FAKE NEWS DETECTION USING NLP

CONTENT :

- ❑ Abstract
- ❑ Introduction
- ❑ Problem statement
- ❑ Data source
- ❑ Data preprocessing
- ❑ Feature extraction
- ❑ Model selection
- ❑ Model training
- ❑ Conclusion

ABSTRAC T :

In the age of digital media, fake news is a serious problem because it spreads misinformation and harms individuals, organizations, and even entire nations which is a challenging aspect. This study proposes a machine learning approach for detecting fake news. In the proposed approach, a categorization model is developed with four different types of machine learning algorithms, evaluating the content and aesthetic components of news stories. The performance of the proposed model is analyzed by using a large dataset of real and fake news articlesand the results show that it outperforms many existing systems. The proposed findings demonstrate the potential of machine learning techniques, such as logistic regression, decision tree, random forest, and neural networks algorithms to address the fake news detection challenges.

INTRODUCTION :

The deliberate spread of incorrect or misleading information through different media is referred to as fake news, also known as disinformation. Fake news has become a widespread issue with the rapid rise of the internet and social media,and it now poses a threat to society in many ways, including by inciting fear and distrust, influencing public opinion and decision-making, and even producingpolitical instability. Therefore, it has become crucial for governments, media outlets and individuals to identify and stop the spread of fake news.

EXAMPLE OF AN IMAGE :

International Research Journal of Engineering and Technology (IRJET)

www.irjet.net

Machine Learning and NLP based Fake News Detection

[illegible]

I. INTRODUCTION

[illegible]

II. LITERATURE SURVEY

[illegible]

III. BACKGROUND

1. Machine Learning

[illegible]

DESIGN THINKING

PROBLEM STATEMENT :

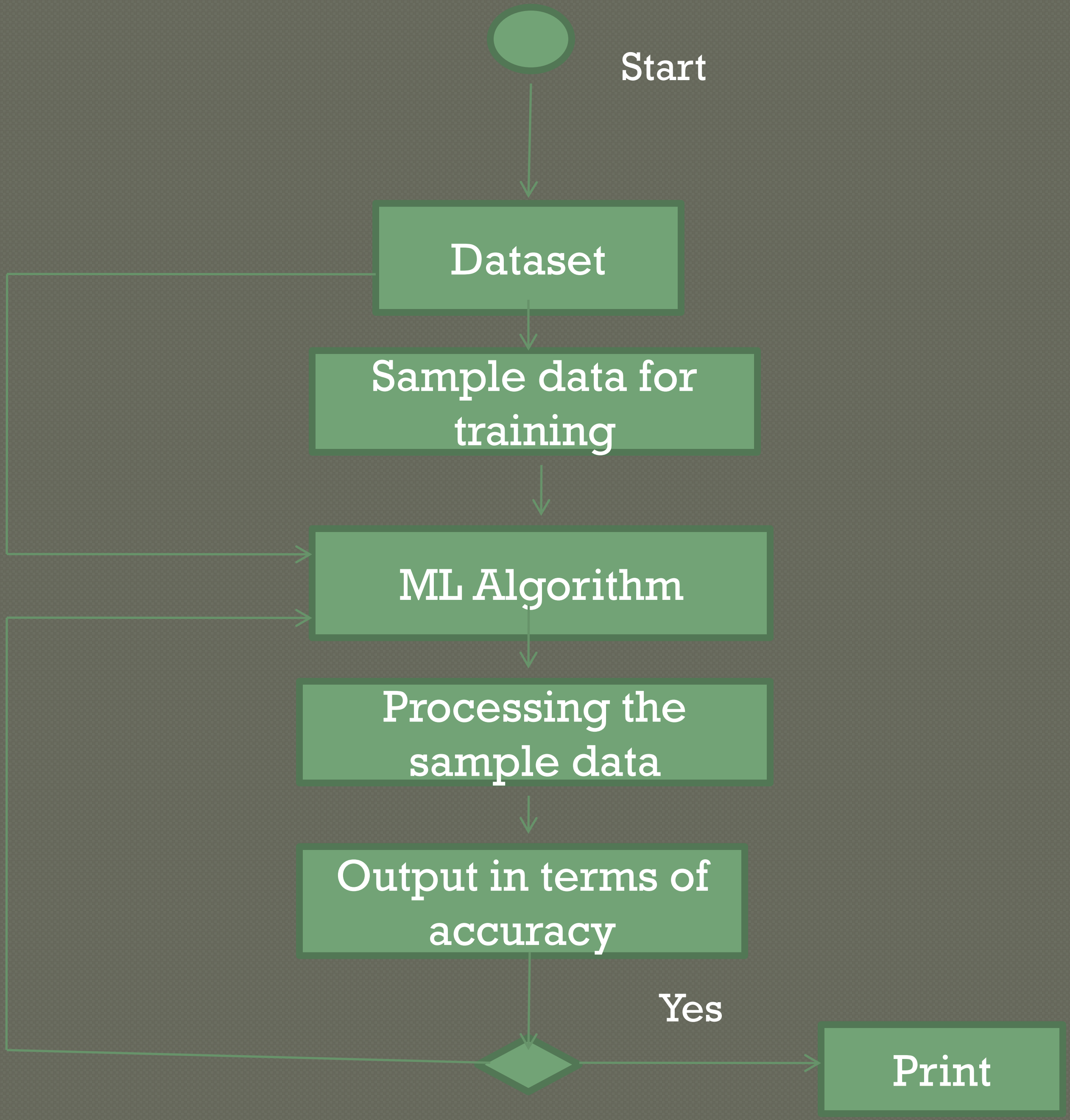
The fake news dataset is one of the classic text analytics datasets available on kaggle . It consists of genuine and fake articles' titles and text from different authors. Our job is to create a model which predicts whether a given news is
real or fake.

DEFINITION :

The problem is to develop a fake news detection model using a dataset givenThe goal is to distinguish between genuine and fake news articles based on theirtitles and text. This project involves using natural language processing (NLP) techniques to preprocess the text data, building a machine learning model for classification, and evaluating the model's performance.

DATA SOURCE :

Choose the fake news dataset available on Kaggle, containing articles titles and text, along with their labels (genuine or fake).



Start

Dataset

Sample data for training

ML Algorithm

Processing the sample data

Output in terms of accuracy

Yes

Print

DATA PREPROCESSING :

Clean and preprocess the textual data to prepare it for analysis.

The data undergoes data pre-processing, feature extraction, dimensionality reduction and finally the data is sent to the classification models i.e. Rocchio classification, Bagging classifier, Gradient boosting classifier and Passive Aggressive Classifier to train the model which is further used to detect the fake news.

VARIOUS TEXT PREPROCESSING STEPS ARE :

- Tokenization.
- Lower casing.
- Stop words removal.
- Stemming.
- Lemmatization

FEATURE EXTRACTIONL :

In natural language processing (NLP), feature extraction is a fundamental taskThat involves converting raw text data into a format that can be easily processed By machine learning algorithms . There are various techniques available for featureExtraction in NLP , each with its own strengths and weakness

FEATURE EXTRACTION METHODS

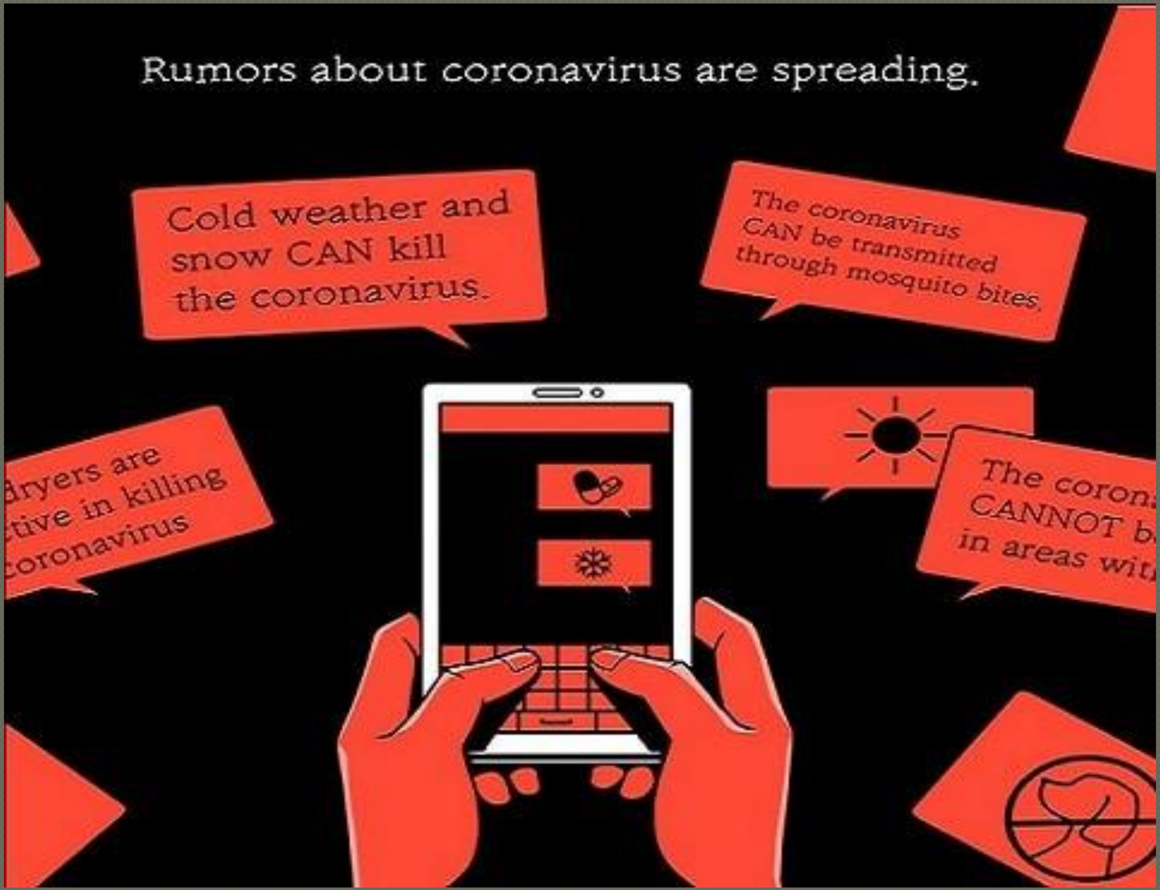
At present,

there are three typical feature extraction methods, namely

bag-of-words (BoW), word2vec (W2V) and large pre-trained natural language processing (NLP) models. BoW is widely used in traditional Machine learning.

MODEL SELECTION

Select a suitable classification algorithm (e.g., Logistic Regression, Random Forest, or Neural Networks) for the fake news detection task.


Rumors about coronavirus are spreading.
Cold weather and snow CAN kill the coronavirus.
The coronavirus CAN be transmitted through mosquito bites.
dryers are tive in killing coronavirus
The coron CANNOT b in areas wit

MODEL TRAINING

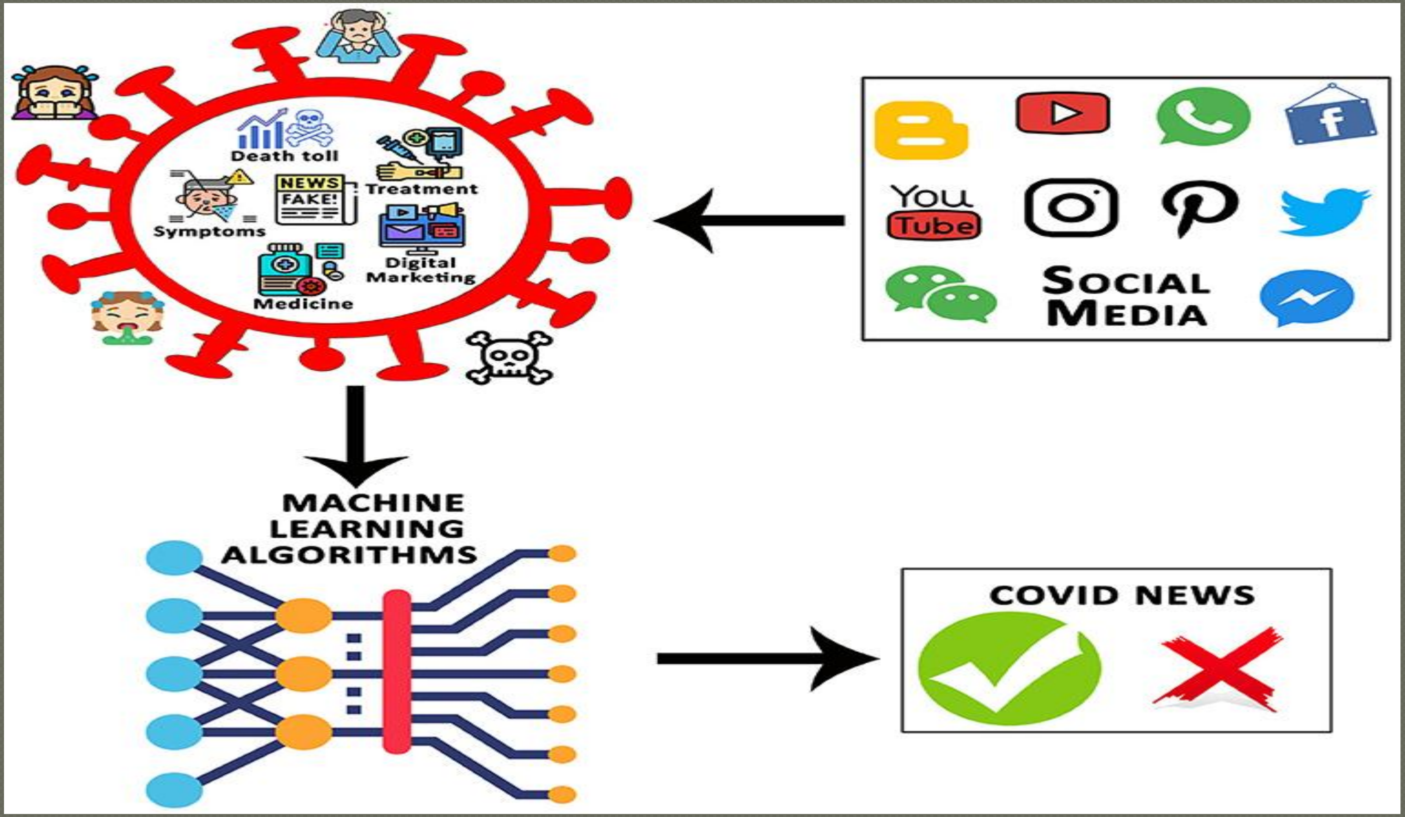
We can train our model now that we have preprocessed our text data . We will use a simple bag-of-words approach, representing each article as a vector of wordfrequencies. We will use the *Count Vectorizer* class from the *sklearn* library to convert the preprocessed text into feature vectors.

Count Vectorizer is a commonly used text reprocessing technique in natural languageprocessing. It transforms a collection of text documents into a matrix of word counts. Each row in the matrix represents a document, and each column represents a word in the document collection.

The Count Vectorizer converts a collection of text documents into a matrix of token counts. It works by first tokenizing the text into word sand then counting the frequency of each word in each document . The resulting matrix can be used as input to machinelearning algorithms for tasks such as text classification.

EVALUATION :

Evaluate the model's performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.



CONCLUSION :

Preprocessing is an essential step in natural languages processing tasks such as text classification, and techniques such as lowercasing, removing stop words, and stemming/lemmatizing can significantly improve the performance of models. Count Vectorizer is a powerful tool for converting text data into a numerical representation that can be used in machine learning models.

The project of fake news detcting using natural language Processing(NLP) at the of conclusion is the more number Text Connects true or fake news to identify correctness of that news