WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fake News Detection Using Machine Learning and BERT Model

Pranali L. Kakad[1], Pooja S. Jagtap[2], Sanskruti S. Mokashi[3], Rutuja V. Sakat[4]

[1]*Assistant Professor Computer Engineering PVPIT, Bavdhan, Pune-411021, India*
[2] *B.E. Computer Engineering PVPIT, Bavdhan, Pune-411021, India*
[3]*B.E. Computer Engineering PVPIT, Bavdhan, Pune-411021, India*
[4]*B.E. Computer Engineering PVPIT, Bavdhan, Pune-411021, India*

https://doi.org/10.5281/zenodo.7978522

ABSTRACT

In the modern era of computing, the news ecosystem has transformed from old traditional print media to social media outlets. Social media platforms allow us to consume news much faster, with less restricted editing results in the spread of fake news at an incredible pace and scale. In recent research, many useful methods for fake news detection employ sequential neural networks to encode news content and social context-level information where the text sequence was analysed in a unidirectional way. Two new fake news datasets are introduced, one obtained through crowdsourcing and covering six news domains, and another one obtained from the web covering celebrities. Our best-performing models achieved accuracies that are comparable to human ability to spot fake content. A bidirectional training approach is a priority for modelling the relevant information of fake news that is capable of improving the classification performance with the ability to capture semantic and long-distance dependencies in sentences. In this paper, we propose a BERT-based (Bidirectional Encoder Representations from Transformers) deep learning approach (FakeBERT) by combining different parallel blocks of the single-layer deep Convolutional Neural Network (CNN) having different kernel sizes and filters with the BERT.

Keywords: Machine Learning, BERT (Bidirectional Encoder Representations from Transformers), NLP (Natural Language Processing), FakeBERT.

1. Introduction: Fake News as a real-world problem

Due to easy access, rapid growth, and proliferation of the information available through regular news mediums or social media, it is becoming easy for people to look for news and consume it. But on the other hand, it is becoming a daunting task to differentiate between false information and true information thus leading to widespread fake news. We can define fake news as a type of deceiving journalism and statements that are used to artifice and mislead people.

Motivation

The situation is dire because many people believe anything they read on the internet and the ones who are amateur or are new to digital technology may be easily fooled. A similar problem is fraud which may happen due to spam or malicious emails and messages. Only then one can look into the different techniques and fields of machine learning (ML), natural language processing (NLP), and artificial intelligence (AI) that could help us fight this situation. “Fake news” has been used in a multitude of ways in the last half a year and multiple definitions have been given.

Problem Definition

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with a supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis.

- Understand the approach (i.e., concepts like BERT, analysis, data mining and, Regression,)
- Broaden our understanding of the chosen domain.
- Understand the actual pros and cons behind every step

2. Introduction to the proposed architecture

In the paper, we will be using Exploratory Data Analysis to understand the dataset and to derive the meaning behind all aspects and features. After initial understanding is gained using exploratory data analysis, we shall be implementing Clustering Analysis as our base algorithm. We can give our chosen features from the CSV dataset as an entry to the system. After taking the input the algorithms apply to that input which is K-Means Clustering. After the accessing data set the operation is performed and appropriate recommendations are produced. The proposed system will add more parameters significant to appropriate accommodations with their individual Point-Of-Interests in mind like, Budget and Distance from the desired location in accord with the priority levels. The accommodation recommendation system is designed to help provide appropriate suggestions related to accommodations

Proposed Method

We propose a BERT-based deep learning approach (FakeBERT) by combining different parallel blocks of the single-layer CNNs with the Bidirectional Encoder Representations from Transformers (BERT). We utilize BERT as a sentence encoder, which can accurately get the context representation of a sentence. This work is in contrast to previous research works where researchers looked at a text sequence in a unidirectional way (either left to right or right to left for pre-training). Many existing and useful methods had been presented with sequential neural networks to encode the relevant information. We use the following models-

1. Logistic Regression
2. Decision Tree
3. Gradient Booster
4. Random Forests
5. BERT(FakeBERT)

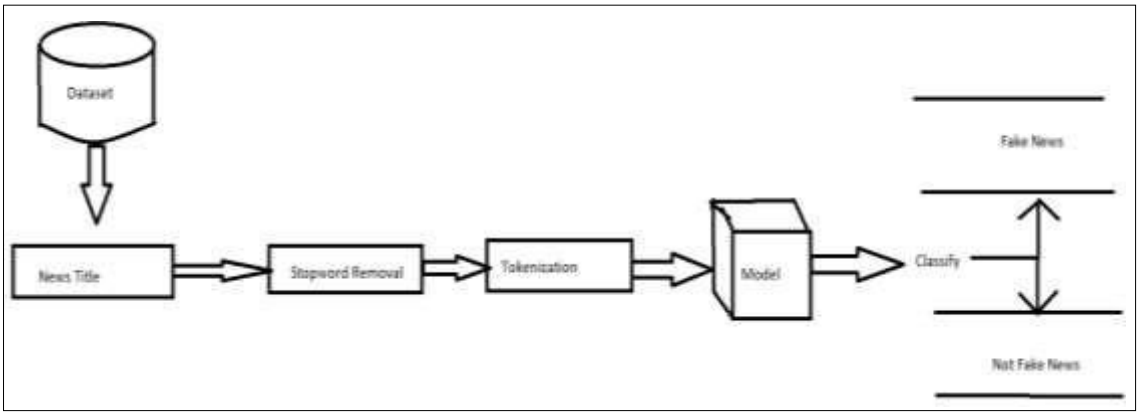

Fig. 2.1.1 Word Flow Diagram

Design Architecture and Implementation

BERT

BERT is an advanced pre-trained word embedding model based on transformer-encoded architecture. We utilize BERT as a sentence encoder, which can accurately get the context representation of a sentence. BERT removes the unidirectional constraint using a mask language model (MLM). It randomly masks some of the tokens from the input and predicts the original vocabulary id of the masked word based only. MLM has increased the capability of BERT to outperform as compared to previous embedding methods. It is a deeply bidirectional system that is capable of handling the unlabeled text by joint conditioning on both the left and right context in all layers.

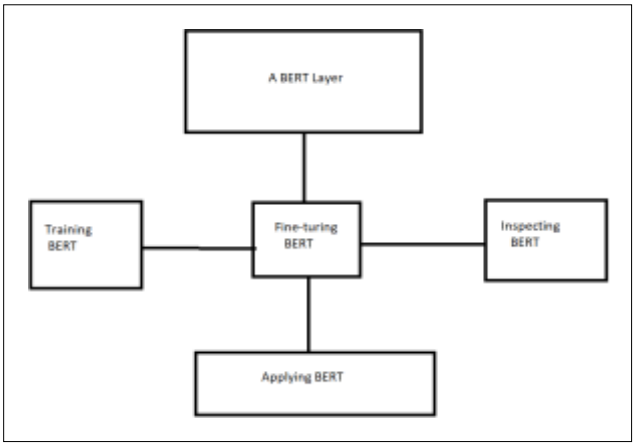

Fig.2.2.1 Conditioning for BERT model

3. Architecture

In our proposed framework, as illustrated in Figure 1, we are expanding on the current literature by introducing ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The ensemble techniques along with Linguistic Inquiry and Word Count (LIWC) feature set used in this research are the novelty of our proposed approach. There are numerous reputed websites that post legitimate news contents, and a few other websites such as PolitiFact and Snopes which are used for fact checking. In addition, there are open repositories which are maintained by researchers [11] to keep an up-to-date list of currently available datasets and hyperlinks to potential fact checking sites that may help in countering false news spread. However, we selected three datasets for our experiments which contain news from multiple domains (such as politics, entertainment, technology, and sports) and contain a mix of both truthful and fake articles. The datasets are available online and are extracted from the World Wide Web.

The corpus collected from the World Wide Web is preprocessed before being used as an input for training the models. The articles' unwanted variables such as authors, date posted, URL, and category are filtered out. Articles with no body text or having less than 20 words in the article body are also removed. Multicolumn articles are transformed into single column articles for uniformity of format and structure. These operations are performed on all the datasets to achieve consistency of format and structure.

Once the relevant attributes are selected after the data cleaning and exploration phase, the next step involves extraction of the linguistic features. Linguistic features involved certain textual characteristics converted into a numerical form such that they can be used as an input for the training models. These features include percentage of words implying positive or negative emotions; percentage of stop words; punctuation; function words; informal language; and percentage of certain grammar used in sentences such as adjectives, preposition, and verbs. To accomplish the extraction of features from the corpus, we used the LIWC2015 tool which classifies the text into different discrete and continuous variables, some of which are mentioned above. LIWC tool extracts 93 different features from any given text. As all of the features extracted using the tool are numerical values, no encoding is required for categorical variables. However, scaling is employed to ensure that various feature's values lie in the range of (0, 1). This is necessary as some values are in the range of 0 to 100 (such as percentage values), whereas other values have arbitrary range (such as word counts). The input features are then used to train the different machine learning models. Each dataset is divided into training and testing set with a 70/30 split, respectively. The articles are shuffled to ensure a fair allocation of fake and true articles in training and tests instances. The learning algorithms are trained with different hyperparameters to achieve maximum accuracy for a given dataset, with an optimal balance between variance and bias. Each model is trained multiple times with a set of different parameters using a grid search to optimize the model for the best outcome. Using a grid search to find the best parameters is computationally expensive. However, the measure is taken to ensure the models do not overfit or underfit the data.

Various ensemble techniques such as bagging, boosting, and voting classifier are explored to evaluate the performance over the multiple datasets. We used two different voting classifiers composed of three learning models: the first voting classifier is an ensemble of logistic regression, random forest, and KNN, whereas the second voting classifier consists of logistic regression, linear SVM, and classification and regression trees (CART). The criteria used for training the voting classifiers is to train individual models with the best parameters and then test the model based on the selection of the output label on the basis of major votes by all three models. We have trained a bagging ensemble consisting of 100 decision trees, whereas two boosting ensemble algorithms are used, XGBoost and AdaBoost. A k-fold ($k = 10$) cross validation model is employed for all ensemble learners.

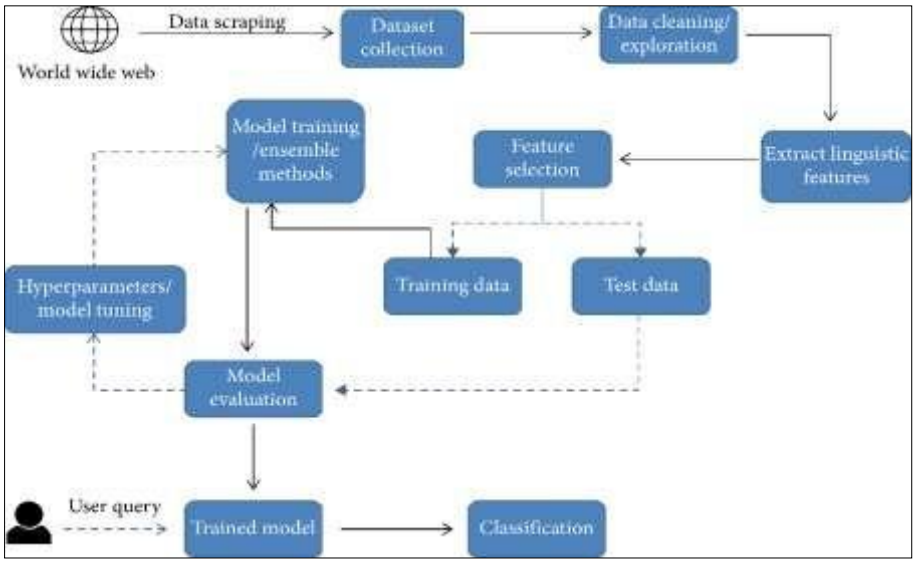

Fig. 3.1 Architecture

4. Algorithm

Tokenization a common task a data scientist comes across when working with text data. It consists of splitting an entire text into small units, also known as tokens. Most Natural Language Processing (NLP) projects have tokenization as the first step because it's the foundation for developing good models and helps better understand the text we have. Although tokenization in Python could be as simple as writing split(), that method might not be the most efficient in some projects.

Logistic Regression classifier is evaluating the parameters of a Logistic model; it is a type of binomial relapse. Scientifically, a twofold calculated model has a reliant variable with two conceivable qualities, for example, pass/fail, win/lose, alive/dead; these are spoken to by a pointer variable, where the two qualities are marked "0" and "1".

Natural Language Processing: The main reason for utilizing Natural Language Processing is to consider one or more specializations of system or an algorithm. The Natural Language Processing (NLP) rating of an algorithmic system enables the combination of speech understanding and speech generation. In addition, it could be utilized to detect actions with various languages.

Data mining techniques are categorized into two main methods, which is; supervised and unsupervised. The supervised method utilizes the training information in order to foresee the hidden activities. Unsupervised Data Mining is a try to recognize hidden data models provided without providing training data for example, pairs of input labels and categories. A model example for unsupervised data mining is aggregate mines and a syndicate base.

Machine Learning (ML) is a class of algorithms that help software systems achieve more accurate results without having to reprogram them directly.

Data scientists characterize changes or characteristics that the model needs to analyze and utilize to develop predictions. When the training is completed, the algorithm splits the learned levels into new data [11]. There are six algorithms that are adopted in this paper for classifying the fake news.

Decision Tree is an important tool that works based on flow chart like structure that is mainly used for classification problems. Each internal node of the decision tree specifies a condition or a "test" on an attribute and the branching is done on the basis of the test conditions and result. Finally, the leaf node bears a class label that is obtained after computing all attributes. The distance from the root to leaf represents the classification rule. The amazing thing is that it can work with category and dependent variable. They are good in identifying the most important variables and they also depict the relation between the variables quite aptly.

Random Forest are built on the concept of building many decision tree algorithms, after which the decision trees get a separate result. The results, which are predicted by large number of decision tree, are taken up by the random forest. To ensure a variation of the decision trees, the random forest randomly selects a subcategory of properties from each group. The applicability of Random forest is best when used on uncorrelated decision trees. If applied on similar trees, the overall result will be more or less similar to a single decision tree. Uncorrelated decision trees can be obtained by bootstrapping and feature randomness.

Gradient Booster is one of the most popular machine learning algorithms for tabular datasets. It is powerful enough to find any nonlinear relationship between your model target and features and has great usability that can deal with missing values, outliers, and high cardinality categorical values on your features without any special treatment.

5. Use Case Diagram

The following is the use case diagrammatic representation of the detection of fake news with highest accuracy.

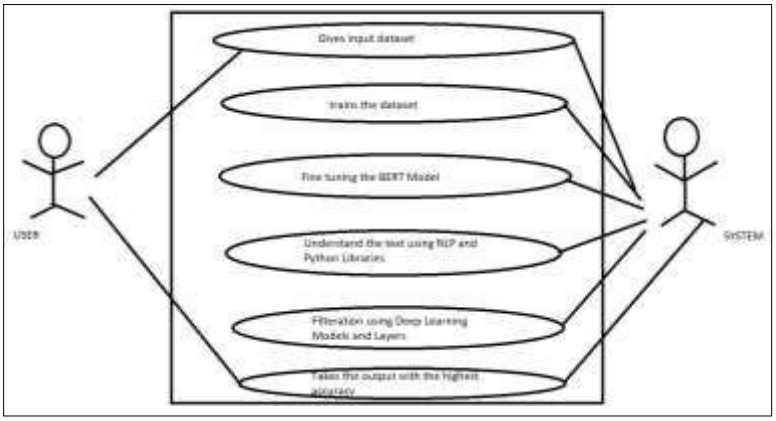

Fig.5 Use case diagram

6. Performance Evaluation

Classification Metric

To check how well the model we use some metrics to find the accuracy of our model. There are many types of classification metrics available in Scikit learn.

1. Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or ” classifier”) on a set of test data for which the true values are known.

A. true positives (TP): These are cases in which we predicted yes (news is fake), and it is fake news.

B. true negatives (TN): We predicted not fake news, and it is not fake news.

C. false positives (FP): We predicted news is fake news, but it is actually real news. (Also known as a ” Type I error.”)

D. false negatives (FN): We predicted real news, but it is actually fake news. (Also known as a ” Type II error.”)

2. Accuracy Score

Accuracy is a metric used in classification problems used to tell the percentage of accurate predictions.

Accuracy = TP+TN/(TP+TN+FN+FP)

3. Precision

Precision is the ratio of actual fake news detected by the model and all the news classified by the model as fake. In terms of the true positives (TP) and false positives (FP), precision (p) can be formulated as the equation precision = TP/(TP+FP).

4. Recall

Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. recall = TP/TP+FN

5. F1-Score

F1-score is one of the most important evaluation metrics in machine learning. It elegantly sums up the predictive performance of a model by combining two otherwise competing metrics — precision and recall. F1-Score = 2* (Precision + Recall) / (Precision + Recall)

Final Outcome

Our model is built on the top of a bidirectional transformer encoder-based pre-trained word embedding model (BERT). Classification results demonstrate that FakeBERT provides more accurate results with an accuracy of 98.90 percentage. The accuracy of FakeBERT is better than the current state-of-the-

art models with real-world fake news dataset: Fake-News.footnotes should be avoided if possible. Necessary footnotes should be denoted in the text by consecutive superscript letters1. The footnotes should be typed single spaced, and in smaller type size (7pt), at the foot of the page in which they are mentioned, and separated from the main text by a one-line space extending at the foot of the column. The Els-footnote style is available in MS Word for the text of the footnote.

7. Conclusion

In this, we have demonstrated the performance of our proposed model (Fake-BERT -a BERT-based deep convolutional approach) for fake news detection. Our model is a combination of BERT and three parallel blocks of 1d-CNN having different kernel-sized convolutional layers with different filters for better learning. Our model is built on top of a bidirectional transformer encoder-based pre-trained word embedding model (BERT). Classification results demonstrate that FakeBERT provides more accurate results with an accuracy of 98.90 percent. The accuracy of FakeBERT is better than the current state-of-the-art models with real-world fake news dataset: Fake-News. This dataset consists of thousands fake and real news articles during the 2016 U.S. General Preseantiaial Election. We evaluated our models with different parameters (Accuracy, FPR, FNR, and Crossentropyloss). We will further study the problem of fake news from the viewpoint of different echo- chambers that exists in social media data, which can consider as a group of personalities having the same opinion on any social concern. The prime motivation to introduce echo chambers is that every user is co-related in a graph-like structure (not in isolation) to any social media platform like a community.

References

[1] Crestani F, Rosso P (2020) The role of personality and linguistic patterns in discriminating between fake news spreaders and fact-checkers.” I n Natural language processing and information systems: 25th international conference on applications of natural language to information systems, NLDB 2020, Saarbrucken, Germany. Proceedings, vol 181. Springer Nature

[2] Alkhodair S A, Ding S H.H, Fung B C M and Liu J 2020 Detecting breaking news rumors of emerging topics in social media” Inf. Process.

Manag. 57 102018 2020

[3] Kaur Prabhjot et al 2019 Hybrid Text Classification Method for Fake News Detection Inf.” International Journal of Engineering and Advanced Technology (IJEAT) 2388-2392

[4] Bondielli A, Marcelloni F (2019) A survey on fake news and rumor detection techniques.” Inform Sci 497:38–55

[5] Chen W, Zhang Y, Yeo CK, Lau CT, Sung Lee B (2018) Unsupervised rumor detection based on users’ behaviors using neural networks.” ’Pattern

[6] De S, Sohan FY, Mukherjee A (2018) Attending sentences to detect satirical fake news. ” In: Proceedings of the 27th international conference on computational linguistics, pp 3371–3380

[7] Ahmed H, Traore I, Saad S (2017) Detection of online fake news using N-gram analysis and machine learning techniques. systems.,” : International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer, Cham, pp 127–138

[8] Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election.” Econ Perspect 31(2):211–36

[9] Granik Mykhailo and Mesyura Volodymyr 2017 First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (Ukraine:

IEEE) Fake news detection using naive Bayes classifier” Journal of Computational and Theoretical Nanoscience., 12. 6334-6342. 10.1166/jctn.2015.4675.

[10] Greff K, Srivastava RK, Koutn´ık J, Steunebrink BR, Schmidhuber J (2016) LSTM” IEEE Trans Neural Netw Learn Syst 28(10):2222–2232, [7] De S, Sohan FY, Mukherjee A (2018) Attending sentences to detect satirical fake news. ” In: Proceedings of the 27th international conference on computational linguistics, pp 3371–3380