# Analysis of Fake News Detection using Support Vector Machine

[1]Satish Chadokar, [2]Anchal Farkade, [3]Kajal Deshmukh, [4]Aanchal Khandelwal, [5]Nisha Barasker

Computer Science & Engineering Department
Shri Balaji Institute of Technology & Management, Betul

***Abstract*: Fake News is an undesirable buzz projected to impact the thoughts of masses. Over the last decade, specifically in India netizens have increased. Along with this admiration of social media is climbing high. It has become suitable to put anything on websites using current technologies. The Internet is perfect for producing hateful and false facts as news. They hold power to change sentiments and the way people should think about subjects. In this paper, we study and propose an idea of system for fake news detection that uses machine learning methodologies to reduce misperception caused because of fake news. We trace the paths of previously proposed models to deeply study and understand essence of the objective to build the model.**

***Keywords*: Machine Learning, Classifiers, Naïve Bayes, Support Vector Machine (SVM), Natural Language Processing.**

## I. INTRODUCTION

Fake News is made-up stuff, masterfully manipulated to look like credible journalistic reports that are easily spread online to large audiences. Most of Netizens are willing to believe twisted facts decorated enough to look authentic. Fake News is triggered whenever propaganda has to be launched among peoples. In situations like war, elections, targeted damage to reputation of organisations, competitor elimination or personal gain; fake news is used like weapon. Fake News can cause havoc and degradation in integrity of communities and can cause harm to Nation's Security. This causes greater deals to go downside. Many times, it had been observed that fake news get so much highlight that well known news sources fall for the lies and share the same evidences to support well knitted fake news.

There are numerous sites which give false data. They deliberately attempt to bring out purposeful publicity, deceptions and falsehood under the pretence of being true news. Their basic role is to control the data that can pass as valid original source. Thus, readers are open to have confidence in it. As long as you can attract more people to read these false claims and create traffic on the website. Attacker is now succeeded in spreading fake information. Attacker achieves power to control data over internet and can dictate every single thought of people reading it who is engaged to judge and make opinions based on the author of fake news. In all this process gaining trust of the reader is important. Once you score for all right things you can start creating much more chaos which is very hard to not believe. They can generate more data based on reading habits and area of spread by knowing how much the content was shared. This data can be again used to improve quality of fake news and create artificial bots to control data and information.

Fake news detection is made to stop the rumours that are being spread through the various platforms whether it be social media or messaging platforms, this is done to stop spreading fake news which leads to activities like mob lynching, this has been a great reason motivating us to work on this project. We have been continuously seeing various news of mob lynching that leads to the murder of an individual; fake news detection works on the objective of detecting this fake news and stopping activities like this thereby protecting the society from these unwanted acts of violence. Fake News Detection mechanism will have human like intelligence in judging authenticity of news articles which is helpful and saves time to brainstorm.

## II. LITERATURE REVIEW

### 2.1. Overview

There have been several initiatives taken to tackle the problem of Fake News and detect the genuine news:

- Jasmine and Rupali [1] of K.J. Somaiya College of Engineering, Mumbai, published their research paper on fake news using Machine Learning. They wrote in their research paper; fake news is not a new problem but today humans believe more in social media which leads to believe in fake news and then spread of the same fake news. They have used support vector machine, Passive Aggressive Classifier, Naive Bayes. Out of these three, support vector machine gives highest accuracy but time required for SVM is high as compared to passive aggressive, naive bayes.
- The Authors in [2] presented a method of detecting fake news using support vector machine. They designed and implemented a solution that uses a dataset of news pre-processed using cleaning techniques, steaming, Ngram encoding, bag of words and TF-IDF to extract a set of features allowing to detect fake news.
- The Authors in [3] have used multiple datasets and used basic algorithms. They applied naïve bayes, passive aggressive and deep neural network for comparative study and implementation. Naïve Bayes method is stable one but DNN outperformed the other two. Neural Networks are good in representing complex and non-linear structures.
- Anjali, Avinash, Harsh and Amit [4] implemented a system for fake news using the four existing approaches: naïve bayes, SVM and NLP. The mentioned system detects the fake news on the based on the models applied. Also, it had provided some suggested news on that topic which is very useful for any user.
- The Authors in [5] discussed a systematic review of fake news detection using machine learning and conducted a resourceful literature survey on the topic. The impact of a false news increases many folds in presence of image. The everchanging nature of fake news in the social media bubble continues to expand. They concluded that deep learning methods could be used to compute hierarchical features. Like natural language processing and modelling in categorization of news posts.

2.2. **Need of time to detect fake news:** The Delhi Government had announced new tougher measures to tackle the spread of fake news in the region, particularly after the riots in Delhi which got an over-enthusiastic communal angle on instant messages and WhatsApp forwards, Facebook posts and Twitter conversations.[6].The Chief Election Commissioner said that a proactive approach to counter fake news would facilitate credible electoral outcomes that would help preserve the 'freedoms', which the social media platforms require to thrive. Social media companies should ensure faster and more comprehensive scans of their platforms for the detection of fake news floated to influence the polls around the world.[8]

2.3. **WhatsApp approach to counter Fake News:** To stop the spread of misinformation, WhatsApp has implemented some security measures and also fake news detection. The fake news on WhatsApp has been in prominence for about a couple of years now. WhatsApp had introduced the first set of limits on message forwards, as well as the clear labelling to differentiate a forward from a personal message.[6] WhatsApp testing 'Suspicious Link Detection' feature: This feature will alert uses by putting a red label on links that it knows to lead to a fake or alternative website/news.[7]

2.4. **Outcome:**

As mentioned in the above section, all top most giants are trying to hide their selves from the rumours and focus should be on true news and authenticated articles. More or less, the approaches follow in the extraction are based on machine learning and Natural language processing. The classifiers, models and analytical algorithms are required to work hand in hand for the authentication of news articles. SVM will be used in the paper by the authors as an existing best suitable approach with Naïve Bayes. SVM is best suited for binary classification. There are various news websites and news blogs which allows to work with RSS feeds and import the references of the news articles. This will help us in finding the news accuracy.[9]

# III. METHODOLOGY

## 3.1 System Architecture

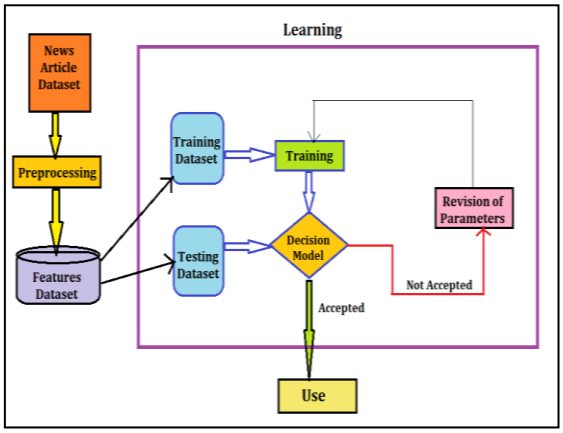


Fig 3.1 System Architecture for fake news detection

The system takes input of news dataset from database. With related information about news, such as date, author and source. It is then pre-processed into a feature dataset for clear interpretations. The pre-processed dataset is split into train dataset and test dataset for learning phase. The training is continued using train dataset and classifier algorithms, such as support vector machine. The training helps machine learning model to build a decision model useful enough for successful testing. If it passes testing phase, the model is accepted. If the model fails to give correct predictions, it is revised using parameters of failure. In real world, there is no decision model with 100% accuracy. So, a model with the most accuracy rate is accepted as novel.

## 3.2 Pre-processing

Text data needs to be pre-processed in order to be transformed into a format suitable for data modelling. There are many widely used methods for converting text data, but the method utilised is Natural Language Processing (NLTK). For news headlines and articles, removed stop words, removed punctuation, and stemmed the data. By removing the accessible extraneous material, the quantity of the actual information will be reduced. Sentences are built with stop words like as, a, the, an, are, as, at, and for. If utilised as a feature in text classification, they are meaningless. Stops Removing stop words is a crucial step in NLP since words can be processed and filtered in various ways. Stop words were eliminated using the Natural Language Toolkit package. Commas and other punctuation simply add meaning to sentences and should be removed from text because of this.

### 3.2.1 Natural language processing [6]:

It is the artificial intelligence-driven process of making human input language decipherable to software. The techniques that Natural Language Processing (NLP) uses to extract data from text are:

**Text Classification:** Text Classification is the organizing of large amounts of unstructured text (raw text data). It takes text dataset then structures it for further analysis. It is often used to mine helpful data from customer reviews.

**Keyword Extraction:** Keyword extraction is the automated process of extracting the most relevant information from text using AI and machine learning algorithms.

**Lemmatization and Stemming:** These refers to the breakdown, tagging, and restructuring of text data based on either root stem or definition.

## 3.3 Learning

In this phase datasets are split for training and testing modules. For the model to form logic of its own the pre-processed data from datasets based on features and attributes is used for categorisation and classification.

There could be features as highlights for a news to be fake or real. Developing this understanding within system is the target and sole objective.
Training: To train the model following algorithms have shown excellent results:

### 3.3.1 **Support Vector Machine Algorithm**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.
Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

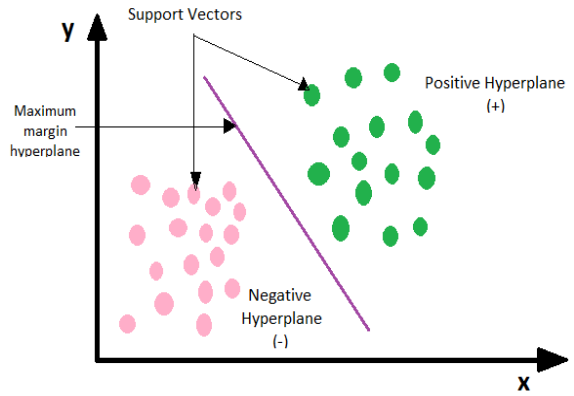


Fig 3.2 Support Vector Machine (SVM): Classification of two distinct categories using hyperplane

[4] The maximum and minimum of the decision of the decision function are calculated during the training phase and used to compute the degree of truth or fallacy by the following function:

$$p = \begin{cases} \dfrac{Dec}{Max(dec)} x\ 100, & if\ Dec > 0 \\ \dfrac{Dec}{Min(dec)} * 100, & else \end{cases}$$

where,

- Dec is the decision function value;
- Max(dec) and Min(dec) are the maximum and minimum values of the decision function;
- P is the percentage of truth or fake.

### 3.4 **Verification and Testing**

Next step after training the model is to check for the validation. The features dataset which is subdivided into two parts, a training part and a test part. Its usefulness consists in avoiding over-fitting, i.e., testing the model on the same training dataset. The subdivision is not random but according to a particular sample. The test dataset is utilized for testing phase. The test dataset contains limited information and input to the model knows nothing about output, unlike in training phase.

### 3.5 **Revision**

For improvement of the current model development and accuracy previous parameters are changed and experimented. This procedure continues until accuracy is acceptable and model gains stability.

3.6 **Usage**

The Model after various training and testing rounds is finally set to predict authentication of news. It can be used on unlabelled news to predict their class: fake or real.

## IV. EXPERIMENT & RESULT

We use SVM classification algorithm to build the proposed model.

A. Used Dataset

We used datasets from open source available in Kaggle website [10]. About the Dataset:

- id: unique id for a news article
- title: the title of a news article
- author: author of the news article
- text: the text of the article; could be incomplete
- label: a label that marks whether the news article is real or fake:
  1: Fake news
  0: real News

B. Result & Discussion

The Experiment is done using Jupyter Notebook. The following table contains the results and percentage of accuracy:

| Article | Accuracy | Implementation Method |
|---|---|---|
| Jain A., Khatter H., Shakya A. (2019) | 93.50% | Naïve Bayes, SVM, NLP |
| **Anchal F., Aanchal K., Kajal D., Nisha B. (2023)** | 94.03% | SVM, NLP |

## V. CONCLUSION AND FUTURE WORK

Machine Learning has become powerful tool in fake news detection process. The usage of Machine Learning in identification of fake news has been implemented significantly many times. This paper has shown that this approach has been the best considering time and cost applied. The usage of Machine learning in identification of fake news is still in its infancy. As it has not been used as application of real-world entity. The concept implemented in this project was an attempt to predict the fake news using learning capability of machine learning algorithm.

Projects like this if upgraded with more advanced features should be integrated on social media to prevent the spread of fake news. This model can also be trained for political news to know the accuracy of the poll results. This model has the scope to train to evaluate truthless regarding entertainment news. Similarly, the model can also be trained for finding the correctness of the medical data.

## REFERENCES

[1] Jasmine Shaikh, Rupali Patil, Department of Electronics and Telecommunication, K.J. Somaiya College of Engineering, Mumbai, India "Fake News Detection using Machine Learning" 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC).

[2] Nihel Fatima Baarir, Abdelhamid Djeffal, Mohamed Khider University of Biskra "Fake News detection Using Machine Learning" 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH).

[3] Rahul R Mandical, Mamatha N, Shivakumar N, Monica R, Krishna A, Department of Computer Science and Engineering SJB Institution of Technology Bangalore, India "Identification of Fake News Using Machine Learning" 2020 IEEE.

[4] Anjali Jain[1], Avinash Shakya[2], Harsh Khatter[3], Amit Kumar Gupta[4], [1,4]KIET Group of Institutions [2,3]ABES Engineering College, Ghaziabad "A SMART SYSTEM FOR FAKE USING MACHINE LEARNING"2019 2nd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT).

[5] Syed Ishfaq Manzoor, Dr Jimmy Singla, Nikita School of Computer Science & Engineering Lovely Professional University, Phagwara Punjab India "Fake News Detection Using Machine Learning approaches: A systematic Review" Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).
[6] https://www.news18.com/news/tech/whatsapps-biggest-move-yet-in-the-war-against-fake-news-message-forwards-limited-to-one-2567797.html
[7] https://www.deccanherald.com/national/cec-wants-social-media-platforms-to-detect-fake-news-faster-1158272.html
[8] https://www.deccanherald.com/national/cec-wants-social-media-platforms-to-detect-fake-news-faster-1158272.html
[9] https://monkeylearn.com/blog/natural-language-processing-techniques/
[10] jru dataset available at kaggle: https://www.kaggle.com/datasets/jruvika/fake-news-detection
[11] Kaggle fake news data set: https://www.kaggle.com/datasets/pontes/fake-news-sample