

Fake News Detection for Hindi Language

Kausthub Thekke Madathil¹, Neeraj Mirji², Charan R³ and Anand Kumar M⁴

¹ Department of Information Technology National Institute of Technology Karnataka Surathkal, India 575025

Abstract

The understanding of the term “Fake news” varies from one individual to the other. If we look into the basic meaning of “Fake news”, it refers to inappropriate and made up news. In most cases, the news is made up of baseless sources and facts. These news generally mislead the reader and are generally published for one’s own benefit or to defame others. In recent years, a large population is active on various social media platforms and hence they have become the major medium through which fake news is circulated. A lot of fake news is been circulated in local languages as well. Also most of the existing work is based on the English language and only very little work is done using resource scarce language for fake news identification like Indic Languages. So this paper focuses to define false news and suggest an effective method for detecting fake news in Hindi using standard machine learning algorithms like Multi-layer Perceptron and Naive Bayes and deep learning techniques like transforms - mainly mBERT.

Keywords

Fake news detection, Indic Languages, Hindi language, Classification Algorithms, Transformers, mBERT

1. Introduction


Internet has become a part of our life. There is no doubt that many young people prefer the internet to get their news rather than newspapers, radio etc. The internet provides many opportunities for us and more number of people rely on it for the majority of their knowledge. But most of the news read from the internet is either less accurate or completely fake. Fake news has quickly become a social problem. Propagating false or rumor information to change people’s behavior to create chaos in society is increasing day by day. Nowadays the word “fake news” has become most common synonym for issues which describe incorrect and misleading stories which are generally published to generate money on the basis of page views. As a result, combating fake news has become more vital, as well as more difficult, in the age of social media. This is a difficult task because even humans have difficulty distinguishing between fake and genuine news. As a result, developing an automated system for detecting fake news becomes critical. In this paper we have developed a method that can predict whether a particular news article in Hindi is fake or not using deep learning techniques.

The International Conference and Workshop on Agglutinative Language Technologies as a challenge of Natural Language Processing (ALTNLP), June 7-8, Koper, Slovenia

✉ kausthubtm.191it125@nitk.edu.in (K. T. Madathil); neerajmirji.191it232@nitk.edu.in (N. Mirji); charan.191it212@nitk.edu.in (C. R); m1_anandkumar@nitk.edu.in (A. K. M)



© 2021 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)

2. Related Work

There are several automated detection approaches for false news and deception posts that are currently in use. Counterfeit news detection has several facets, ranging from utilising chatbots to promote disinformation to exploiting click bait to spread rumours. Many click baits are available on social media networks, including Facebook, which encourages people to share and like messages, propagating false information. There has been a lot of effort put into detecting falsified data. Authors have introduced various detection techniques. Following Detection Methods have been proposed by the authors of the paper [3]. First being, Linguistic foundation Modeling deception, second, Grouping, next method is Predictive Analytics, fourth, Methods based on content cues and the last method, Nontextual cue-based methods. The writers have compiled a list of False news strategies for various sorts of fake news as and is shown in paper [3] figure 1. The accuracy of these models, according to the authors, is only 63 percent to 70 percent.

In [4] N. J. Conroy et al where the paper deals with methods for detecting fake news utilising automated detection. Linguistic Cue Approaches for Machine Learning, Rhetorical Structure and Discourse Analysis, Network Analysis Approaches, and SVM Classifiers are among the approaches addressed. These models are only text-based and offer no improvement over previous approaches.

Most of the literature focuses on fake news spread through social media. [16] S. Helmstetter and H. Paulheim (2018, August): Weekly Twitter false news detection for supervised learning. The authors classed every tweet/post in this article as a binary classification task. The source of the post/tweet is the sole criterion for classification. The authors used Twitter API and DMOZ to manually collect data sets. The results show that 15% of the tweets are false, 45 percent are genuine, and the rest is undecided. Stanford University's Samir Bajaj released a work on false news identification where fake news detection using an NLP viewpoint was proposed and they apply deep learning system. They have also considered a genuine data set from the Signal Media News dataset. Facebook and WhatsApp are also focusing on detecting fake news according to a recent article. Facebook mentioned in an article that they are attempting to stop the spread of fake news in two important areas, because most false news is motivated by money, the first step is to disrupt economic incentives. The second is to create innovative features to combat the spread of incorrect information. The following are Facebook's precautionary measures: Ranking advancements: News Feed rankings help to limit the predominance of inaccurate news items. Easier Reporting: Identify what is and is not accurate.

Most of the existing work is based on the English language and only very little literature is available for resource scarce languages for fake news identification like Indian language - Hindi. [17] D. K. Sharma et. al. proposes a dataset for Hindi Language. The dataset i.e., Hindi news is collected using a web scraping utility called the Parsehub. They have conducted multiple experiments by using standard existing machine learning algorithms and have able to achieve better results varying from 1%-5% improvement. The proposed results reflect the effectiveness of their proposed dataset. Our dataset is a combination of two datasets from BBC Boomlive and from Cornell University Hostility Detection Dataset. The BBC Boomlive dataset is formed by processing the raw dataset collected from Boomlive and BBC news which includes removing

Fake News Type	Fake news detection Method					
	<i>Linguistic Modeling</i>	<i>Deceptive</i>	<i>Clustering</i>	<i>Predictive Modeling</i>	<i>Content Cues</i>	<i>Non-Text Cues</i>
Visual-based	No	No	No	No	No	Yes
User Base	No	No	No	Yes	Yes	Yes
User Post Based	Yes	Yes	Yes	Yes	No	Yes
Social Network Based	No	No	No	No	Yes	No
Knowledge Based	No	No	Yes	No	No	No
Style Based	Yes	No	No	Yes	Yes	No
Stance Based	No	No	No	No	No	No

Figure 1: [3] False News Strategies

null values, unwanted columns and stop words. Whereas the Cornell dataset consisted of news which were labeled under the categories "hostile", "offensive", "fake" from which we considered the news under the label of "fake" only for our dataset.

3. Problem Statement

In a world where the media has more influence over its consumers than ever before, the news plays a major role in providing them with the information about current affairs taking place around the globe. But, many times they are provided with wrong information and are misled either knowingly or unknowingly. To prevent such issues, it is necessary to filter out all fake news from reaching the audience before they could lead to any sort of misunderstandings or wrong assumptions. Machine learning can be used as an aid to achieve this feat and make the news consumption more cleaner and correct. . Also, since fake news is not just limited to the English language we are targeting Hindi language and exploring different methods to overcome the problem of limited dataset and text processing techniques. This paper focuses on defining fake news and developing and comparing machine learning models that can accurately predict whether a given news article in Hindi is fake or not. Also, our aim is to propose a method to expand the limited dataset size, process the Hindi texts, train and compare standard machine learning models and discuss the results.

4. Dataset Description

The dataset was collected from two sources which were BBC Boomlive and from Cornell University Hostility Detection Dataset. The BBC Boomlive dataset consisted of 1250 fake news and 720 true news taken from popular newspapers and news channels. The Cornell University Hostility Detection Dataset News consisted for news which were labelled under the categories "hostile", "offensive", "fake" from which we considered the news under the label of "fake" only for our dataset. The number of fake news collect from this dataset was 1010. In total the combined



Figure 2: Word cloud for the dataset

dataset consists of 3020 news out of which 760 are true news and 2260 are fake news. There are a total of 4 attributes for each data row which includes the news title, short description of the news, long description of the news, and the labels. For this paper we have considered only the short description and rest of the columns and attributes are dropped off.

Table 1
Dataset Summary

DATASET	FAKE NEWS DATA	TRUE NEWS DATA
BBC-BOOMLIVE	1250	720
CORNELL UNIVERSITY HOSTILITY DETECTION	1010	0
TOTAL	2260	720

5. Methodology

This paper’s focus would be various machine learning models and its performance to identify fake news specifically for Hindi Language. Since the problem under consideration is text classification, we would be considering standard classification algorithms such as Naive Bayes Classifier and Multi-Layer Perceptron as these are the standard algorithms for text based processing. Also, we have discussed transformers, which is effective for text classification because of their self attention mechanism and a better understanding of word features. mBERT model with fine-tuning is discussed to draw comparisons and results. Since, the dataset collected was really small we have employed certain data augmentation techniques to increase the dataset for better prediction. The proposed model consists of crucial steps of Data Augmentation, Text Preprocessing, Tokenization and Model Architecture.

5.1. Data Augmentation

This section deals with two data augmentation techniques that we have employed on the dataset to increase the size of our training dataset. We won’t be augmenting the testing dataset because it might lead to false answers. The two techniques used are

- Back translation : The Hindi texts are translated to English and retranslated back to Hindi which might lead to certain change in the original text which might be synonyms that might be induced during the translation. For this we have considered the python library called google-trans.
- iNLTK : generate similar sentences is a function borrowed from iNLTK package which generates sentences that are similar to the original sentences by keeping the meaning still intact.

After data augmentation the size of our training set increased from close to around 2400 texts to around 10000 texts. The testing set size remained the same of around 600 texts since we did not apply augmentation to the testing data set.

5.2. Text Preprocessing

The dataset is cleaned by clean-text library from python and the texts with null values are dropped. Other columns other than the "short description" column is removed since these are not necessary for our model. This is mainly because the short description contains the gist of the entire text and is sufficient for our study, whereas the long description is too big and may introduce errors and reduce the performance. Also, due to the lack of a huge dataset it is not favorable to train our model with long description. Additionally, expanding our dataset is of prior importance in our future work and will make sure to include long description. The special characters and non-alphabetic characters are replaced by white-spaces and the text is converted to lowercase. Next the stop words and five most frequently occurring words are removed. Further the labels are added to the texts as 0 or 1 which corresponds to "Fake" or "Real".

5.3. Tokenization

Since we are dealing with sentences which cannot be directly feed into the models we must convert them into tokens. Considering Hindi language texts, we have utilized indic tokenize for tokenizing for Naive Bayes and MLP models whereas we have considered BertTokenizer for the mBERT model .

5.4. Model Specific Preprocessing

For MLP and Naive Bayes models the tokens further were lemmatized using the WordNet lemmatizer from the NLTK package. The countVectorizer from the SKLearn package then extracts features from these lemmatized tokens, returning a matrix of token counts. To keep the computational complexity acceptable, the vectorizer is configured to use just the top 4000 features/tokens. In case of mBERT, utility function padsequences() is borrowed from python to make the size of input vector same. Further attention masks are created i.e., if a token iD is 0, then it's padding, set the mask to 0 or else if a token iD is greater than 0, then it's a real token, set the mask to 1. These attention masks help our model train better.

5.5. Models and its architecture

5.5.1. Multinomial Naive Bayes

The Multinomial Naive Bayes Algorithm is a set of probabilistic algorithms based on the popular Bayes' theorem and the "naive" assumption of conditional independence between all pairs of features. The Bayes theorem basically estimates the probability $P(x/y)$, where x is the class of probable outcomes and y is the supplied case to be identified, which represents some specific characteristics. In natural language processing (NLP) challenges, naive Bayes is commonly utilised. We have basically followed a pipe line like architecture. In this type of work flow, the output of the first stage is obtained and is provided as the input for second. We will be using a powerful machine learning tool Pipeline. Its a python module that basically helps in handling such pipes. It takes two parameters as input. The Step list which is basically a list of tuples that are chained in a sequence with the last object being a estimator and the verbose.

5.5.2. Multi-Layer Perceptron

A multi-layer perceptron is a feed-forward artificial neural network (ANN). It uses forward propagation to make predictions and back propagation based on the loss function and computed gradients to train the parameters. The number of hidden layers present varies based on whether it's a shallow network or a deep network. We have constructed a basic feed-forward ANN with an input size of 4000 (matching to the size of the matrices supplied by the preprocessing function), three hidden linear layers with ReLU activation functions, and a size 2 output layer. Finally, because we are doing a binary classification job, we utilise the Adam optimizer with a rather low learning rate of 0.001, and the error function used is cross entropy loss. .

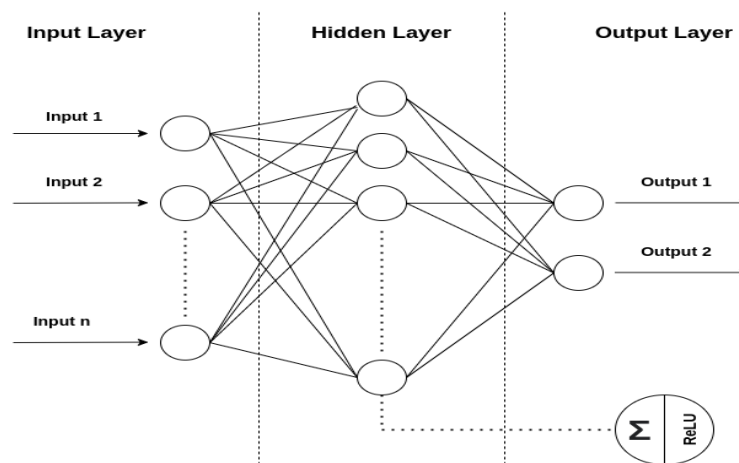


Figure 3: Multi-Layered Perceptron architecture

5.5.3. mBERT

mBERT is Multilingual Bidirectional Encoder Representations from Transformers . Its architecture is Transformer-based and is pre-trained over a large collection of unlabelled text dataset. It is “deeply bidirectional” model. Hence during the training phase it learns from both left and right side of the token’s context. One of most interesting feature of mBERT is that we can fine tune it just by adding couple of output layers. Following that, the pre-trained BertForSequence-Classification model is loaded, its parameters are locked, and a tiny multi-layer perceptron is added as a binary classifier, which is then trained. Finally, we employ a stochastic gradient descent (SGD) optimizer with a mean squared error (MSE) loss function and a 0.01 learning rate.

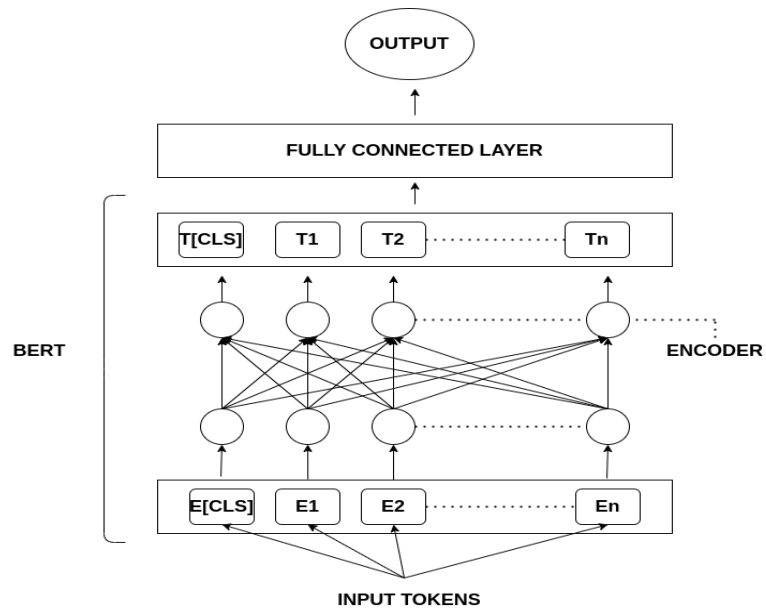


Figure 4: mBERT architecture

6. Results

This section discusses the results obtained while classifying the news into fake or real. The data is processed, and the finalized dataset is used to train individual models which are Multinomial Naive Bayes, Multi-Layered Perceptron and mBERT. The confusion matrix is considered to evaluate the models. Generally a high accuracy model represents a good model in many cases but since we are dealing with a binary classification model, A news that was forecasted as false but contained accurate facts might have cause negative implications similarly a news predicted as true but was really untrue (false positive) might can cause trust concerns. Therefore, there's a need to used other metrics into account other than Confusion Matrix. We have used considered 3 other metrics into account i.e., precision, recall, and F1-score.

We have used a training and testing split of 0.20 [There is no separate test set. We split the training data itself into train and test sets with 80% of data points belonging to the train set and the rest 20% belonging to the test set.] for the training data, the results before and after data augmentation is shown in tables 2 and 3 respectively. The average and maximum token lengths is 43.33 and 786 respectively. Hence, we have considered a maximum token length of 100 for our model.

MODEL	PRECISION	RECALL	F1 SCORE	ACCURACY
MLP	0.94	0.77	0.81	0.89
NAIVE BAYES	0.93	0.92	0.94	0.92
mBERT MODEL	0.95	0.92	0.93	0.93

Table 2
Results on Test Data before Data Augmentation

MODEL	PRECISION	RECALL	F1 SCORE	ACCURACY
MLP	0.97	0.79	0.82	0.92
NAIVE BAYES	0.94	0.96	0.95	0.96
mBERT MODEL	0.95	0.94	0.97	0.97

Table 3
Results on Test Data after Data Augmentation

We have observed better results with data augmentation. MLP has the best precision of 0.97, Naive Bayes has the best Recall of 0.96, mBERT has the best F1 score of 0.97 and the best accuracy of 0.97. Overall, the best performing model is mBERT which mainly because it uses transformers, which is effective for text classification because of their self-attention mechanism and a better understanding of word features.

7. Conclusion and Future Work

The work of manually categorizing news demands in-depth knowledge of the domain as well as expertise in identifying abnormalities in the news. We examined the challenge of classifying false news items for the Hindi language using machine learning models in this study. The data we utilized in our analysis was gathered from BBC-Boomlive and from Cornell University Hostility Detection Dataset. We have used data augmentation to increase the dataset size and we compared the two results before and after data augmentation. Clearly, after the data augmentation, we got better results. We used multiple performance metrics to compare the results for each model. When compared to the other models, the mBERT model scored higher on most performance metrics except for F1 score before data augmentation where Naive Bayes outperformed with a score of 0.94 and precision and recall after data augmentation where MLP has a better precision score of 0.97 and Naive Bayes with recall score of 0.96. The better performance of mBERT model is mostly due to the fact that it employs transformers, which are useful for text categorization due to its self-attention mechanism and a better comprehension of word aspects. This study could be further extended and few of the ideas are to :

- collect more data from various sources and news to expand our dataset in order to improve the accuracy of the model.
- Express the conclusion as a result as the probability of a news text to be fake or true. For example, it is more informative for the user to know that this news text was detected 65 percent as a Fake. This means that there are 35 percent probability to be true.
- A long term goal would be to implement and study fake news detection for various other Indian languages.

References

- [1] Mehta D, Dwivedi A, Patra A, Anand Kumar M (2021) A transformer-based architecture for fake news classification. *Soc Netw Anal Min* 11:39. <https://doi.org/10.1007/s13278-021-00738-y>
- [2] Hariharan RamakrishnaIyer LekshmiAmmal, Anand Kumar Madasamy, " Ensemble Transformer Model for Fake News Classification " NITK_NLP at CheckThat! 2021.
- [3] Parikh, Shivam B. and Pradeep K. Atrey. "Media-Rich Fake News Detection: A Survey." 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (2018): 436-441.
- [4] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.
- [5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, Association for Computational Linguistics, 2012, pp. 171–175.
- [6] Gilda, Shlok. "Evaluating machine learning algorithms for fake news detection." 2017 IEEE 15th Student Conference on Research and Development (SCORED) (2017): 110-115.

- [7] Calvillo DP, Ross BJ, Garcia RJB, Smelter TJ, Rutchick AM. Political Ideology Predicts Perceptions of the Threat of COVID-19 (and Susceptibility to Fake News About It). *Social Psychological and Personality Science*. 2020;11(8):1119-1128. doi:10.1177/1948550620940539
- [8] G. K. Shahi, D. Nandini, FakeCovid – a multilingual cross-domain fact check news dataset for covid-19, in: *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media*, 2020.
- [9] Q. Liao, H. Chai, H. Han, X. Zhang, X. Wang, W. Xia, Y. Ding, An Integrated Multi-Task Model for Fake News Detection, *IEEE Transactions on Knowledge and Data Engineering* 4347 (2021) 1–12.
- [10] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, B. W. On, Fake news stance detection using deep learning architecture (CNN-LSTM), *IEEE Access* 8 (2020) 156695–156706.
- [11] F. Monti, F. Frasca, D. Eynard, D. Mannion, M. M. Bronstein, Fake news detection on social media using geometric deep learning, 2019.
- [12] Ajao, Oluwaseun et al. “Fake News Identification on Twitter with Hybrid CNN and RNN Models.” *Proceedings of the 9th International Conference on Social Media and Society* (2018): n. pag.
- [13] M. G. Sherry Girgis and E. amer, “Deep learning algorithms for detecting fake news in online text,” in *Proceedings of the ICCES*, pp. 93–97, Cairo, Egypt, July 2018.
- [14] J. Y. Khan, M. T. I. Khondaker, A. Iqbal, and S. Afroz, “A benchmark study on machine learning methods for fake news detection,” pp. 1–14, 2019.
- [15] A. M. P. Braşoveanu and R. Andonie, “Integrating machine learning techniques in semantic fake news detection,” *Neural Processing Letters*, vol. 52, no. 2, 2020.
- [16] Stefan Helmstetter, H. Paulheim, “Weakly Supervised Learning for Fake News Detection on Twitter”, 2018 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*
- [17] 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.
- [18] <https://journals.sagepub.com/doi/full/10.1177/0002764219878224>
- [19] https://pytorch.org/hub/huggingface_pytorch-transformers/
- [20] https://scikit-learn.org/stable/modules/naive_bayes.html