a model that uses the social media texts and classifies it into multiple levels of depression. By using natural language processing, machine learning, and data mining techniques, researchers have made significant strides in automating the detection of depression in social media data. These include traditional machine learning algorithms like SVMs, Random Forest or Naive Bayes algorithms in the past and advanced Deep Learning algorithms in the recent times like Recurrent Neural Networks(RNN), Convolutional Neural Networks(CNN), Long Short-Term Memory (LSTM) and Generative Adversarial Networks (GAN) to name a few (William and Suhartono, 2021). In this paper we present our model for detecting depression for the competition **Shared Task on Detecting Signs of Depression from Social Media Text** at **LT-EDI@RANLP 2023**. This paper has 6 sections. Section 2 contains a short summary of the related works on depression detection using social media text referred to by the authors. Section 3 contains the details of the datasets used and any pre-processing done on them. Section 4 is on the methodology of our model. Section 5 shows the experiments done and the results obtained. Section 6 gives the conclusion of the paper and future work.

2 Related Works

Most of the work right now is focused on detecting depression from social media texts. One of the very first papers on depression detection in social media texts was written by De Choudhury et al.(2013), where the authors used a statistical classifier to estimate the level of depression. Chatterjee et al.(2019) used Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). Kim et al.(2020) used Convolutional Neural Networks (CNN) and XGBoost in their model of text classifier in an attempt to use ensemble learning. Amanat et al.(2022) created a hybrid model using Long Short-Term Memory (LSTM) and Re-

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 262–265, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_040

PID	Text data	Label
$train\_pid_1$	My life gets worse every year : That's what it feels like anyway....	moderate
$train\_pid_2$	Words can't describe how bad I feel right now : I just want to fall asleep forever.	severe
$train\_pid_3$	Is anybody else hoping the Coronavirus shuts everybody down?	not depression

Table 1: The data files are in Comma Separated Values (csv) format with three columns namely PID, Text data and Label. The above table shows a sample for each of the labels.

current Neural Networks (RNN). (Poświata and Perełkiewicz, 2022) presented the winning solution for Shared Task on Detecting Signs of Depression from Social Media Text at LT-EDI@ACL 2022. In their paper they have prepared a new pre-trained language model, DepRoBERTa. They have presented three models for the competition: RoBERTa-large, DepRoBERTa and ensemble model. They achieved a macro-averaged F1-score of 0.583. For feature extraction many works have made use of the Linguistic Inquiry and Word Count (LIWC) and models based on contextual word embeddings (Tadesse et al., 2019). In recent times more emphasis has been on deep learning and large pre-trained transformer-based language models. Deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown more promise in detecting depression from social media text (Aswathy et al., 2019). These techniques are able to learn complex patterns in language, which can help them to distinguish between genuine expressions of depression and other types of emotional expression. However it is important to note that this depends on the quality and the quantity of the data we have with us. The quality of datasets impact the accuracy of the models a lot which makes the balancing of datasets crucial. Chawla et al.(2002) introduces SMOTE, a method for oversampling the minority class in imbalanced datasets. The paper evaluated SMOTE on a variety of imbalanced datasets and found that it was able to improve the performance of classifiers on these datasets.

3 Dataset

The datasets provided for the competition includes Train data, Development data and Test data (Sampath et al.). Train data and Development data contain reddit posts in English language and each post is annotated with one of the labels: ‘not depression’, ‘moderate’ and ‘severe’. The first label indicates absence of any sign of depression in the post. The second label indicates the presence of some signs of depression and the third one indicates severe depression. The distribution indicates an imbalance among the classes, with the ’moderate’ label being the most prevalent, followed by ’not depression’, and ’severe’ being the least represented category. Class imbalance can pose challenges during model training, as the minority class (in this case, ’severe’) may have fewer samples to learn from, potentially resulting in biased predictions. To address this class imbalance we used SMOTE oversampling for Training dataset on the class ’severe’. SMOTE works by creating synthetic minority class examples that are located between existing minority class examples. This is done by randomly selecting a minority class example and then selecting one of its k nearest neighbors. A new synthetic example is then created at a point along the line connecting the two selected examples. To understand the datasets better and prepare the data for analysis, several preprocessing steps were undertaken. Firstly, duplicates were removed from both the Training and Development datasets to ensure the uniqueness of the samples. Additionally, the word ’[removed]’ was eliminated from the datasets as it indicated that certain posts were removed by Reddit moderators for various reasons, such as containing inappropriate or banned content. We will be using the Development dataset as Test dataset for the sake of the paper to compare the models on the basis of F1-score which won't be possible for test data as the posts are not labelled.

Table 2: Label Distribution in the Dataset

Labels	Number	Percentage
Not Depression	2755	38
Moderate	3678	51
Severe	768	11

4 Methodology

For depression detection from text a few steps are involved. This includes data preprocessing which involves cleaning and preprocessing the collected data, feature extraction which involves identifying the most informative features that are likely to be associated with depression and model development which involves choosing an appropriate algorithm to develop a predictive model. For data preprocessing we are cleaning the data and using SMOTE to balance the dataset.

4.1 SVM model with TextBlob(TF-IDF)

In this model we are using a LinearSVM model for classification, TextBlob for sentiment analysis and TfidfVectorizer is used to convert the pre-processed text data into a numerical representation using TF-IDF (Term Frequency-Inverse Document Frequency). The sentiment scores are calculated using the polarity and subjectivity scores of the text. The sentiment scores and the TF-IDF representations are combined by concatenating them together to form a single feature vector for each text data point. This combined feature vector is then used as input to the model. All the models were downloaded from the Python Package Index. The results of the model will be discussed in section 5.

4.2 CNN model with TextBlob(TF-IDF)

In this model, we will utilize a convolutional neural network (CNN) for classification. TextBlob will be used for sentiment analysis, and TF-IDF (Term Frequency-Inverse Document Frequency) will be employed to convert the pre-processed text data into a numerical representation. The sentiment scores are calculated using the polarity and subjectivity scores of the text. The sentiment scores and the TF-IDF representations are combined by concatenating them together to form a single feature vector for each text data point. This combined feature vector is then used as input to the model. The Python Package Index will be used to download the necessary models. The results of this model will be discussed in section 5.

4.3 Bi-LSTM model with TextBlob and GloVe embeddings

In this model, we utilize a bidirectional long short-term memory (Bi-LSTM) model for classification, TextBlob for sentiment analysis, and GloVe embeddings for word representation. The input text data is tokenized using TextBlob, and GloVe embeddings are employed to convert the tokens into numerical representations. This model uses two input layers: an embedding layer uses the pre-trained GloVe embeddings as initial weights and a Bi-LSTM layer. The Bi-LSTM architecture is capable of capturing contextual information from the text data by processing it in both forward and backward directions. The results of this model will also be discussed in section 5.

4.4 Experimentation

In addition to the models mentioned above, we conducted further experimentation to explore different approaches and techniques for detecting signs of depression from social media text. We recognized the importance of considering alternative methods and evaluating their performance to find out the best model for the given dataset.

During our experimentation, we explored the utilization of Word2Vec embeddings in combination with other machine learning algorithms. We observed that the SVM model with TextBlob (TF-IDF) outperformed the other variations. This finding indicates that the combination of SVM, TextBlob sentiment analysis, and TF-IDF representation effectively captures the necessary information from social media text to detect signs of depression.

5 Results and Dicussion

Table-3 shows the results of the three models. The SVM model with TextBlob (TF-IDF) was the best in terms of accuracy (0.603) and F1-score (0.479). Following that was the CNN model with TextBlob (TF-IDF) with an accuracy of 0.584 and F1-score of 0.442. Finally the model based on Bi-LSTM model with TextBlob and GloVe embeddings had an accuracy of 0.47 and F1-score of 0.34. The macro F1-score for each of them were 0.48, 0.44 and 0.34 respectively. The main challenge faced in this shared task was the imbalanced dataset. The reason for the better performance of the SVM model could be because of the simplicity of the problem. To compare the models the metrics used during the experiments are accuracy, precision, recall and macro F1-score and weighted F1-score for all the classes. The macro F1-score is the main metric to evaluate the models.

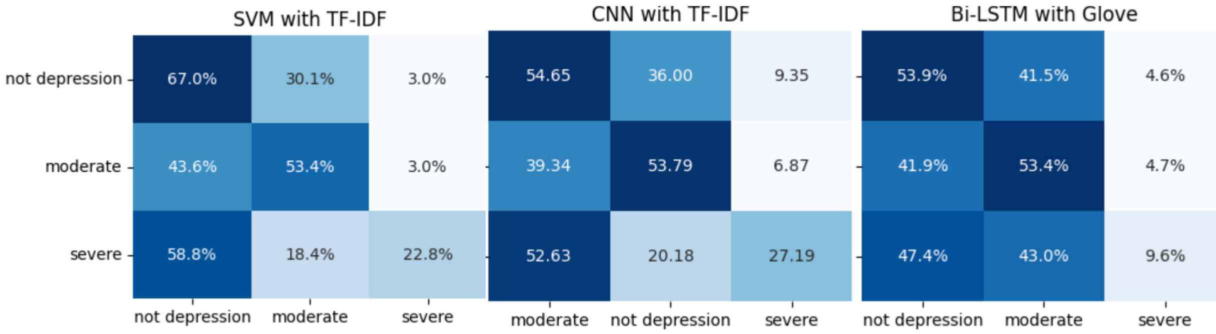

Figure 1: Confusion matrices for the the three models

Model	Accuracy	F1 score
SVM model with TextBlob(TF-IDF)	0.603	0.479
CNN model with TextBlob(TF-IDF)	0.584	0.442
Bi-LSTM model with TextBlob and GloVe embeddings	0.47	0.34

Table 3: Results of each model on the development set.

6 Conclusion and Future Works

In this paper, we presented three solutions to the Shared Task on Detecting Signs of Depression from Social Media Text at LT-EDI@RANLP 2023. The SVM model with TextBlob for feature extraction proved to be the best one among the three and the CNN model with a close second. For future works we plan to implement models better suited for both feature extraction and classification and use a better split training data and testing data. Also the accuracy could be improved by balancing the dataset better by using other methods.

References

Amna Amanat, Muhammad Rizwan, Abdul Rehman Javed, Maha Abdelhaq, Raed Alsaqour, Sharnil Pandya, and Mueen Uddin. Deep learning for depression detection from textual data. *Electronics*, 11 (5):676, 2022.

KS Aswathy, PC Rafeeque, and Reena Murali. Deep learning approach for the detection of depression in twitter. In *Proceedings of the International Conference on Systems, Energy Environment (ICSEE)*, 2019.

Ankush Chatterjee, Umang Gupta, Manoj Kumar Chinnakotla, Radhakrishnan Srikanth, Michel Galley, and Puneet Agrawal. Understanding emotions in text using deep learning and big data. *Computers in Human Behavior*, 93:309–317, 2019.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 128–137, 2013.

Jina Kim, Jieon Lee, Eunil Park, and Jinyoung Han. A deep learning model for detecting mental illness from user content on social media. *Scientific reports*, 10 (1):1–6, 2020.

Rafał Poświata and Michał Perełkiewicz. Opi@ lt-edi-acl2022: Detecting signs of depression from social media text using roberta pre-trained language models. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 276–282, 2022.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and booktitle = "Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion" month = may year = Mahibha C, Jerin ". Findings of the shared task on Detecting Signs of Depression from Social Media.

Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7:44883–44893, 2019.

David William and Derwin Suhartono. Text-based depression detection on social media posts: A systematic literature review. *Procedia Computer Science*, 179:582–589, 2021.