HINDI FAKE NEWS DETECTOR

Dr.A.Siva Kumar, Adishree Solapurkar, Akshit Khamesra

Dept. of Computer Science and technology, SRM Institute of Science and Technology, KTR, Chennai, Tamil Nadu, India

Abstract

Fake news has proliferated on the internet in recent decades. More people than ever before are creating and sharing knowledge because of social networks, many of which have no connection to reality. This has led to the rapid dissemination of false information used for various political and business objectives. Finding reliable news sources has become more difficult due to online newspapers. In this work, we gathered news articles of Hindi text from various news sources. Techniques for pre-processing, feature extraction, classification, and prediction are all extensively covered. A "Fake news detection system" has been developed in this project. Various Hindi news articles have been collected from multiple sources to help diversify the dataset and train the model better. The project first pre-processes the dataset and uses the pre-trained Bert model for feature extraction. Then, the data is classified from the dataset and prediction processes are employed on the dataset. Various machine learning algorithms and deep learning models like Naïve Bayes, Long Short-Term Memory, Logistic Regression have been employed in previous works for the purpose of detecting fake news. Pre-processing steps include data cleaning, stop words removal, tokenizing, stemming. The testing and training of the dataset include using the BERT for sequence classification model. The model is trained and tested against the validation dataset.

Keywords: Fake News Detection, Machine Learning, Hindi News, BERT model

1. Introduction

Fake news is the intentional or unintentional dissemination of unverified, fabricated, or unauthentic information via social media. In today's world, anyone can upload information on the internet. Unfortunately, fake news gains a lot of attention especially through the various social media platforms available. Individuals become misdirected and do not reconsider before directing such mis-educational pieces to the farthest reaches of the arrangement.

With the rapid rise of social media all across the globe, information and knowledge available on the internet also has increased vastly. Everything is at the tip of our fingers because of the internet. But it's not necessarily a boon. Due to the easy accessibility of information, many parties, organizations, committees and companies take advantage of this and spread counterfeit

or fake agendas, news, and schemes that benefit their personal purposes. The internet and especially social media is now being used to manipulate the general audience by misinforming the readers using bots, cyborgs and scammers. Because of the exponential rise in the spread of fake news, it is now becoming a big challenge to detect it. Many researchers turn to machine learning to classify and determine whether the given information is fake or not but have not been quite successful.

The goal of this research is to determine whether the provided news is fake or not using Hindi text-based news. Since there haven't been many advancements in the field of identifying or detecting fake news in Hindi, this project takes on the issue and addresses it with the aid of numerous research papers that have been carried out in this field. Different classification models on the dataset will be compared and a model will be finally proposed which can accurately determine whether the provided news is fake or real with the use of BERT.

---

2. Literature survey

Many projects have been developed in the field of fake news detection in English. Detecting fake news can be a difficult task as there are many parameters to be considered. To find the right model is another big challenge as there are a lot of machine learning and deep leaning models available and to pick the right model, proper research and understanding of machine learning is required. Overcoming language barriers is next. Due to the challenge of language barriers and the lack of datasets available, fake news detection in local languages is still a huge issue to tackle and not many projects and research has been done in this field. This project takes up on this issue and uses BERT model [9] to successfully detect fake news in Hindi with the help of various research papers and projects conducted in this field.

Data is collected either by web scraping [1], a website can be used to scrap fake news and true news and various classifiers like SVM [1] [7,13], random forests [1,3,4,8] or the combination of both like in [13] where SVM-RF produces a higher accuracy can be used. Kaggle and other websites can be used to collect datasets but in this project, web scraping is done. Various classifiers and deep-learning models have been used in different projects all over the world. Naïve-Bayes is also a famous classifier that is used to calculate probability and how much certain features influence the result that can be used but as observed in [1,4,9] it produces a comparatively lower accuracy, but like in [9] BERT + Customised NVB can be used to acquire a higher accuracy. This combination combines the feature extraction from BERT and the probability calculation from naïve-bayes. This project uses the BERT model [9,10] to train and classify the dataset. Using the BERT model has a huge advantage over the readability of the input as it understands ambiguity. Many projects have been made using RNN and CNN [4,10] but it is found out that BERT is more suitable and produces more accuracy from the related work and research papers. Graph Transformer network (GTN) [2] can also be used but the limitation here is emotions and sentiment analysis has not been taken into consideration which vastly effects the results. TF-IDF and count vectorizer [3,5,7,8] is used for feature extraction and a weighing scheme [7] that assigns weights to the features involved can be implemented

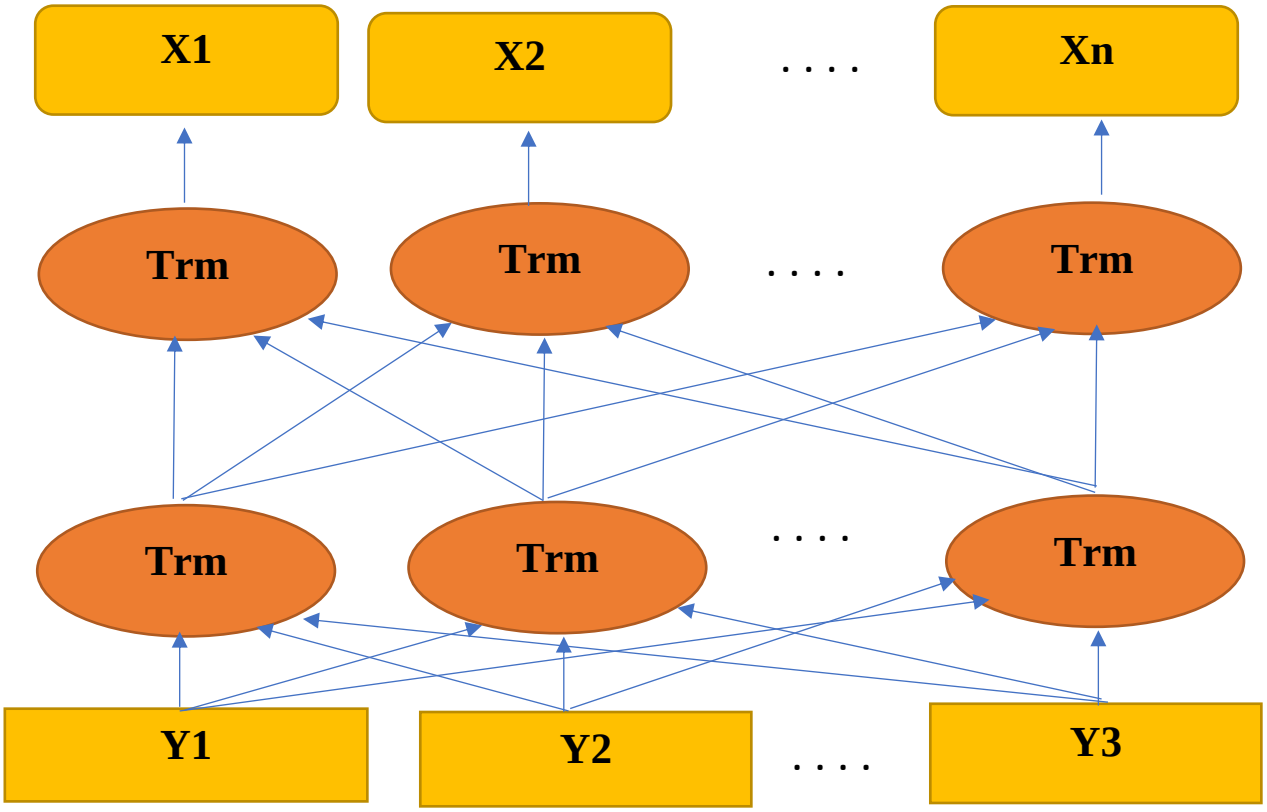
for feature extraction. A BERT based deep learning approach [10] is implemented and python libraries are used for the same. BERT is a pre-trained model that helps in feature extraction and it is a deep learning model in which each output data is connected to each input. Instead of using regex [13] for pre-processing, this project uses tokenization and generalization [3,5] using BertTokenizer for assigning numbers to each tokens. After the dataset is trained and tested, a confusion matrix [9,6] is implemented to obtain the values of accuracy, precision, recall [2,9,10] which will be the result.

---

3. Proposed Model

There are various different approaches that can be used for the identification of fake news. Multiple projects have been made using different machine learning and deep learning models. In this project, the BERT model has been chosen to classify the given dataset into fake news and real news. The BERT model is essentially an open-source framework of machine learning for natural language processing (NLP). BERT was created by Google researches to help computers in understanding the meaning of ambiguity by establishing context using text. The BERT framework has already been pre-trained with information taken from the source: Wikipedia in the form of text and can be tested with question-and-answer datasets.

BERT is an abbreviation for Bidirectional Encoder Representations from Transformers. It is a deep learning model in which each output data is connected to each input data. Old models like RNN and CNN could read the text from right to left and vice versa in the past, but not both at the same time. The BERT model is unique in that sense that it is designed to read and interpret from both sides. This ability prelude to transformers is known as bidirectionality. BERT is a transformer-based model and does not use recurrent connections, but only attention and feed-forward layers. BERT is different from transformers in such a way that BERT only has an encoder while the transformer has both encoder and decoder.



We want the model to be able to process patterns in a language, and BERT has the ability to do so, allowing us to train the model in an efficient manner and categorize the input dataset into fake and real news. Due to bidirectionality of the BERT model, it produces a higher accuracy than the other deep learning models like convolutional neural networks (CNN) and recurrent neural networks (RNN) when it comes to a much larger and complicated datasets. BERT understands the contextual relationship between different words of the sentence. For BERT, only an encoder of the transformer is enough.

The major advantage of BERT over the directional models like RNN, LSTM, CNN is that BERT is non-directional and reads the whole sentence as the input instead of sequential ordering. It has a significant benefit over the approaches discussed above. It can be quickly implemented within the parameters of already available models, providing non-specialist fake-news detection systems available over the internet with understanding of the model's decision-making process. The model's classification feature, the tokenizer, and the sample are used in this project. The input for this project is a dataset which will be pre-processed, trained and tested using TensorFlow, pytorch and transformers. This project uses a BERT model of 12 layer with an uncased vocab.

4. Proposed Solution

4.1. Dataset

The lack of a high-quality dataset presents a serious obstacle to fake news detection. For the English language, there are a ton of datasets available, however there are hardly any for the Hindi language. The dataset for this project has been collected across two platforms. Fo fake news, boomlive has been used and jagranjosh is used for true news. This dataset consists of 2010 rows of various Hindi articles collected across different platforms of which 760 rows are of fake news and the rest 760 are of true news.

4.2. Pre-processing

Since machines cannot understand text, the raw text has to be translated into numbers before being utilised as training data. To make the raw data suitable for the classification by the model used, it should be cleaned and this step is called Pre-processing which makes it suitable for testing and training in the later stages. The various processes of pre-processing are: Data cleaning followed by stop words removal. This is followed by extracting the base word using stemming and TF-IDF. Then, BertTokenizer is used for tokenization.

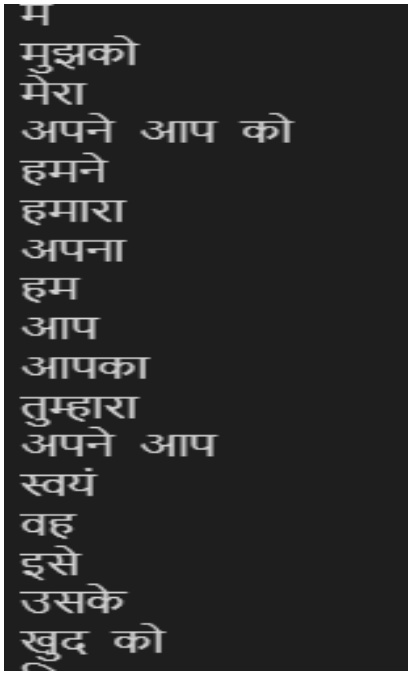
These steps are briefly discussed below:

4.2.1. Data Cleaning

Data cleaning is used to get rid of unnecessary and irrelevant data that won't be used in training the model. This involves removing null values, improperly formatted data or noises from the dataset. Making sure the dataset is cleaned is one of the most vital steps as it boosts the efficiency of the model.

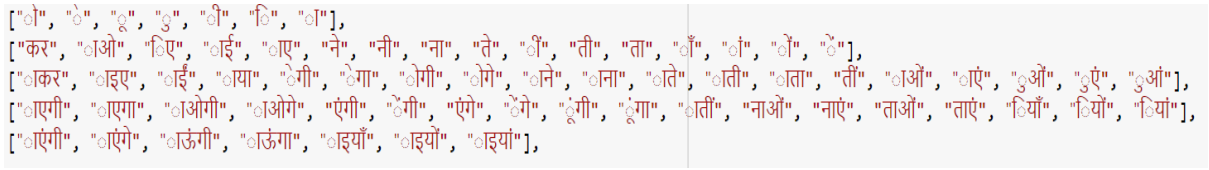
4.2.2. Removing Stop-words

Stop words are the words in a **stop list** which are frequently used and are of no significance as they have no influence over the rest and hence can be ignored. These words are removed from the dataset for easier training. Some of the stop-words in Hindi can be seen in the below fig.



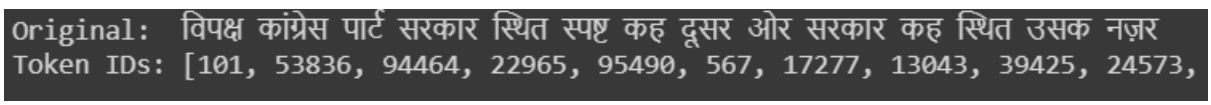
4.2.3. Stemming

Stemming involves discarding the suffixes from the tokens to extract the base form of the word. For example, for a token named “Speaking”, it will be converted to “speak” and “ing” will be discarded. Some of the suffixes in Hindi are given in the below fig.



4.2.4. Tokenization

Tokenization marks the start of the NLP process. It divides a sentence into multiple parts which consist of understandable words. In this project, we will be using BERT tokenizer to encode a sentence. It will tokenize all sentences in the dataset and map the tokens to their respective ids.



4.2. Training and Testing

The dataset is divided into 90% training and 10% validation set. BertTokenizer is used to tokenize sentences and to map the token to their id’s.

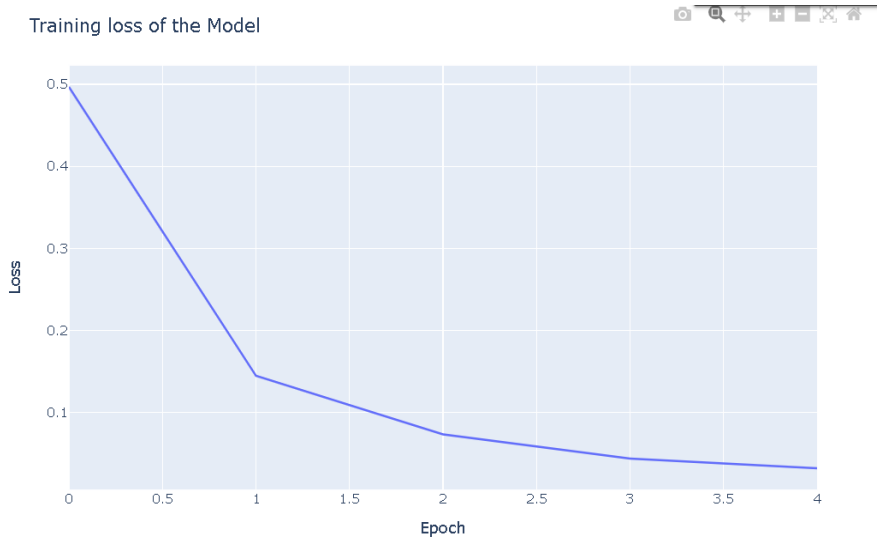
The scikit-learn library is used and train_test_split command is used to split our data into train and validation datasets for training the model. Then, we convert all inputs and labels into torch tensors, the required data type for our model.

We use the DataLoader module from the PyTorch library. DataLoader creates an iterable around the Dataset to enable easy access to the samples. The batch size for training should be known as the DataLoader takes that information. Here, we specify the batch size as 32. Then, we create the DataLoader for our training and validation set. Attention masks are created so that the BERT model knows which tokens contain padding and which don't. 1 indicates that the model will pay attention to that token and 0 indicates that the model will not pay any attention.

From the transformers module, we import BertForSequenceClassification which is the pre-trained BERT model with a single linear classification layer on top.

AdamW from the hugging face instead of pytorch is used to finetune the BERT model. The number of training steps is calculated with the formula: number of batches*number of epochs. The number of epochs this project is using is 5 and the number of batches is the len(dataloader). Each training data batch contains three parts: input id's, attention masks and labels. Loss is a tensor word that determines how much the predicted values differ from the actual values in the training data. Forward pass is implemented to calculate the loss. The training loss obtained is accumulated over all the batches which is 32 in this project and the average loss is calculated at the end. Now, backward pass is performed to obtain the gradients. Average loss over the training dataset is calculated and is used for plotting the learning curve.

Each epoch is run and after each epoch is run successfully the performance is measured on the validation dataset. Then, the model is put into the evaluation mode and logits are calculated instead of loss. Logits and labels are moved to the CPU. Accuracy of each batch is calculated at the end. We accumulate the accuracy and the number of batches is tracked using a counter variable. The final accuracy is reported for the validation run.



As it is observed from the above plot, the loss of the model is reduced after each epoch and it is almost negligible when epoch=4. The scikit-learn library is used to calculate the results in terms of the accuracy score, precision score, recall score and f1 score.

5. Results

In the field of fake news detection, numerous papers have been researched and published. In this field, there is still a lot of room for experimentation. The BERT model was used as the machine learning model in this paper. The accuracy given by the Bert model we used in this work is 97.69%. On the other hand, the other models such Naive Bayes Classifier and Logistic Regression Classifier give an accuracy of 88.23% and 89.15% respectively. However, the LSTM model gives a slightly better accuracy of 92.36%.

```
[ ]  from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score


     print(accuracy_score(actual, guesses))
     print(recall_score(actual, guesses))
     print(precision_score(actual, guesses))
     print(f1_score(actual, guesses))
```

```
0.9769736842105263
0.9868421052631579
0.967741935483871
0.9771986970684038
```

In conclusion, we can say that the BERT model gives the best accuracy overall.

6. Conclusion

In the field of fake news detection, numerous papers have been researched and published. In this field, there is still a lot of room for experimentation. The BERT model was used as the machine learning model in this paper. The project can be modified and can be used for other local language datasets. After going through multiple research papers, a deep learning model BERT is implemented and it gives the best accuracy overall when compared to other supervised learning algorithms like Naive-bayes, Support vector machines, LSTM etc.

6. References

[1] D. K. Sharma and S. Garg, "Machine Learning Methods to identify Hindi Fake News within social-media," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9580073.

[2] H. Matsumoto, S. Yoshida and M. Muneyasu, "Propagation-Based Fake News Detection Using Graph Neural Networks with Transformer," 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), 2021, pp. 19-20, doi: 10.1109/GCCE53005.2021.9621803.

[3] Dr. S. Rama Krishna, Dr. S. V. Vasantha, K. Mani Deep, 2021, Survey on Fake News Detection using Machine learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICACT – 2021 (Volume 09 – Issue 08) https://www.ijert.org/survey-on-fake-news-detection-using-machine-learning-algorithms

[4] Waqas Haider Bangyal, Rukhma Qasim, Najeeb ur Rehman, Zeeshan Ahmad, Hafsa Dar, Laiqa Rukhsar, Zahra Aman, Jamil Ahmad, "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches", Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 5514220, 14 pages, 2021. https://doi.org/10.1155/2021/5514220

[5] M. Gahirwal, S. Moghe, T. Kulkarni, D. Khakhar, J. Bhatia, Fake news detection, Int. J. Adv. Res. Ideas Innov. Tech. (2018)

[6] Thoudam Doren Singh, Apoorva Divyansha, Vikram Singh, Anubhav Sachan, Abdullah Faiz Ur Rahman Khilji, Debunking fake news by leveraging speaker credibility and BERT based model, in: International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2020, pp. 960–968.

[7] Rubin, Victoria L. et al. “Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News.” (2016).

[8] K. Harshitha , P. Veejendhiran , Aditya V , Dr. M R Sumalatha, Dr. P. Lakshmi Harika, 2022, Fake News Detection, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 11, Issue 07 (July 2022)

[9] S. Liu, H. Tao and S. Feng, "Text Classification Research Based on Bert Model and Bayesian Network," 2019 Chinese Automation Congress (CAC), 2019, pp. 5842-5846, doi: 10.1109/CAC48633.2019.8996183.

[10] Kaliyar, R.K., Goswami, A. & Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 80, 11765–11788 (2021). https://doi.org/10.1007/s11042-020-10183-2

[11] Zhou X, Zafarani R (2018) Fake news: a survey of research, detection methods, and opportunities. arXiv:arXiv-1812

[12] Yang F, Liu Y, Xiaohui Y, Yang M (2012) Automatic detection of rumors on Sina Weibo. In: Proceedings of the ACM SIGKDD workshop on mining data semantics

[13] Y Zhou, B Xu, J Xu et al., "Compositional Recurrent Neural Networks for Chinese Short Text Classification[C]", Ieee/wic/acm International Conference on Web Intelligence

[14] S. Aphiwongsophon, P. Chongstitvatana Detecting Fake News with Machine Learning Method 2018 15th International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI- CON) (2018), pp. 528-531