Fake News Detection using Naive Bayes classifier

Table of Content

Introduction

The ease of gaining news through social media has been ever more convenient than reliable. Because humans value the convenience of distributed information, these news aren’t verified or challenged. Resulting, in people acting on fake news and further repeat this cycle of misinformation as they continue to spread false information.

Motivation

1. Hot Topics/Issues
   a) Covid-19 [6]
   b) Russia vs Ukraine War
2. Concepts that are important to specific populations or topics of interest
   a) Political views
   b) Religious views
3. Uninformed on how to verifying sources

Problem statement

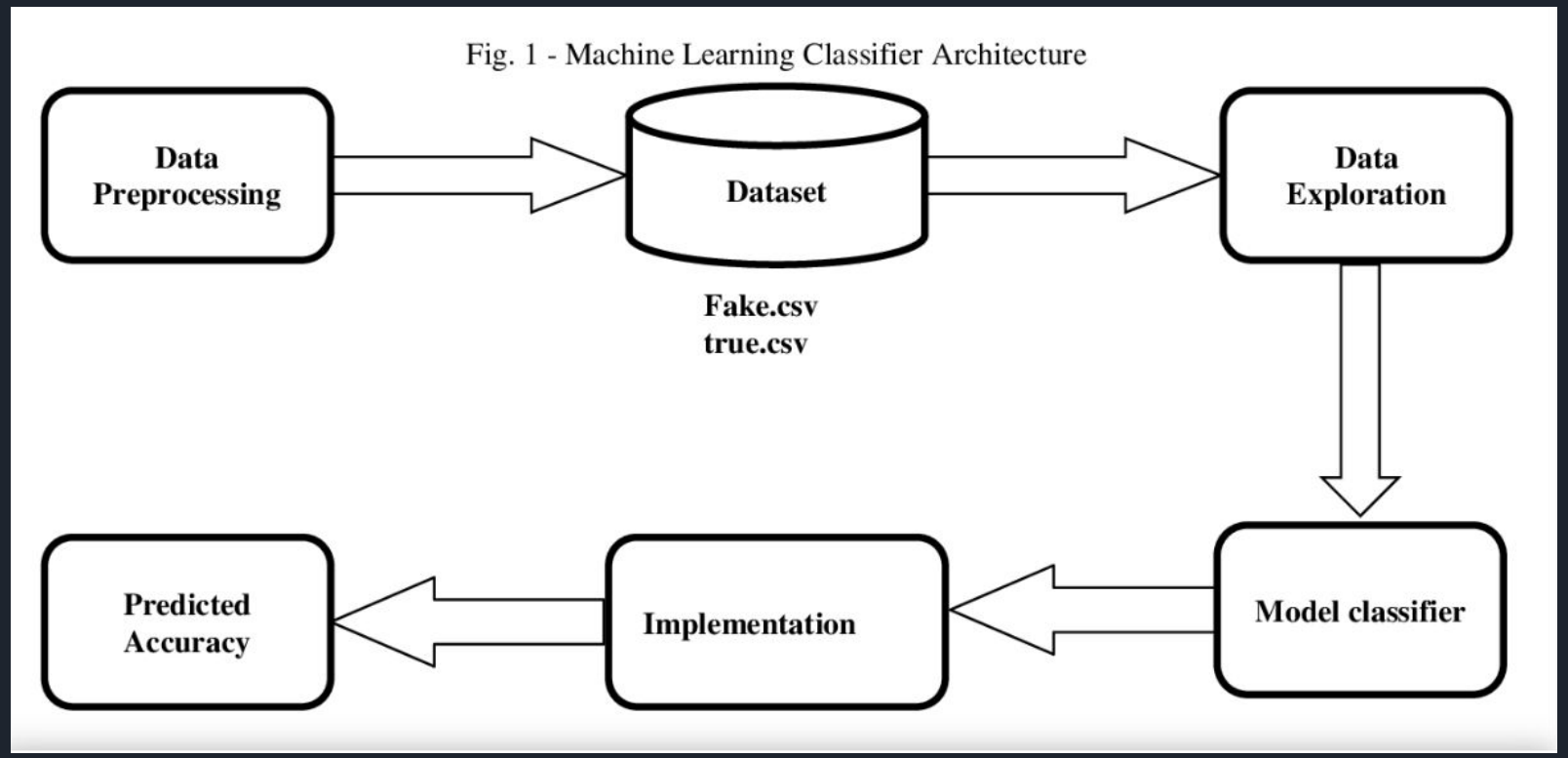
The problem at hand is to design a system that can accurately identify fake news articles from a given dataset. The objective is to employ the Naive Bayes classifier, a popular machine learning algorithm, to classify news articles as genuine or fake based on a set of predefined features. By leveraging the principles of probability theory and the independence assumption, the Naive Bayes classifier can provide a probabilistic assessment of the likelihood of an article being fake.

Proposed Work

Data Pre processing:

1. Data Collection:
2. Data Pre processing :filter and manipulate columns, adding labels (fake or real), and merge on selected columns
3. Pre processing the text: NLP Approach Text to features conversion: data cleaning ,removing stop-words, removing punctuation
4. Model Implementation :Naive Bayes
5. Check on Model Performance: Classification Report

System overview



Higher Classification of Fake Political News Using Decision Tree Algorithm Over Naive Bayes AlgorithmT. Dinesh, Dr.T. Rajendran

Implementation: Naive Bayes classifier

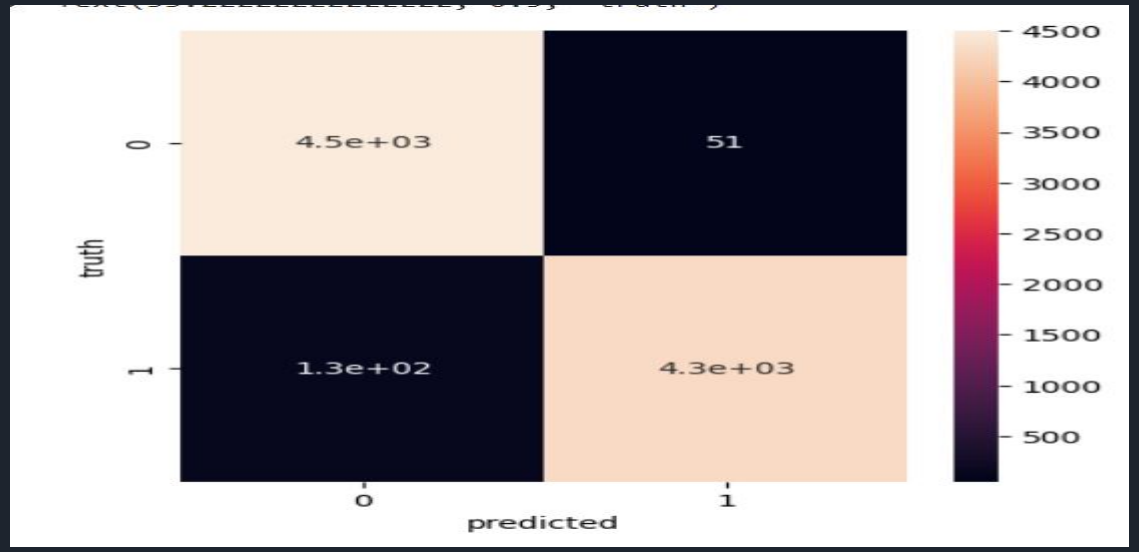
Classification technique based on Bayes' theorem with an assumption of independence among predictors

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

1. Convert data set into a frequency table
2. Create likelihood table by finding probabilities
3. Use Naive Bayesian equation to calculate posterior probability for each class

Result

Confusion matrix: is used to define the performance of a classification algorithm.




```
Confusion Matrix
[[4505  51]
[ 127 4297]]
```

Conclusion

The Naive Bayes classifier is a popular and widely used algorithm due to its simplicity, speed, and good performance in many text classification tasks. However, for more advanced NLP problems like fake news detection that require capturing complex linguistic relationships, other algorithms like SVM,Random Forest,LSTM,recurrent neural networks (RNNs), convolutional neural networks (CNNs), or transformer-based models may be more suitable.

Limitations

- Independence assumption
- Limited handling of out-of-vocabulary words
- Difficulty with handling large feature spaces

Future scope

a)we can use image,video as well as audio data
b)We could compare various other machine learning algorithms.
c) Testing the proposed method in this paper on a larger dataset to check for accuracy and problems associated.
d) Can improve the webpage by using animations and wallpapers and making it more attractive.
e) After the successful implementation and removing all problems, can try for making this in the form of a mobile app.

Thank you