

We loaded the Fake news & True news data.

Then we added the label column to both dataset

Label for fake news - 0

Label for real news - 1

Then we merged both datasets and did Exploratory Data Analysis

The primary focus is over the content of the news and the subject of it.

So we can drop the columns like "title" and "date" to have relevant data only.

Next we checked for null values in the dataset and there is no null values to be filled.

Based on the subject of the news majority of the news are on politics.

We went ahead with converting the data into lower case and removing any special characters etc.

We did the process of Tokenisation, stemming & Lemmatization. Then we did the separation of features & labels. And converted the data into numeric so that model building can be carried out.

We split the data into training & testing and since it's a regression model we used Logistic Regression.

Logistic Regression gave us an accuracy score of 99%