PROJECT REPORT FAKE NEWS CLASSIFIER

Problem statement:

The authenticity of Information has become a longstanding issue affecting businesses and society, both for printed and digital media. On social networks, the reach and effects of information spread occur at such a fast pace and so amplified that distorted, inaccurate, or false information acquires a tremendous potential to cause real-world impacts, within minutes, for millions of users. Recently, several public concerns about this problem and some approaches to mitigate the problem were expressed.

In this project, you are given a dataset in the fake-news_data.zip folder. The folder contains a CSV files train_news.csv and you have to use the train_news.csv data to build a model to predict whether a news is fake or not fake. You have to try out different models on the dataset, evaluate their performance, and finally report the best model you got on the data and its performance.

Data- Description:

There are 6 columns in the dataset provided to you. The description of each of the column is given below:

"id": Unique id of each news article

"headline": It is the title of the news.

"news": It contains the full text of the news article

"Unnamed:0": It is a serial number

"written_by": It represents the author of the news article

"label": It tells whether the news is fake (1) or not fake (0).

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

Total value of rows and columns (20800,6)

After dropping nan values real dataset is (18825,6)

By Dropping Label column the dataset reamin (18825,5)

- Data Sources and their formats

Kaggle is the data source for the fake news classifier detection .

The format of data solving is data preprocessing and data wrangling.

After this steps we will use EDA and finally Model Evaluation.

- Data Preprocessing Done

Data preprocessing is done by step by step.

- Data Inputs- Logic- Output Relationships

Data with normalization and without normalization in fake news classifier dataset.

The hyperparameter will be used for best results after model deployment.

Model/s Development and Evaluation

MultinomialNB and multinomial classifier with hyperparameter which gives us a best result at 0.90%.

Passive aggressive algorithm which is also gives a accuracy of 0.90% best model deployment in fake news classifier.

CONCLUSION

The fake news classifier project concluded that there is more real and fake news available by using best model evalution in natural language processing we can detect the fake news as well as real news.

Best accuracy on MultinomialNB is 90%.