

BITS F464 (MACHINE LEARNING)

# FAKE NEWS DETECTION PROJECT REPORT

---



## Submitted By:

1. Akshat Goyal : 2018B3A70864H
  2. Geet Sahitya Dulam : 2018B4A70797H
  3. Isha Gohel : 2018B4A70454H
  4. Nived Damodaran : 2018B5A71003H
  5. Ritika Garg : 2018B3A70258H
  6. Yash Amin : 2018B5A71044H
-

## **Acknowledgement**

We would like to express our sincerest gratitude to Dr. Paresh Saxena, for giving us an opportunity to work under him for this project and taking his valuable time to provide us the required guidance wherever required. His input proved to be very vital for the project. We would like to thank him for providing us with such a wonderful opportunity to apply our course knowledge on real life data and get hands-on experience. We are indebted for all his help and guidance throughout the course and this assignment.

**Birla Institute of Technology and Science-Pilani,  
Hyderabad Campus**

---



---

## Contents

<b>Introduction</b>	<b>4</b>
<b>EXPERIMENTS CONDUCTED SO FAR</b>	<b>5</b>
PAPER USED	5
RESULTS FROM THE PAPER	6
<b>RESULTS THAT WE GOT</b>	<b>6</b>
<b>PROPOSED CHANGES</b>	<b>8</b>
<b>WORK DISTRIBUTION BETWEEN GROUP MEMBERS</b>	<b>9</b>



---

## Introduction

“The world is more connected than ever”- this is the sentence we have heard repeatedly from international organizations to famous politicians, but with this statement comes an underlying problem of “Fake” news

Fake news is false or misleading information presented as news. A US government survey of 2019 reported that 89% of the respondents didn't know that the news was fake when a fake news was purposely spread and back traced to get the extent of the problem. Ever since then, the debate around fake news has heated up, laws are getting stricter and firms are spreading big bucks to solve this problem, making this topic an interesting and important one, and hence we are presenting our part of analysis for the problem.

**Link to the paper implemented:** [click here](#)

**Paper Title:** Multiclass Fake news Detection Using Ensemble Machine learning;

**Authors:** Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang

## **EXPERIMENTS CONDUCTED SO FAR**

### **PAPER USED :**

Fake news on social media is widespread and has become a matter of concern as it has the ability to cause a lot of social and national damage. This paper analyzes the research related to fake news detection and explores commonly used machine learning models in order to create a model with the best supervised machine learning algorithm that can classify fake news as "real" or "fake". We do so by making use of tools like python scikit-learn. This process results in feature extraction and vectorization. Then, we implement various machine learning algorithms to choose the classifier with maximum accuracy on the basis of results obtained from confusion matrix results.

### **RESULTS FROM THE PAPER :**

In the paper various machine learning algorithms have been implemented including Random Forest Classifier, Naive Bayes, Gradient Boosting Classifier, Decision tree, Logistic Regression and Linear SVM. According to the paper, the maximum accuracy of 86% is achieved for the Gradient Boosting Classifier.

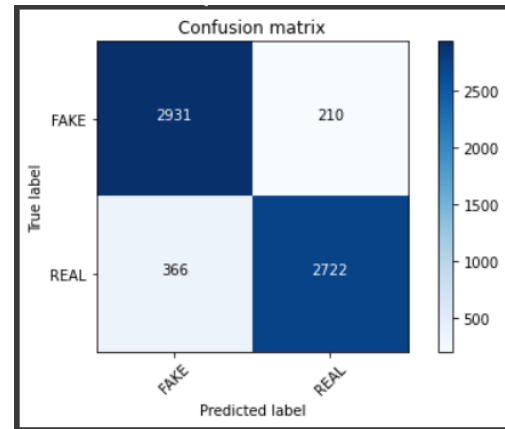
We have done a comparative analysis of the six machine learning algorithms and the results have been shown below.

## RESULTS THAT WE GOT

**Naive Bayes:** Accuracy - 90.8%

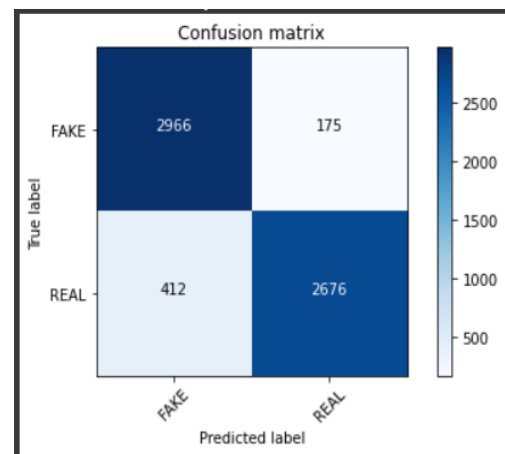
About: Theorem is based in Bayes Theorem.

Naive Bayes classifiers are a family of probabilistic classifiers based on the Bayes Theorem.



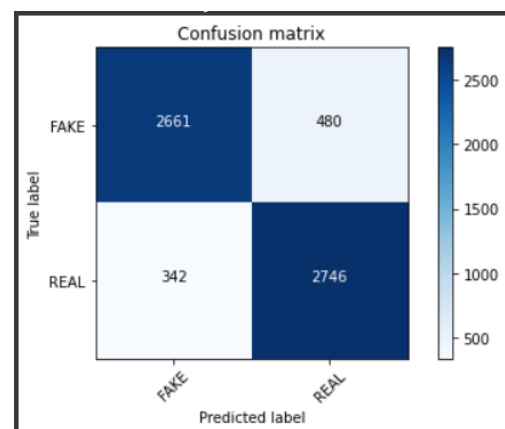
**Random Forest:** Accuracy-90.6%

Random Forest is an ensemble machine learning method which uses a multitude of decision trees at training time.



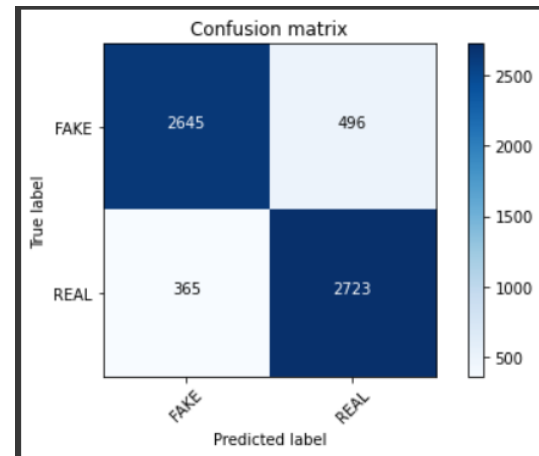
**Gradient Boost:** Accuracy-87.7%

It gives a prediction in the form of an ensemble of weak prediction models, usually decision trees.



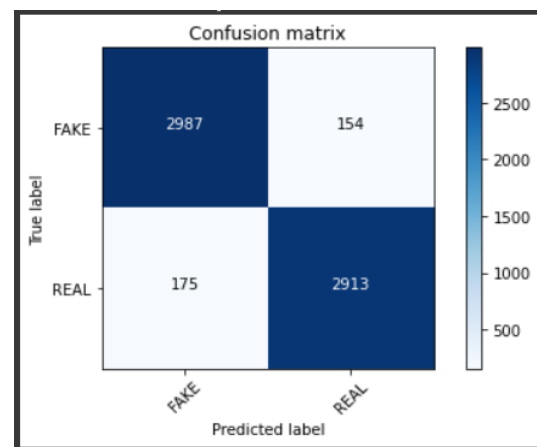
### Decision Tree: Accuracy-86.2%

It uses a tree to go from observations about a data item to conclusion about its target value.



### Logistic: Accuracy-94.7%

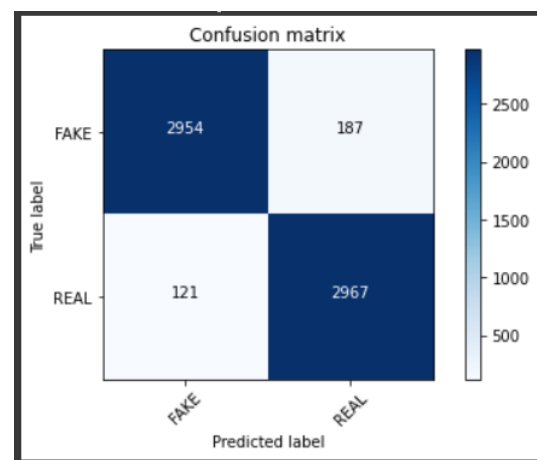
It is a binary model that models probability of one event given two alternatives.



### SVC: Accuracy-95.1%

SVC is an acronym for Support Vector Classifier, its objective is to fit to the data that we use.

It is an unsupervised algorithm which attempts to find natural clustering of the data into groups.



## PROPOSED CHANGES

**Preprocessing** : Preprocessing the text data by removing special symbols, removal of stopwords and converting the word to its root form(stemming). All text data needs to be cleaned before building the models.

Results before/after preprocessing:

Classifier	Before Preprocessing Accuracy	After Preprocessing Accuracy
Naive Bayes	90.3	90.8
Random Forest	90.1	90.6
Gradient Boost	86.6	87.7
Decision Tree	88.1	86.2
Logistic	94.7	94.7
SVC	95.1	95.1

Conclusion : Preprocessing did not have a huge impact on the accuracy of our models. One of the reasons can be that the data is already cleaned.

**Cross-Validation:** We used K-Fold cross validation with K=5. Our models were then run on this data and the results were obtained.

Classifier	Accuracy
Naive Bayes	91.19
Random Forest	89.91
Gradient Boost	88.05
Decision Tree	87.74
Logistic	94.66
SVC	94.87

An improvement of 1% was seen in Naive Bayes and Decision Tree models.





---

## WORK DISTRIBUTION BETWEEN GROUP MEMBERS

Source Code: [Zeudon/Fake News Classification: Fake News classification project done as part of the ML Course. \(github.com\)](https://github.com/Zeudon/Fake_News_Classification)

Student Name	Contribution
Akshat Goyal	Improvements in paper, Project report
Geet Sahitya Dulam	Analysis of selected paper, Selection of papers
Isha Gohel	Improvements in paper, Project report
Nived Damodaran	Analysis of selected paper, Selection of papers
Ritika Garg	Improvements in paper, Project report
Yash Amin	Analysis of selected paper, Project report