

Detecting Fake News with Machine Learning Model

Prof. Sanjivanee R Kale, Kartik D Borkar, Mayur P Bokad, Ayush G Joshi, Himanshu A Wadhonkar.

Computer Science and Engineering

ABSTRACT

Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites.[1] This kind of fake news when spread may have a severe effect. Hence it is extremely essential that certain measures should be taken in order to reduce or distinguish between real and fake news. This paper presents a survey on fake news detection based on various supervised, unsupervised and semi supervised machine learning techniques. With the help of machine learning it is tried to determine whether the news is real or fake using Support Vector Machine.

Keywords: Machine Learning, Naïve Bayes Classifier, Support Vector Machine(SVM).

INTRODUCTION

News is basic part of our life. In regular daily existence current news are valuable to improve data what happen all over. So most of society favor watching news by far most of the society all around slant toward examining paper quickly in the initial segment of the day getting an accuse out of cup of tea. If news is phony that will trick social orders occasionally counterfeit word used to get out gossipy goodies about things or it will impact some political pioneer positions taking into account counterfeit news. By far most of the serious cell phone customers need to scrutinize the news by methods for online media over web[2]. The main objective is to detect the fake news, which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can differentiate between “Real” news and “Fake” news. This leads to consequences in social networking sites like Facebook, Instagram, microblogging sites like Twitter and instant messaging applications like WhatsApp, Hike where these fake news gets a major boost and gets viral among people, around the country and globe[1].

Therefore, noticing fake news becomes a vital problem attracting huge research efforts. Detection of fake news from social media always creates a new challenge. It is written on social media to mislead readers. In

the 2016 US presidential election, fake news propagated more on Facebook than authentic news[3]. So it is very difficult for readers to understand the motto of news whether it is released for entertainment purposes or any other purpose. That's why it is very necessary to develop a such model which can easily indicate the motto of news so that readers will not get distracted[3].

RELATED WORK

In [4], authors have proposed a model that creates a whole network for learning a depiction of news, reports, authors, and titles simultaneously. To achieve better accuracy, they have used several ML algorithms: Support Vector Machine, Convolution Neural Networks, Long Short-Term Memory, K- nearest neighbors, and Naive Bayes. The author proposed a model that was first tested initially on CNN based machine learning algorithm that provides accuracy with 94% in a combined dataset (Liar and Kaggle) but, it has seen that using the KNN model only predicts 70% that's the very worst model. When they have examined their dataset using SVM with an accuracy of 73% that was almost similar to their previous algorithm, then they have to demonstrate their model after using Naive Bayes with 91% accuracy which was much better. The author has performed their models on data that had taken from large datasets with 25680 data information. Finally, it has lastly achieved better accuracy of 97% by using LSTM. In this paper, a goal has been set for the author to examine small sentences and news in concern and construct the reliability count with the news by putting serially feature extraction and credibility scores. Finally, they have evaluated their overall very high accuracy of 94% with a combination of three algorithms as Naïve Bayes, CNN, and LSTM, and based on their accuracy, recall, and F-1 count on each step, they would be achieved their performance and effectiveness[3].

In [1], Jain et al. (2019) have demonstrated a model with the support of ML and NLP techniques to assemble articles using a Support Vector Machine (SVM) and resolve whether the news is real or fake. They have used a support vector machine algorithm for binary classification to systemize the articles and based on that model works to categorize the articles either real or fake. They have used three main modules for refining their articles or contents in their proposed models as an aggregator, authenticator, and suggestion or recommendation system. In this paper, they have also used the Naïve Bayes algorithm to test the articles either true or false and for obtained with an accuracy of 93.50% achieved by the combination of these three algorithms i.e. Naive Bayes, SVM, and NLP[3].

In [5] A. Uppal et al., propose and execute a strategy for modernized misleading recognition. The proposed system uses significant learning in talk level structure examination to detail the structure that isolates phony and real news. The standard model achieved 74% precision[2].

In [6][7] V. M. Krešňáková, and C. K. Hiramath et al., introduced the location of phony news from the printed information utilizing profound learning methods. thought was to prepare various kinds of neural organization models utilizing both whole messages from the articles and to utilize only the title text. The models were prepared and assessed on the Phony News dataset acquired from the Kaggle rivalry. Likewise proposed counterfeit news discovery framework dependent on characterization, for example, Strategic relapse (LR), Innocent Bayes (NB), Backing vector machine (SVM), Arbitrary woods (RF) and profound neural organization (DNN) and think about all AI strategies for distinguishing counterfeit news [2].

MODELING AND ANALYSIS

Using this model, a tool is implemented for detecting fake news. In this model supervised machine learning used for classifying the data. The first step in this classification problem is dataset collection phase, followed by pre-processing, implementing features selection, then perform the training and testing of dataset and finally running the classifiers. Figure 3.1 describes the proposed model methodology, The methodology is based on conducting various experiments on dataset using the algorithms described in the previous section named Random forest, SVM and Naïve Bayes, majority voting and other classifiers. The experiments are conducted individually on each algorithm, and on combination among them for the purpose of best accuracy and precision.

The main goal is to apply a set of classification algorithms to obtain a classification model in order to be used as a scanner for a fake news by details of news detection and embed the model in python application to be used as a discovery for the fake news data [16][17].

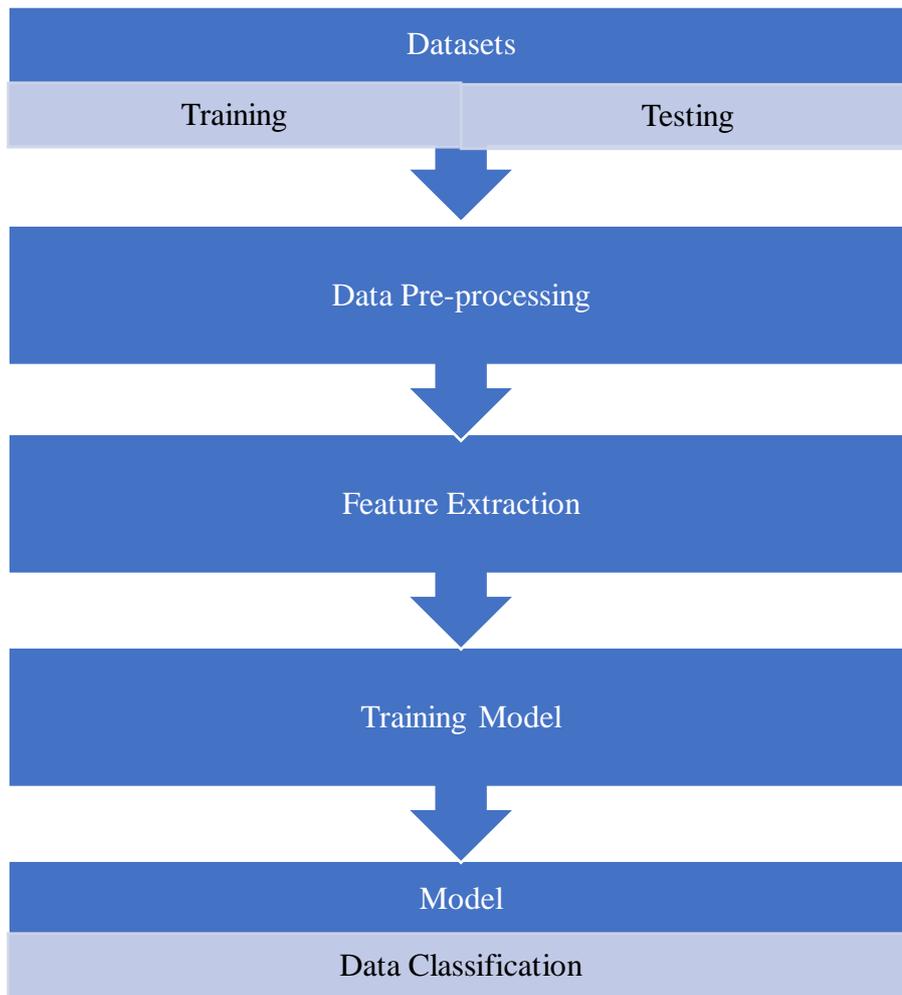


Figure.1 Flowchart for proposed work

As shown in the figure 1, the dataset is applied to different algorithms in order to detect a fake news. The accuracy of the results obtained are analyzed to conclude the final result.

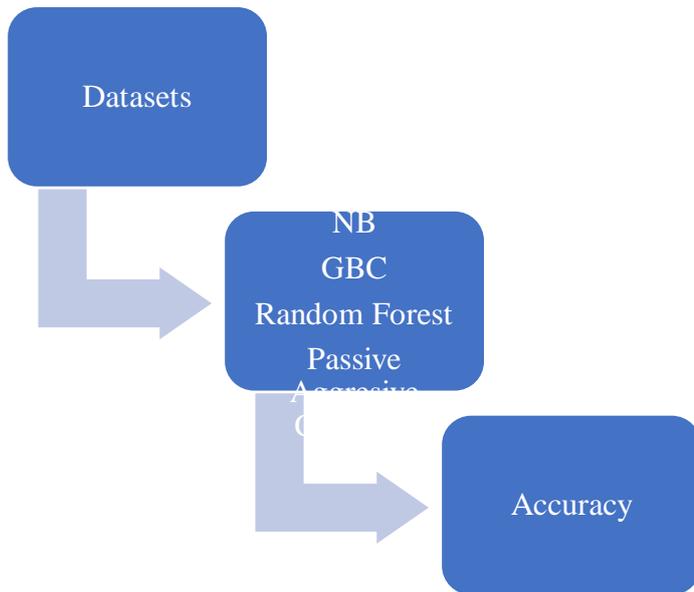


Figure 2: Classification of algorithms

Full dataset is created with half of the data being fake and half with real articles, thus making the model's reset accuracy 50%. Random selection of 80% data is done from the fake and real dataset to be used in our complete dataset and leave the remaining 20% to be used as a testing set when our model is complete. Text data requires pre-processing before applying classifier on it, so we will clean noise, using Stanford NLP (Natural language processing) for POS (Part of Speech) processing and tokenization of words, then we must encode the resulted data as integers and floating point values to be accepted as an input to ML algorithms. This process will result in feature extraction and vectorization; the research using python sklearn library to perform tokenization and feature extraction of text data, because this library contains useful tools like Count Vectorizer and TFIDF-Vectorizer.

METHODOLOGY

Fake news has a spacious effect on our daily life. Detecting fake news is now becomes very important for human beings, for that we generally use machine learning algorithms those are:

- 1) Naïve Bayes
- 2) Support Vector Machine
- 3) Neural Network
- 4) N-Gram Analysis

1) Naïve Bayes

It is an ML algorithm [8] works on Bayes' theorem. It is another supervised learning algorithm that is used for finding classification difficulty as earlier it is mentioned that it is working on Bayes' theorem, so the formula of Bayes' theorem is :

The probability of X is divided by the probability of Y

then it is equal to = $P(X/Y) * P(X)/P(Y)$

Where Probability of X is divided by probability of Y is the posterior probability

$P(X/Y)$ is a Likelihood probability

$P(X)$ is Prior Probability

$P(Y)$ is Marginal Probability

2) Support Vector Machine

It is a strong and supple supervised algorithm, generally for classification and reverting SVM is used[8]. SVM was discovered in 1960 and then again in 1990 it was modified. It is used to differentiate the datasets into classes to find hyperplanes. It is used to reduce the error with the help of a hyperplane.

RESULT AND ANALYSIS

It is evident that the maximum accuracy achieved on Random Forest which is 99.95%. The next highest accuracy is achieved on Decision Tree which is 99.91%. The next highest accuracy is achieved on Gradient Boost Classifier of 99.90%. The next highest accuracy is achieved on Passive Aggressive Classifier which is 99.80%. The least accuracy is achieved on Logistic Regression which is 94.45%. Below Table Represents the name of the classifier and accuracy achieved by classifier.

CLASSIFIER	ACCURACY
Passive Aggressive Classifier	99.80%
Gradient Boost Classifier	99.90%
Random Forest	99.95%
Logistic Regression	99.45%
Decision Tree	99.91%

Table: Accuracy Of Algorithms

Confusion Matrix:

A confusion matrix is a summary of prediction results on a classification problem .The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows the ways in which your classification model is confused when it makes predictions. It gives you insight not only into the errors being made by your classifier but more importantly the types of errors that are being made. In each of the confusion matrix given below for the machine learning algorithms.

- Logistic Regression:**

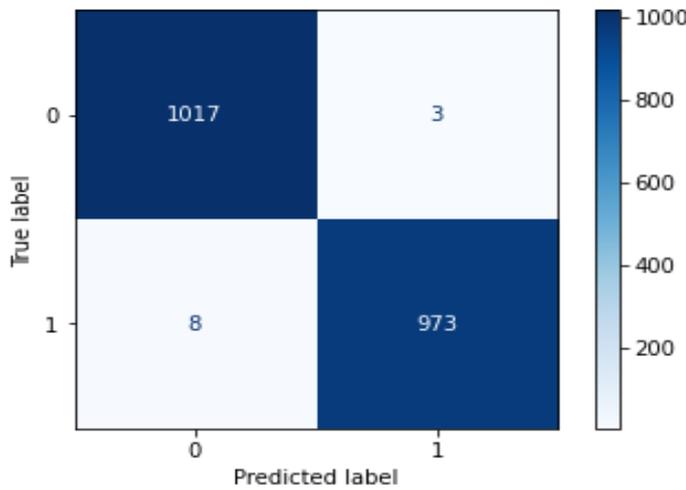


Figure 5.2.1 Confusion matrix for LR

In the above confusion matrix numbers in blue squares represents the True predictions of logistic regression algorithm and the numbers in white squares represents False predictions done by logistic regression algorithm.

- Decision Tree:**

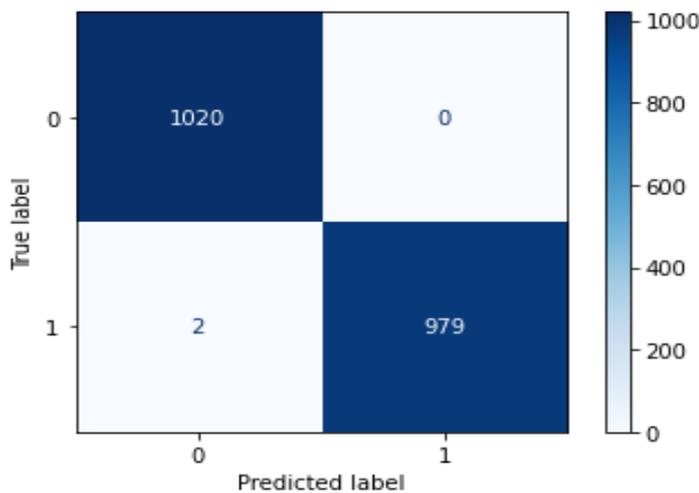


Figure 5.2.2 Confusion matrix for DT

In the above confusion matrix numbers in blue squares represents the True predictions of decision tree algorithm and the numbers in white squares represents False predictions done by decision tree algorithm.

- **Random Forest:**

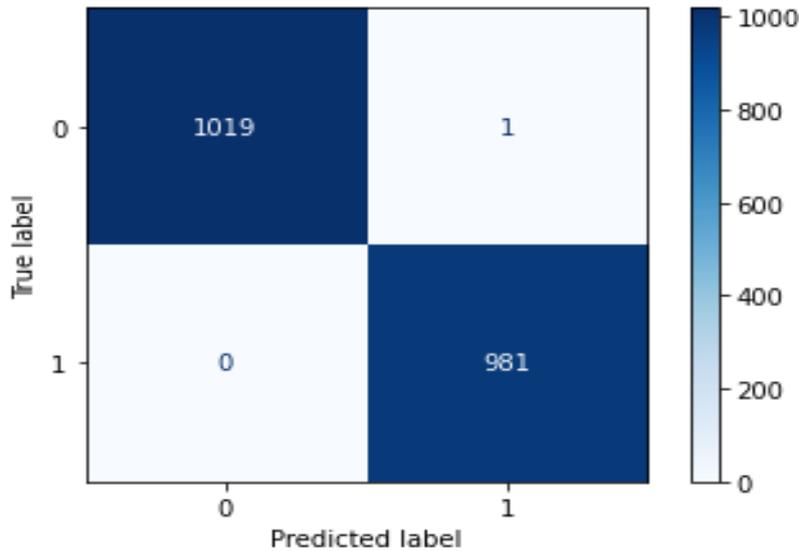


Figure 5.2.3 Confusion matrix for RF

In the above confusion matrix numbers in blue squares represents the True predictions of random forest algorithm and the numbers in white squares represents False predictions done by random forest algorithm.

- **Passive Aggressive Classifier:**

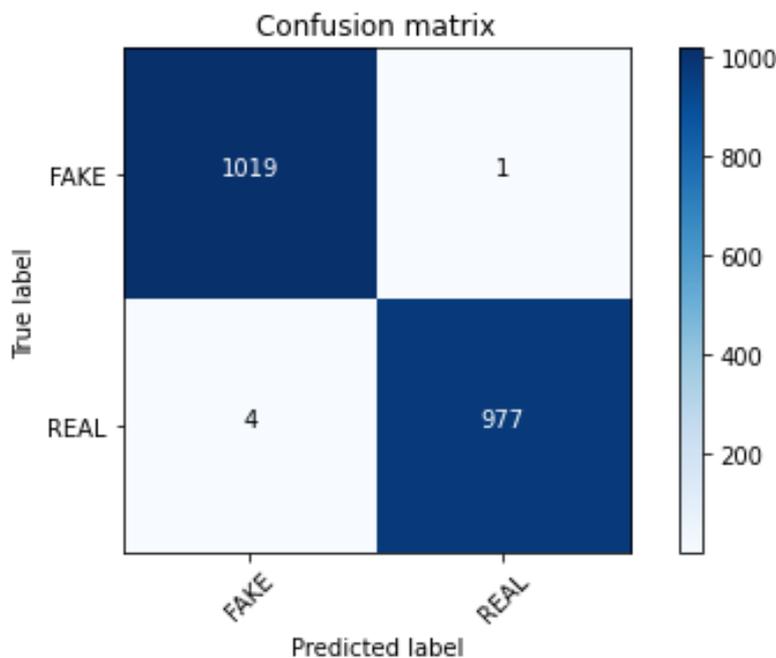


Figure 5.2.4 Confusion matrix for PAC

•Gradient Boost Classifier:

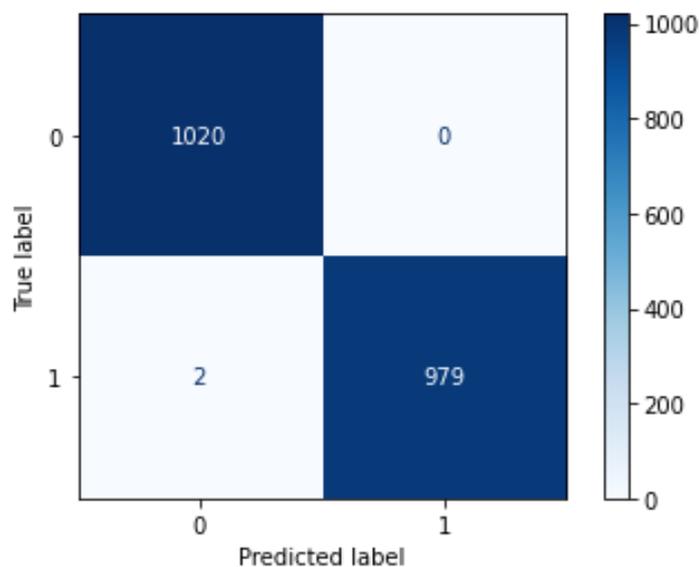


Figure 5.2.5 Confusion matrix for GBC

In the above confusion matrix numbers in blue squares represents the True predictions of gradient boost algorithm and the numbers in white squares represents False predictions done by gradient boost algorithm.

CONCLUSION

The review study discusses pioneering existing work in the field of false news detection. Machine learning-based classification algorithms play a very important role in the detection of fake news or rumors from social media, which is a very complicated and difficult process due to the diverse political, social and economic, and many other related factors. This review discusses various machine learning approaches such as NLP, SVM, Artificial Neural Networking, and many more[3].

REFERENCES

- [1] Anjali Jain, Harsh Khatter, Avinash Shakya, "A *SMART SYSTEM FOR FAKE NEWS DETECTION USING MACHINE LEARNING*" International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). Vol. 1. IEEE, 2019.
- [2] Atul Suryawanshi, Himanshu Yadav, Chetan Agrawal " *A Review of Fake News Detection using Machine Learning*" International Journal of Advanced Research in Computer and Communication Engineering Vol. 9, Issue 9, September 2020.
- [3] Murari Chaudhari, Prashant, Shashank Jha, Ashutosh Kumar Singh, Deepika Saxena " *A Review of Fake News Detection Methods using Machine Learning*" 2nd International Conference for Emerging Technology (INCET) Belgaum, India. May 21-23, 2021.
- [4] Agarwal, Arush, and Akhil Dixit. "Fake News Detection: An Ensemble Learning Approach." 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020.
- [5] A. Uppal, V. Sachdeva and S. Sharma, "Fake news detection using discourse segment structure analysis," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India.
- [6] V. M. Krešňáková, M. Sarnovský and P. Butka, "Deep learning methods for Fake News detection," 2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo), Szeged, Hungary, 2019, pp. 000143-000148, doi: 10.1109/CINTI-MACRo49179.2019.9105317.
- [7] C. K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 411-415, doi: 10.1109/ICAIT47043.2019.8987258.
- [8] Shetty, Shruthi S., et al. "Fake News Detection Using Naive Bayes and Support Vector Machine Algorithm."
- [9] Sharma, Karishma, et al. "Combating fake news: A survey on identification and mitigation techniques." ACM Transactions on Intelligent Systems and Technology (TIST) 10.3 (2019).
- [10] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1).
- [11] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19.1 (2017).

- [12] Khanam Z. and Agarwal S. Map-reduce implementations: Survey and Performance comparison, International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 4, August 2015.
- [13] Zhang, Jiawei, Bowen Dong, and S. Yu Philip. "Fake detector: Effective fake news detection with deep diffusive neural network." 2020 IEEE 36th International Conference on Data Engineering(ICDE). IEEE, 2020.
- [14] K Ludwig, M Creation 2020 "Dissemination and uptake of fake-quotes in lay political discourse on Facebook and Twitter" J. Pragmat, 157, 101–118.
- [15] Cardoso Durier da Silva, F., Vieira, R., & Garcia, A. C. (2019, January). Can machines learn to detect fake news? a survey focused on social media. In Proceedings of the 52nd Hawaii International Conference on System Sciences.
- [16] Bovet, Alexandre, and Hernán A. Makse. "Influence of fake news in Twitter during the 2016 US presidential election." Nature communications 10.1 (2019): 1-14. The science of fake news.
- [17] <https://science.sciencemag.org/content/359/6380/1094.summary> Science 09 Mar 2018:Vol. 359, Issue 6380, pp. 1094-1096 DOI: 10.1126/science.aao2998.
- [18] Barbara Probierz, Piotr Stefanski, Jan Kozak "Rapid detection of fake news based on machine learning methods" 25th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. *Barbara Probierz et al. / Procedia Computer Science 192 (2021) 2893–2902*
- [18] IJSREM Sanjivaneer R. Kale *1, Kartik D. Borkar*2, Mayur P. Bokad*3, Ayush G. Joshi*4, Himanshu A. Wadhonkar*5 "FAKE NEWS DETECTION USING MACHINE LEARNING" Volume: 06 Issue: 03 | March - 2022