



**MIT**

Academy of  
Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

# Fake New Detection

Aditya Karpe (166)

Saurabh Keskar (114)



# Table Of Content

- Algorithms Used
- Libraries
- Exploring the Datasets
- Outputs

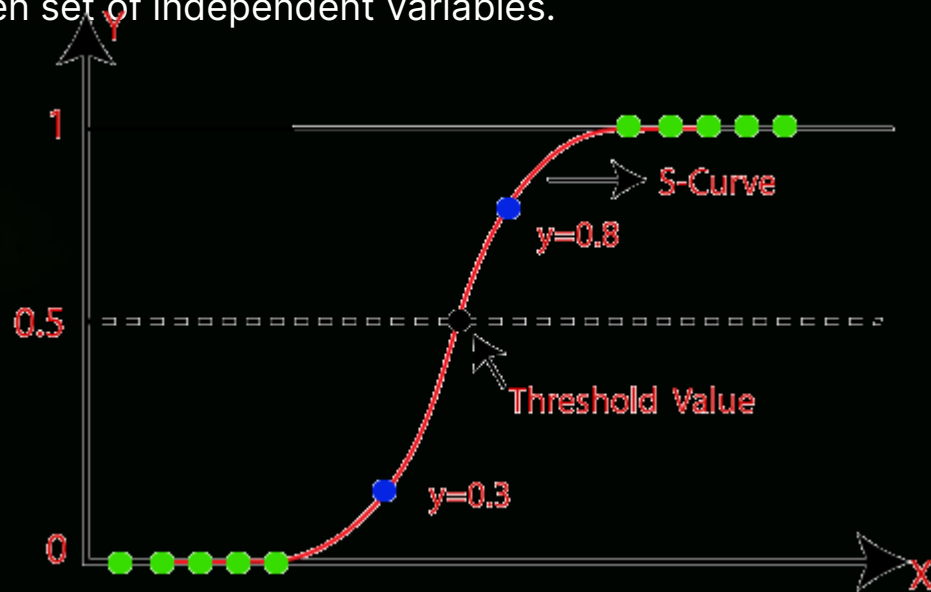


# Algorithm Used

- Logistic Regression
- Decision Tree
- Random Forest
- SVM(Support Vector Machine)

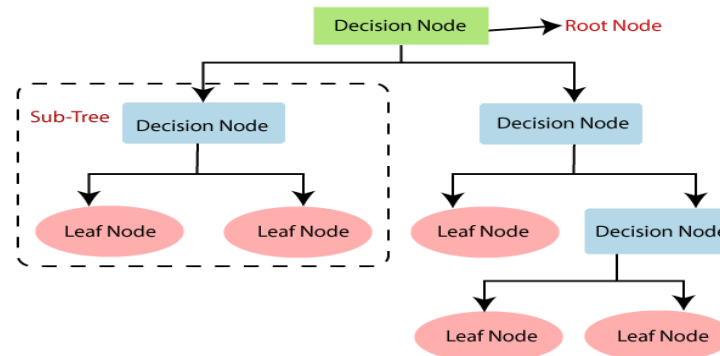
# Logistic Regression

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.



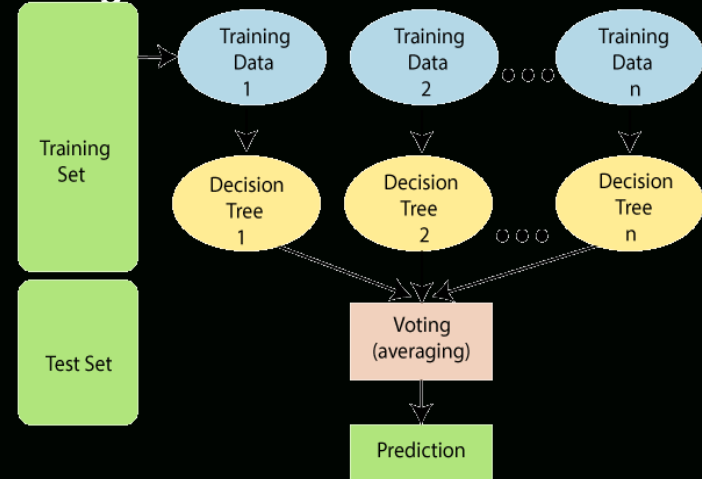
# Decision Tree

- Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome**.
- In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node**.



# Random Forest

- *Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.*
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

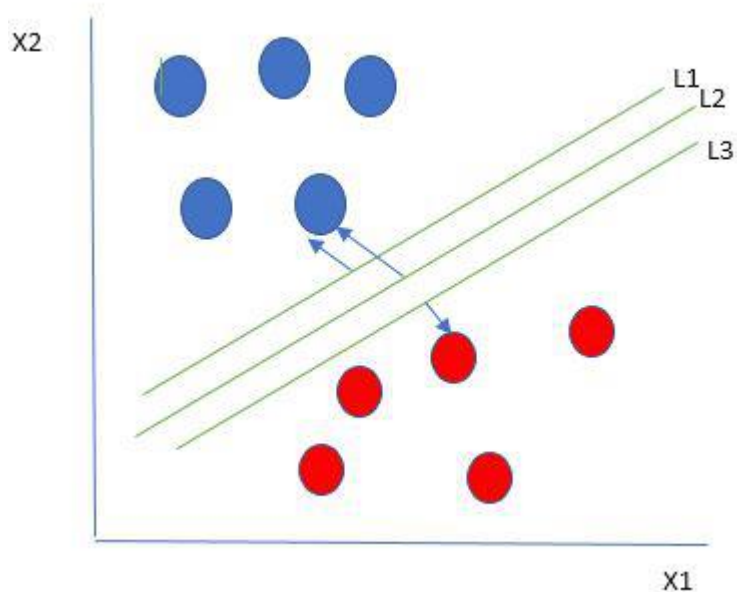


# SVM

SVM is also called Support Vector Machine.

If the number of input features is two, then the hyperplane is just a line.

If the number of input features is three, then the hyperplane becomes a 2-D plane.





# Libraries

- `import pandas as pd`
- `import numpy as np`
- `import matplotlib.pyplot as plt`
- `import seaborn as sns`
- `from sklearn.feature_extraction.text import CountVectorizer`
- `from sklearn.feature_extraction.text import TfidfTransformer`
- `from sklearn import feature_extraction, linear_model, model_selection, preprocessing`
- `from sklearn.metrics import accuracy_score`
- `from sklearn.model_selection import train_test_split`
- `from sklearn.pipeline import Pipeline`





# Exploring The Data Sets

## **Fake News Data Sets**

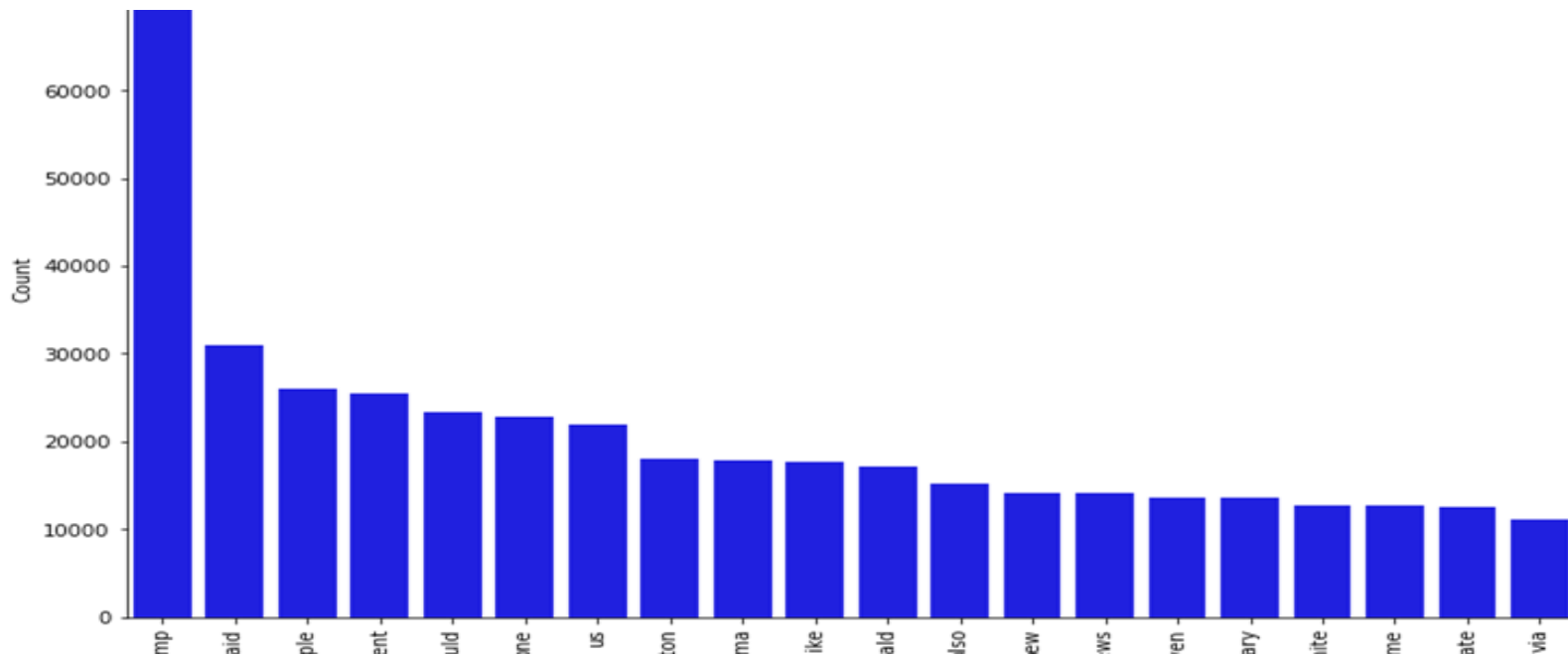
- Title - Heading of news
- Text - whole news
- Subject – Region news
- Date – Date of news

## **True News Data Sets**

- Title – Heading of news
- Text – whole news
- Subject – Category  
(politics, worldwide,etc)
- Date – date of news

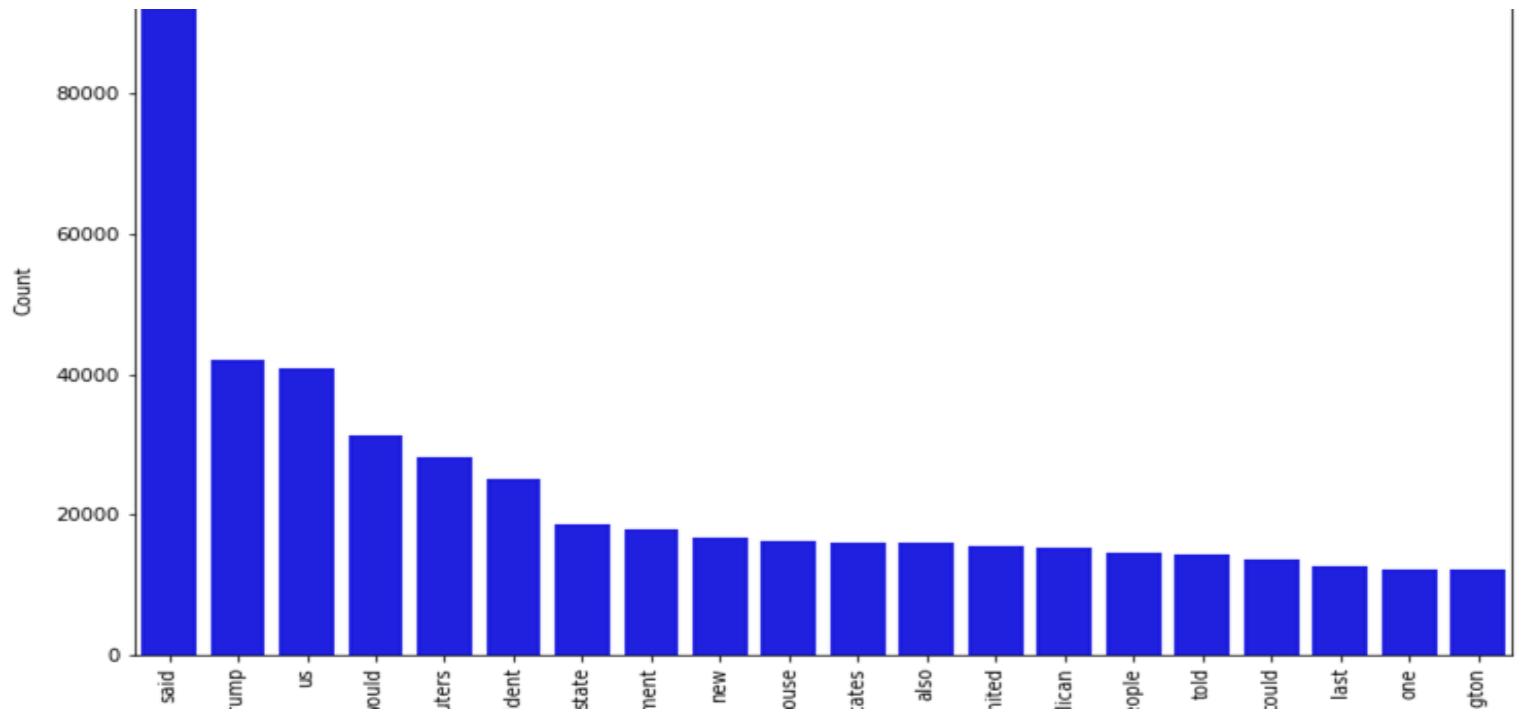


## Most Frequent words in Fake News



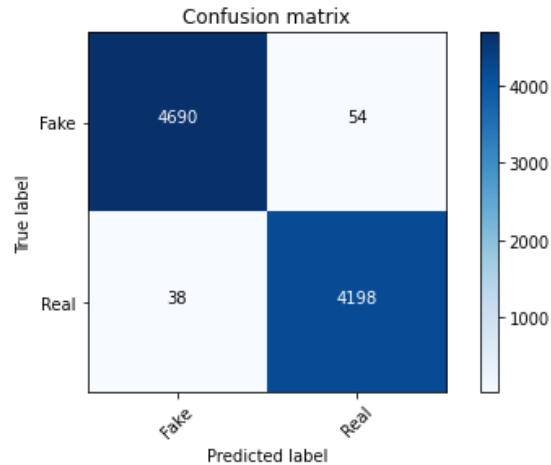


# Most Frequent words in True News

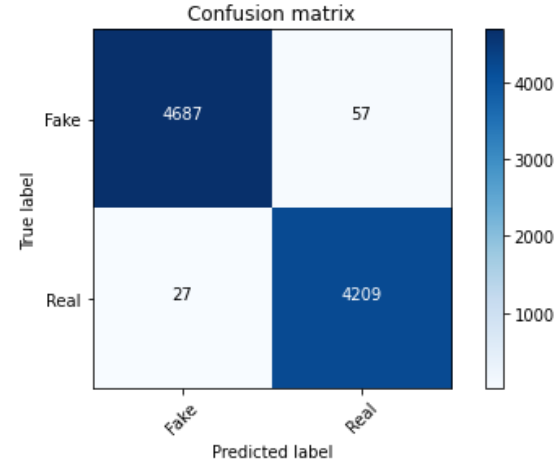


# Confusion Matrix

**Confusion matrix without normalization in logistic regression**

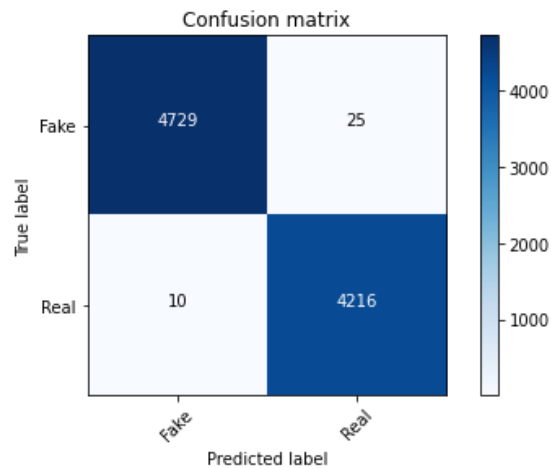


**Confusion matrix without normalization in Random Forest**

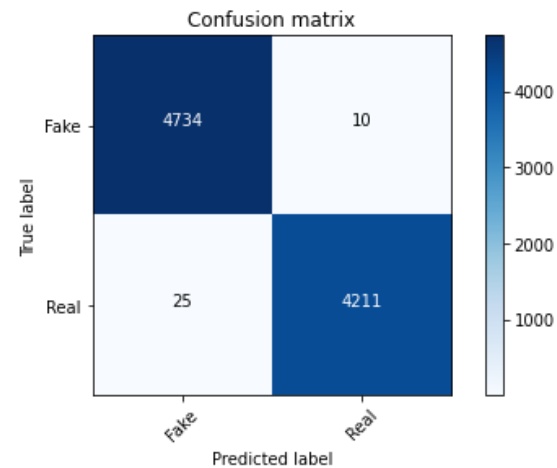


# Confusion Matrix

**Confusion matrix without normalization in SVM**

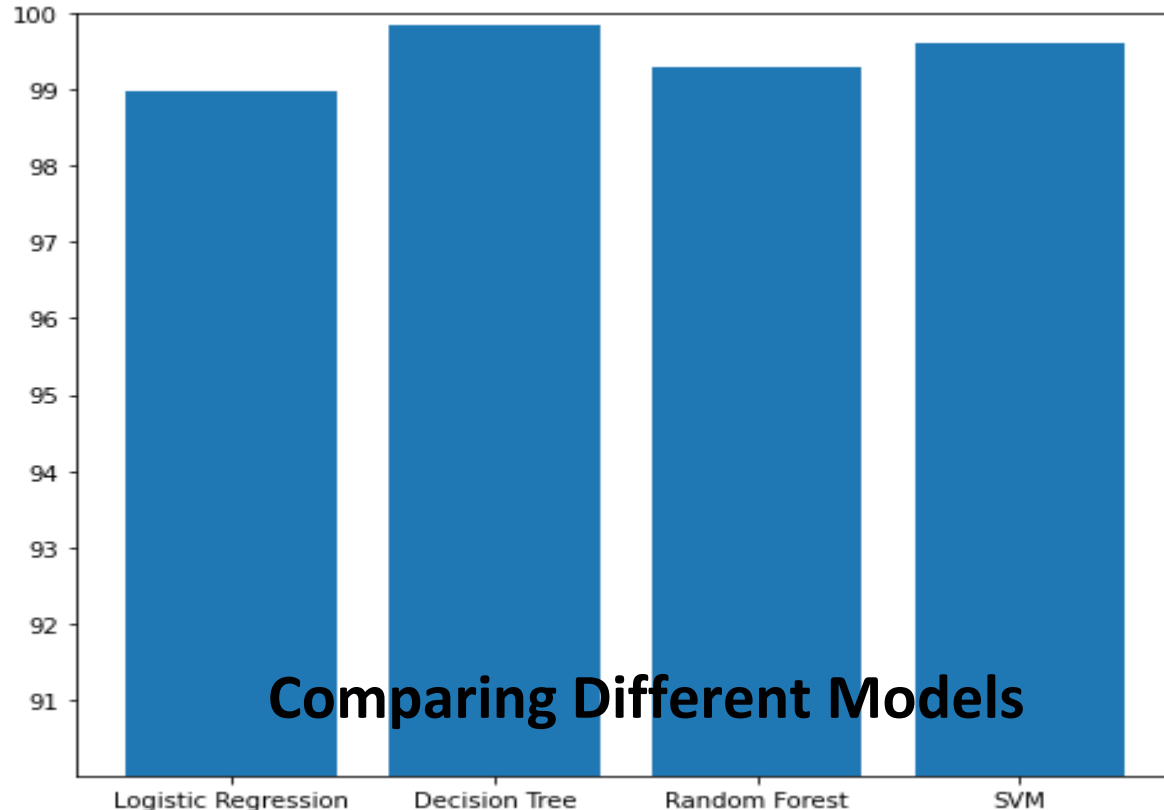


**Confusion matrix without normalization in Decision Tree**





## Comparing Different Models





Thank You...