FAKE NEWS DETECTION

PHASE-4

INTODUCTION

▶ *Imagine a scenario where a false news story spreads rapidly on social media, claiming that a particular medication is a cure for a deadly disease. People start hoarding the medication, causing scarcity and preventing those who need it from accessing it. This example scenario shows one of the several real-world risks of fake news.*

➢ **Detecting fake news using machine learning techniques would mean having an automatic detection system that looks at a piece of text (tweets, news articles, a WhatsApp message) and determine how likely it looks like a piece of false news.**

Table of Contents

Top 5 Fake News Detection

- **FakeNewsNet: Dataset of Political and Gossip Tweets**
- **Fake News Corpus**
- **FakeHealth**
- **Constraint COVID-19 Fake News Dataset**
- **FNC-1 (Fake News Challenge Stage 1)**

1.FakeNewsNet: Dataset of Political and Gossip Tweets

This [dataset](#) uses the Twitter official API to fetch tweets from Twitter users, including metadata and social context. But to begin with, they provide a sizeable amount of clean data that can be readily used to test models.

2.Fake News Corpus

This open-source [dataset](#) comprises millions of news articles scraped from a curated list of 1001 domains. The dataset includes over 9,400,000 articles from over 700 domains scraped from multiple domains, such as NYTimes and WebHose English News Articles.

3.FakeHealth

FakeHealth is collected to address Fake Health News detection challenges from sources like news content, reviews, and social media. As proposed in the paper, "Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository," the data can be fetched from Twitter using the Twitter API and the IDs provided in the GitHub repository- EnyanDai/FakeHealth.

4.Constraint COVID-19 Fake News Dataset

This dataset contains social media posts related to COVID-19 and vaccinations from various popular platforms. The posts are annotated as real and fake news, and the dataset can be better understood in its paper 'Fighting an Infodemic: COVID-19 Fake News Dataset' and the data, as well as the proposed baseline benchmark using attention models and heuristic-based post-processing

5.FNC-1 (Fake News Challenge Stage 1)

FNC-1 was designed as a stance detection dataset containing 75,385 labeled headline and article pairs. The pairs are either agree, disagree, discuss, or unrelated. Moreover, the baseline estimates the stance of a text from a news article relative to a headline. Specifically, the text may agree, disagree, discuss, or be unrelated to the headline. The FNC-1 dataset and the proposed baseline both take advantage of this.
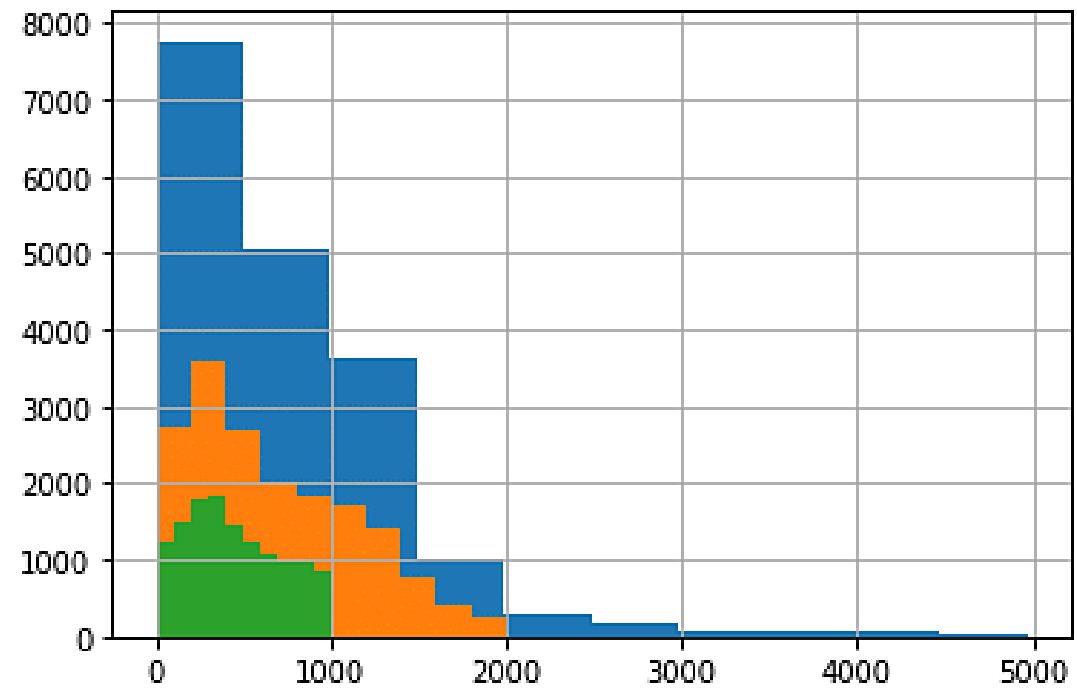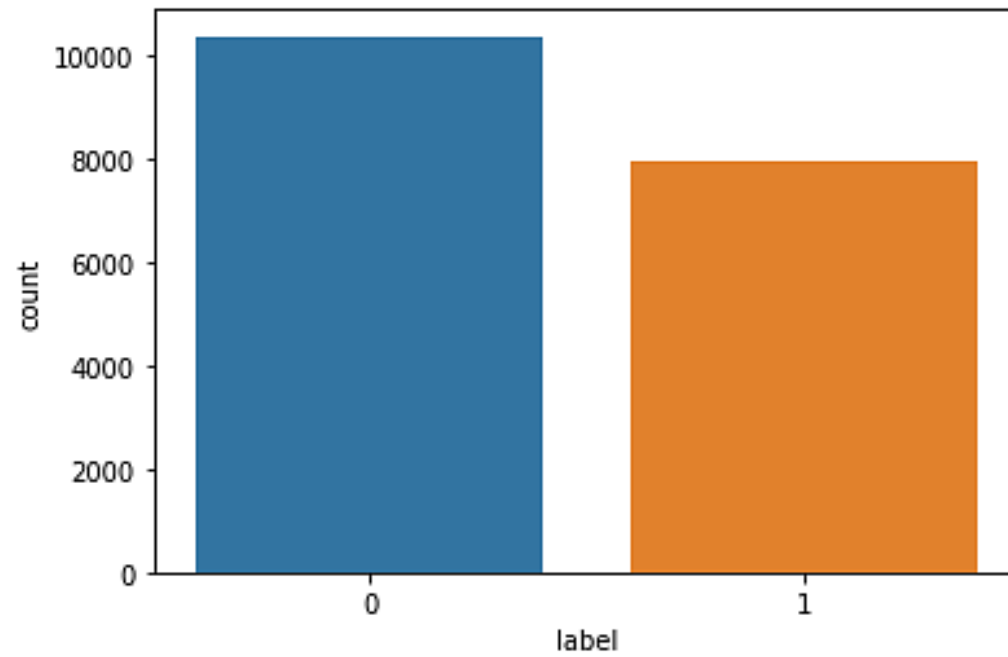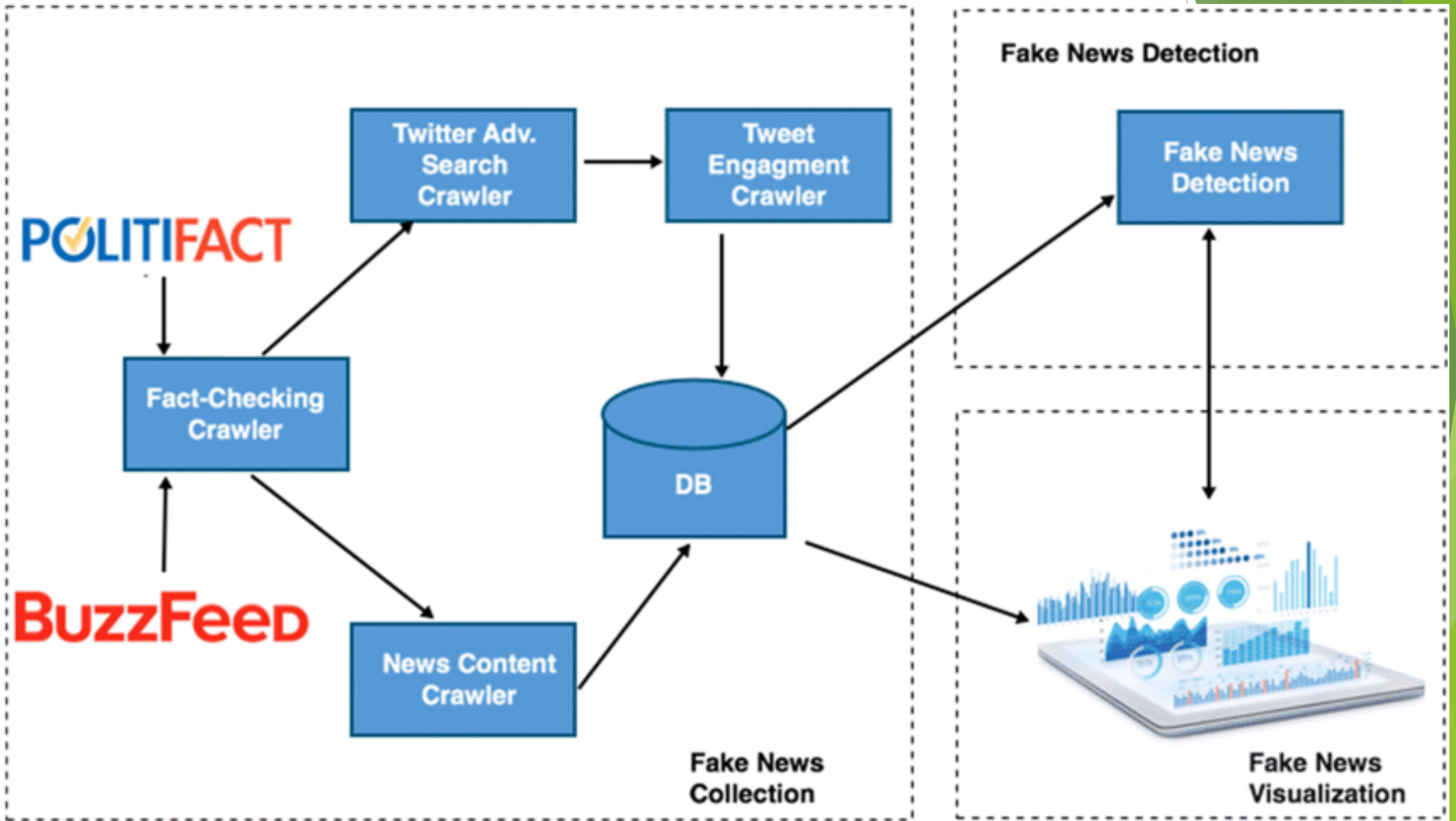
METHODOLOGY

The aim of this research is to create a fake news detection model utilising three machine learning techniques. Because the focus is on model building in machine learning using jupyter notebook, this isn't continually developing new usual package systems. Machine learning normally necessitates a significant amount of time for model training and testing, as well as a large volume of high-quality dataset. In other words, if the model yields predictable results, such as the prediction of fake and actual news, we might consider it to be fairly accurate.

❖ Data Administration

This section collects a body of knowledge (dataset), which could be a collection of report articles, stories, news, or blog postings. Once the dataset has been collected, nltk is used to identify a collection of written or spoken material stored on a computer and used to discover how language is utilised: the data is investigated to obtain a better understanding of its structure, which means stopwords are deleted.

❖ Data Exploration

The charting of graphs according to the fake and true news anticipated by the machine learning algorithm is the major focus of the information exploration section. Word clouds are created, which are essentially visual image ways for conveying text information in which the magnitude of each word represents its frequency or importance.

❖ Model Training

The machine learning model can then be trained after the data has been adequately analysed and controlled. During the Model Training phase, many methodologies are considered, and a learning problem that is a prediction task is determined. Whatever possibilities are available within the training data set are then investigated.

❖ Model Assessment

The output of the model generated is measured in many ways while evaluating it. The model's correctness is graded using performance measures such as F1 score, precision, recall, and accuracy rate, which are based on the confusion matrix report.

CONCLUSION & FUTURE SCOPE

Spreading false information has always had a harmful influence on society. When it comes to distinguishing between false and genuine news, there is still a lot of ambiguity in our culture. Fake news is a false alarm for anybody since it constantly misleads the audience, leaving them perplexed and unable to respond appropriately. They see their daily lives with their own eyes. So, this is when our research can be used to anticipate whether a given piece of news is false or not! People will be more conscious of fake news propagation if they consider the philosophy of our research report. This system was completed in the final year, but it will undoubtedly benefit from more enhancements in the near future, such as the use of a flask.

Disadvantages Of Detecting Fake News Using ML

1.Bias

2.Limited Accuracy

3.Limited Domain
Knowledge
4.Privacy Concerns

5.Maintenance

REFERENCES

1. https://www.researchgate.net/figure /Pseudocode-for-KNNclassification_fig7_260397165

2. https://www.analyticsvidhya.com/b log/2017/09/common-machinelearning-algorithms/