

Department of Computer Science and Engineering
(II Year) (2024-25)

Fake News Detection

using Python and machine learning

Project Domain: Machine Learning

S U B M I T T E D T O :
M S . A S H I M A Y A D A V

S U B M I T T E D B Y :
A N I S H S H A R M A
(E 2 3 C S E U 0 2 9 2)
A R Z O O R A I
(E 2 3 C S E U 2 4 1 9)



TABLE OF CONTENT

- Introduction
- Major Problem
- Purpose
- Work Flow of Project
- System Requirements
- Libraries Used
- Technologies Used
- Results Conclusion
- References
-

What is a Fake News?



- Fake News is a news designed to deliberately spread hoaxes, propaganda, and disinformation.
- Fake News Stories usually spread through social media sites like Facebook, Twitter, Instagram, etc.
- The wide spread distribution of fake news has the potential to have highly negative effects on people and culture.

Major Problem



- Fake news influences people's perception.
- Fake news are used as clickbait for the profit of content publishers. More clicks contributes more money to them.
- The rise of fake news has become a global problem that will impact people, culture as a whole and in general, doubt the authenticity of journalism as well

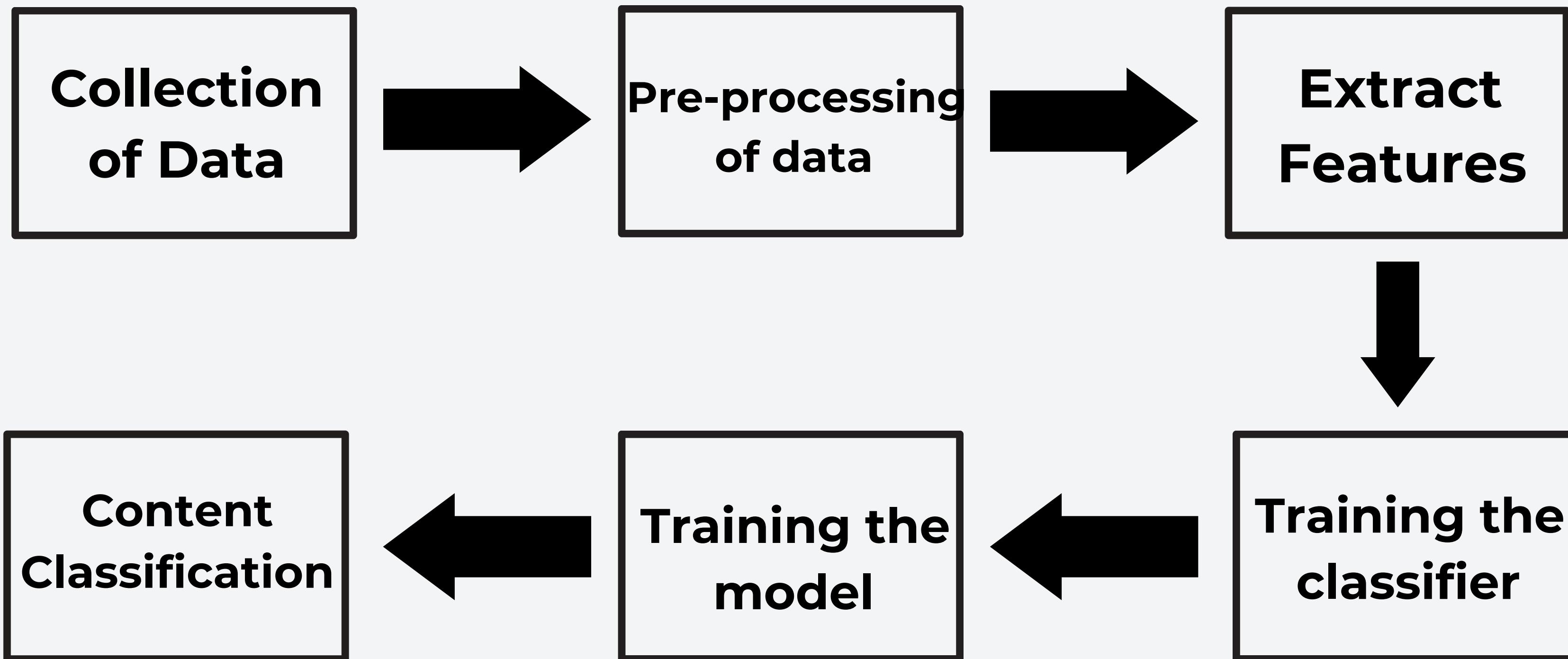
Purpose



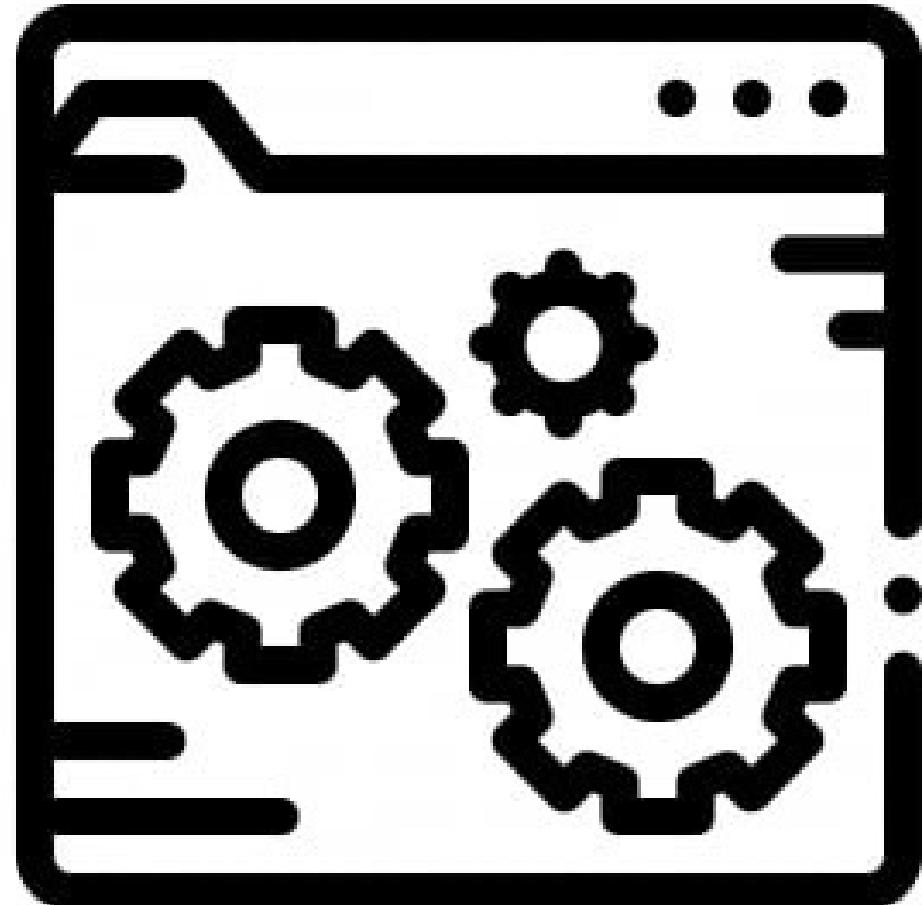
- This project aims to develop a method for detecting the news stories using machine learning. The main goal of this project is to
- identify and predict whether a given news article is a fake news or not. We gathered our data, preprocessed the text, and translated
- our articles into supervised model features.

Delimitations: Our project does not guarantee 100% accuracy.

Work Flow of Project



System Requirements



Hardware :

1. 4GB RAM
2. i3 Processor
3. 500mb free space

Software :

1. Anaconda
2. Python

Libraries Used

- Pandas Numpy
- Seaborn Matplotlib
- Sklearn Regular
- Expression
-
-



Techniques Used

Logistic Regression

Logistic regression is a statistical method used for binary classification problems, where the goal is to predict the probability of a binary outcome (whether the new is fake or not) based on one or more input variables (also known as independent or predictor variables).

It models the relationship between the input variables and the binary output variable using sigmoid function, which transforms the output of the model to a probability between 0 and 1.

Decision Tree Classifier

A decision tree classifier is a type of algorithm used for classification tasks that builds a tree-like model by recursively partitioning the data based on the input variables and their values to predict the class label of new data.

Techniques Used

Gradient Bound Classifier

A gradient bound classifier is a type of machine learning algorithm used for classification tasks. It updates a set of weights to minimize the difference between predicted and actual outcomes. The "gradient bound" term means it has a constraint that prevents the weights from becoming too large. This helps prevent numerical instability or overfitting.

Random Forest Classifier

A random forest classifier is a type of machine learning algorithm used for classification tasks. It combines many decision trees to make a prediction. Each tree looks at a random subset of the data and features, which helps to prevent overfitting. The final prediction is based on the most commonly predicted class by all the trees. Random forest classifiers are accurate and can handle large and complex datasets.

Results

The outcome of our project is as follows:

- **Logistic Regression**: Out of all the test cases we have tested on the model, it gave 100 percent accuracy on trained data and more than 85% accuracy on untrained data(predicted data).
- **Decision Tree Classifier**: Out of all the test cases we have tested on the model, it gave 100 percent accuracy on trained data and more than 60% accuracy on untrained data(predicted data). **Gradient Bound Classifier**: Out of all the test cases we have tested on the model, it gave 100 percent accuracy on trained data and more than 70% accuracy on untrained data(predicted data).
- **Random Forest Classifier**: Out of all the test cases we have tested on the model, it gave 100 percent accuracy on trained data and more than 90% accuracy on untrained data(predicted data).

Conclusion

The outcome of this project shows the capability of machine learning to be fruitful in this Task. We have tried to build a model that helps in catching many intuitive indications of real and fake news. We have used multiple performance metrics

References

1. <https://www.geeksforgeeks.org/fake-news-detection-using-machine-learning/>
2. <https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/>

THANK YOU