FLIP ROBO

FAKE NEWS DETECTION PROJECT

Submitted By:

Lakshmi Rajendra Thute

ACKNOWLEDGEMENT

I am very much Thankful to FlipRobo Technologies for giving me the opportunity to work with them and to work on this project and also, I am very grateful to Data Trained Education Team for their support and help to understand each and every concept of machine learning which helped me a lot while working on this project. I thought, I am fortunate to become a part of FlipRobo Technology.

References:

Google website

Stack overflow

Analytics Vidya

Medium

Data trained notes

INTRODUCTION

Fake news's simple meaning is to incorporate information that leads people to the wrong path. Nowadays fake news spreading like water and people share this information without verifying it. This is often done to further or impose certain ideas and is often achieved with political agendas.

For media outlets, the ability to attract viewers to their websites is necessary to generate online advertising revenue. So, it is necessary to detect fake news.

ANALYTICAL PROBLEM FRAMING

We got the dataset available from client, and we must build a model that is able to detect the fake news. We have data in two different sets, that are fake and true. We have to combine both of them together to build a mode, after combining the size of dataset becomes 44898. The dataset contains total 5 columns, they are title, text, subject, date, isfake.

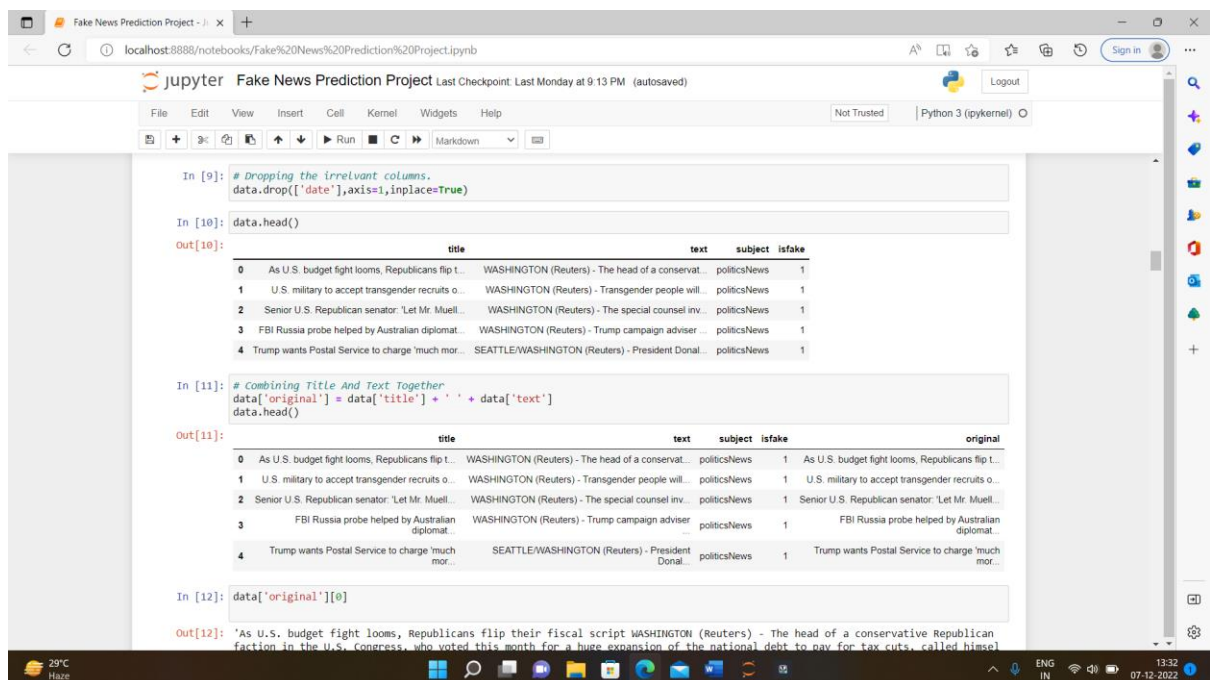## • Mathematical/ Analytical Modelling of the Problem

Here I firstly read the dataset in jypyter notebook for cleaning the dataset I use pre-processing techniques, then did the Exploratory Data Analysis, then Encoding and lastly model Building and Evaluation.

• Data Sources and their formats

I got the dataset in CSV format, and I read the data in Jupyter Notebook using pandas data frame.

• Data Pre-processing Done

The dataset contains some irrelevant columns firstly we dropped that columns, then we add some important **columns as they are need for further procedure.**

• Hardware and Software Requirements and Tools Used

Here for this project, I used Jupyter notebook and libraries such as pandas and NumPy for mathematical operations, matplotlib and seaborn for various type of data visualizations and to explore and for better understanding of the dataset. Also, we use NLP for analysing huge number of unstructured data.

• Identification of possible problem-solving approaches (methods)

The statistical summary shows the total count of 44898 rows then mean, min value, max value, standard deviation, and quartiles shows up and down.
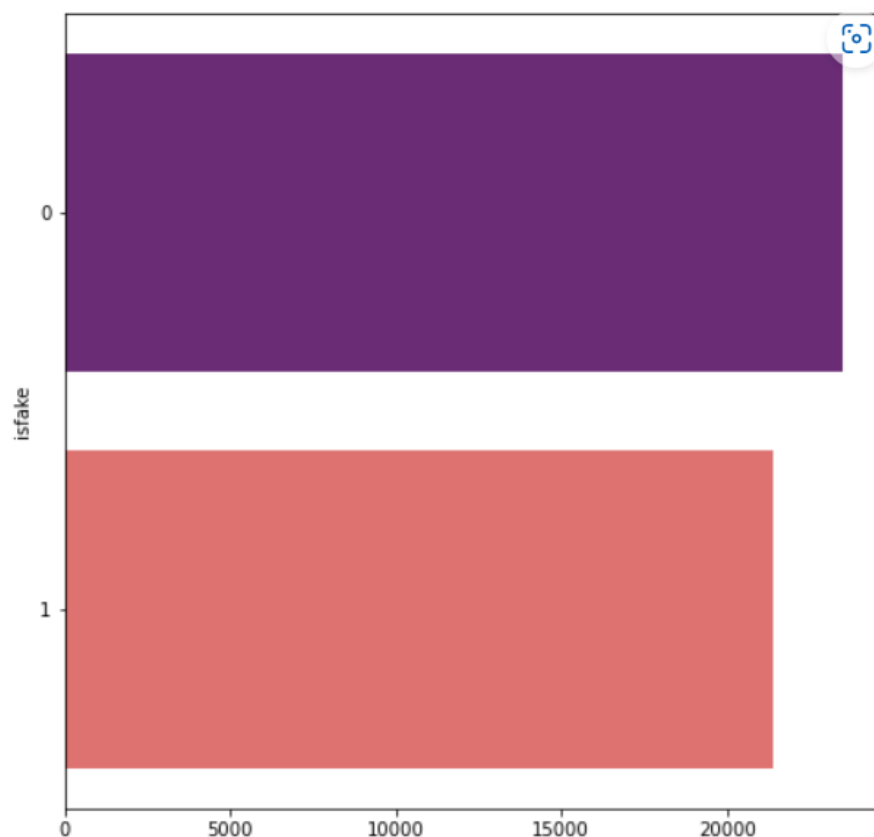
• Key Metrics for success in solving problem under consideration

A company's units can use these dashboards to create milestones and monitor their progress by tracking all the most relevant metrics in one location.

• Visualizations

```python
In [35]: plt.figure(figsize = (8, 8))
         sns.countplot(y = "isfake", data = data,palette='magma')
```
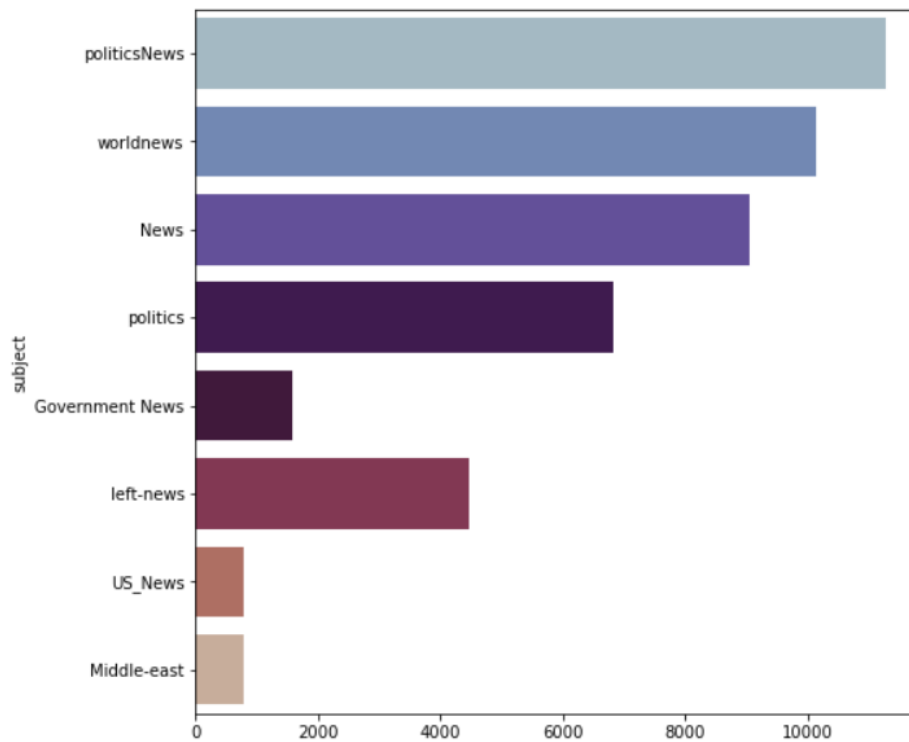
```
Out[35]: <AxesSubplot:xlabel='count', ylabel='isfake'>
```

```
In [33]: # Plotting The Number Of Samples In "Subject"
         plt.figure(figsize = (8, 8))
         sns.countplot(y = "subject", data = data,palette='twilight')
```

Out[33]: <AxesSubplot:xlabel='count', ylabel='subject'>

- **Observations**
- From the above bar graphs, we observe that the number of fake news are more as compared to true news.
- The greatest number of news include political news, then world wide news and least no. of news are of middle-east and US-News.

- **Interpretation Of Results**
  After Pre-processing, Exploratory Data Analysis and model building, we observed that there is more no. of news that are inspired form true events that means true news.