FAKE NEWS DETECTION SYSTEM

BSCS 6-A

SEMESTER PROJECT PROPOSAL

By

MANAHIL SHAMIM (02-134201-055)

IMAAN SHAHID (02-134201-009)

MUHAMMAD AWAIS ARSHAD(02-134201-038)

Proposed Project Supervisor(s):

Miss Salas Akbar

Sir Tariq Siddique

Department of Computer Science

Bahria University Karachi Campus

ABSTRACT

Recently, fake news has been incurring many problems to our society. As a result, many researchers have been working on identifying fake news. Most of the fake news detection systems utilize the linguistic feature of the news. However, they have difficulty in sensing highly ambiguous fake news which can be detected only after identifying meaning and latest related information. Fake news and hoaxes have been there since before the advent of the Internet. The widely accepted definition of Internet fake news is: fictitious articles deliberately fabricated to deceive readers". Social media and news outlets publish fake news to increase readership or as part of psychological warfare. In general, the goal is profiting through clickbaits. Clickbaits lure users and entice curiosity with flashy headlines or designs to click links to increase advertisements revenues. This exposition analyzes the prevalence of fake news in light of the advances in communication made possible by the emergence of social networking sites. The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information. We use simple and carefully selected features of the title and post to accurately identify fake posts.

The area of this project fake news detection system with face is text processing. The software requirements for this project is Jupyter Notebook. The algorithms we implemented in this project are naïve bayes, random forest, SVC and logistic regression and then later we will check whose accuracy is more perfect.

Content

CHAPTER 1

Introduction

1.1 Problem Area

To come up with a solution that can be utilized by users to detect and filter out sites containing false and misleading information and the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

1.2 Scope

We aim to provide the user with the ability to classify the *news* as *fake* or real and also check the authenticity of the website publishing the news.

1.3 Expected Time

The research project is expected to be completed in 3 weeks.

1.4 Specific Aims and Deliverables

Software Project Proposal.
Project progress
Project report
Team member's work as per their contribution.
To choose different algorithms and check the accuracy and efficiency of them.

1.5 Tools and Benefits

The tool that will be used is ANACONDA (Jupyter or Pycharm) etc.

1.6 Description

The aim of the system testing process was to determine all defects in our project. The program was subjected to a set of test inputs and various observations were made and based on these observations it will be decided whether the program behaves as expected or not.

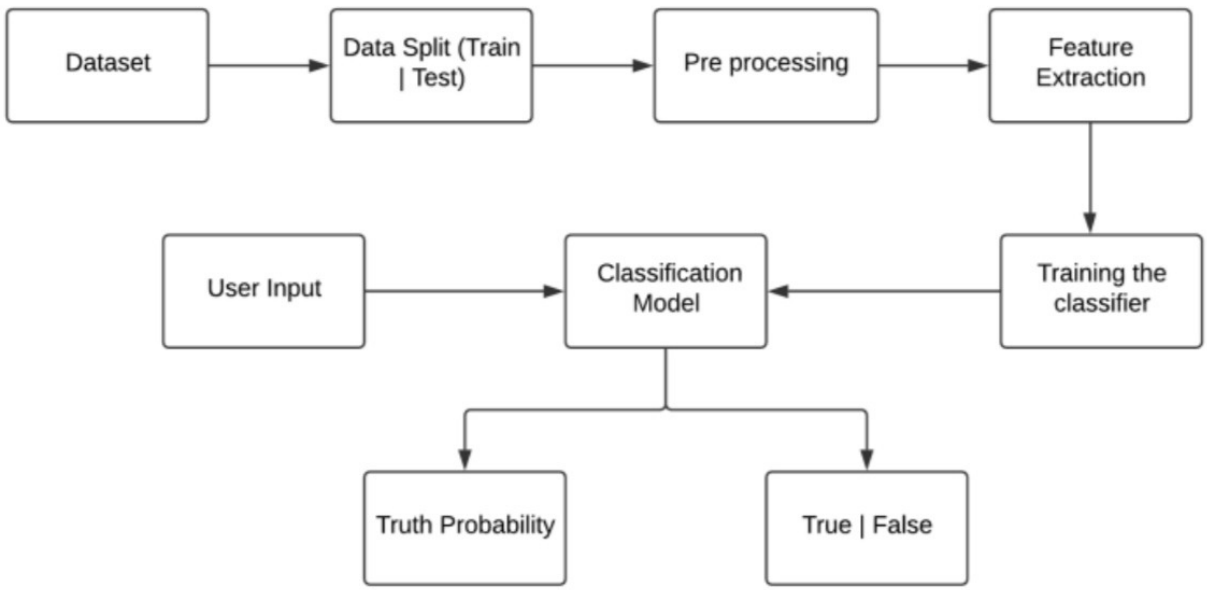


Figure 1.1: The three general stages in sentiment analysis

1.7 Motivation

User is able to detect fake and real news by using few clicks of mouse and few search keywords thus saving his valuable time. Social media and news outlets publish fake news to increase readership which creates troubles in order to solve this problem and to understand Artificial Intelligence from a deeper view we chose this project.

CHAPTER 2

Literature Review

2.1 Description of Related Work

Introduction

Fake news has become a problem of great impact in our information driven society because of the continuous and intense fake content distribution. Information quality in news feeds is under questionable veracity calling for automated tools to detect fake news articles. Due to many faces of fake, creating such tool is a challenging problem. In this work, we propose a model for fake news detection using content-based features and Machine Learning (ML) algorithms.

Methodology

Studying and training the different models with 4 different classifiers and choose the best classifier for final execution.

Implementation

We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the fact-checking websites. On the Internet, there are a few publicly available datasets for Fake news classification like Buzzfeed News, LIAR [15], BS Detector etc. These datasets have been widely used in different research papers for determining the veracity of news.

Then the dataset can be preprocessed, cleaned, tokenized, punctuation removal, vectorized etc.

Algorithms used

Different classifiers can be investigated to predict the class of the text. Four different machine learning algorithms have been used to predict which one has the best accuracy – Multinomial Naïve Bayes Passive Aggressive Classifier and Logistic regression.
The implementations of these classifiers are done using Python library Sci-Kit Learn.

Brief introduction to the algorithms

1. Naïve Bayes Classifier:

This classification technique is based on Bayes theorem, which assumes that the presence of a particular feature in a class is independent of the presence of any other feature. It provides way for calculating the posterior probability.

$$P(x) = \frac{P(c) * P(c)}{P(x)}$$

2. Random Forest

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

3. Logistic Regression

It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression.

4. SVC

SVC, or Support Vector Classifier, is a supervised machine learning algorithm typically used for classification tasks. SVC works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.

Evaluation Matrices

Evaluate the performance of algorithms for fake news detection problem; various evaluation metrics have. These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives. Specifically, accuracy measures the similarity between predicted fake news and real fake news.

References:

https://www.ijert.org/research/fake-news-detection-using-machine-learning-algorithms-IJERTCONV9IS03104.pdf

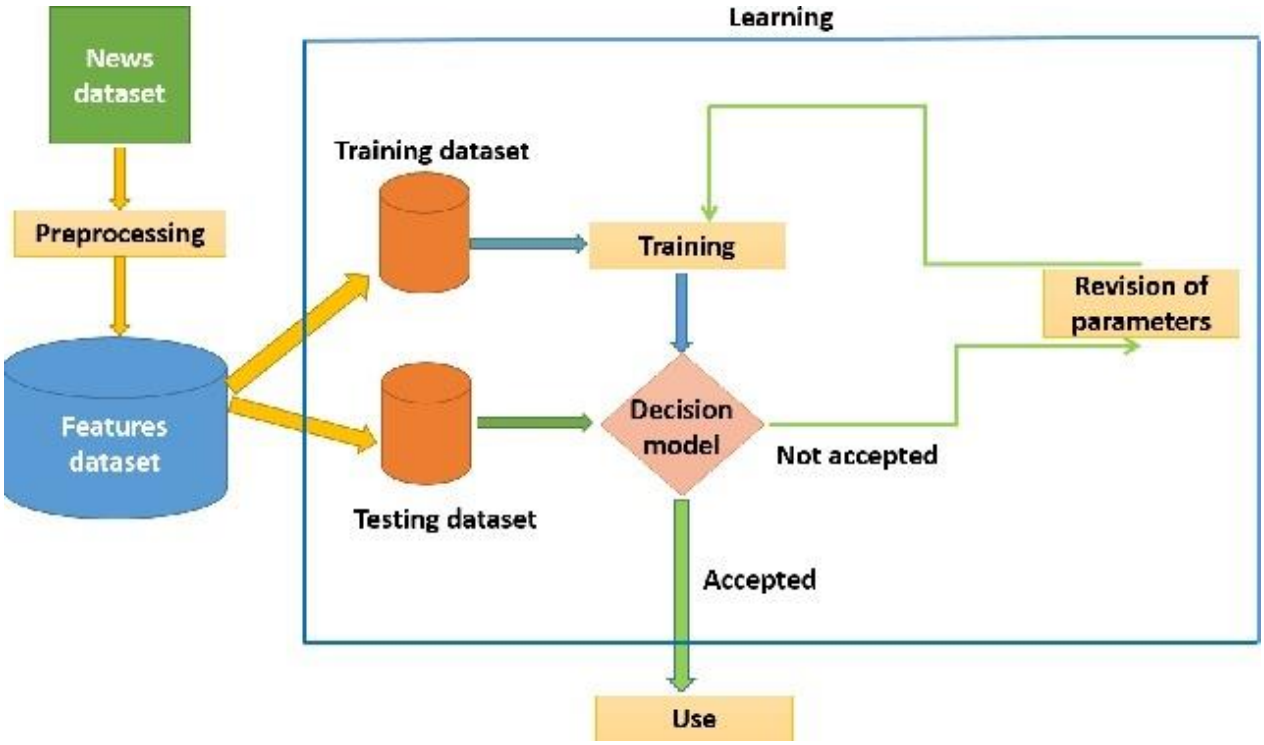
https://d1wqtxts1xzle7.cloudfront.net/66254531/fake_news_detection_using_machine_IJERTCONV9IS03104-libre.pdf?1618308329=&response-content-disposition=inline%3B+filename%3DIJERT_Fake_News_Detection_using_Machine.pdf&Expires=1671820258&Signature=KnO1XfaYvyT30KkMilPIJsTW0Ilnu6Yw6FmUjsY9aSB4CznzGRWeAA6kph4swmywtzV1bWc529RlKZif4hGSlu1dpF7mWyU9hWSurA-6T5FR~4v0VlXiikpRNT1bKAYsDfNpMT73ipgNYJAF5pqjWjyZ-p5iMVe4GgyJIAojXu0dL1LWp~ksBQmI9~hih8-KSg-fk7nfA33gcmlXEY3yIzyRasjfzKWZCHwCPR1gnQ4Nt~radfbNBlpZ9jT57~FPoEh9HxDaL2YX5ItDw-lDITQqL940dlGCFxKt~IIqWWDvfg2KJrc6uu6Hwp93iGCJNeb9wYBXVfDBmD3ydI64zg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

CHAPTER 3

Problem Definition

3.1 Title or problem selected

Recently, fake news has been incurring many problems to our society. The problem is to check the truthfulness of major claims in a news article to decide the news veracity. It is a classic text classification problem with a straight forward proposition implemented with multiple algorithms step wise to see which can work and provide better results in detecting the fake news. This advanced python project of detecting fake news deals with fake and real news and provides users a friendly environment to detect it.



Steps in the aspect-based sentiment analysis workflow

3.2 Activity Plan

Name	Start Date	End Date
Project planning	Dec 01, 2022	Dec 01, 2022
Data collection	Dec 02, 2022	Dec 02, 2022
Data preprocessing	Dec 06, 2022	Dec 06, 2022
Training data and feature extration	Dec 08, 2022	Dec 13, 2022
Testing model	Dec 14, 2022	Dec 19, 2022

References:

https://towardsdatascience.com/fake-news-detection-with-machine-learning-using-python-3347d9899ad1

https://scikit-learn.org/stable/