

A Project Report

CLASSIFICATION OF REAL AND FAKE JOB ADVERTISMENT USING MACHINE LEARNING

**Submitted in partial fulfilment of the requirement
for the award of the degree of**

B.tech/ CSE



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Mr. DAMODHARAN.D

Submitted By -
SHADAB ALI (19SCSE1010372)

SCHOOL OF COMPUTING SCIENCE & ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER, 2021



**SCHOOL OF COMPUTING SCIENCE AND
ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA**

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis/project/dissertation, entitled “**CLASSIFICATION OF REAL & FAKE JOB NOTIFICATION USING MACHINE LEARNING.**” in partial fulfilment of the requirements for the award of B. TECH submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of JULY, 2021 to DECEMBER ,2021 under the supervision of MR. DAMODHARAN. D PROFESSOR, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me for the award of any other degree of this or any other places.

SHADAB ALI (19SCSE1010372)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

DR DAMODHARAN.D
PROFESSOR

TABLE_OF CONTENTS

S. No.	Particulars	Page No.
1	Abstract	4
2	List of Tables	5
3	Introduction	6
4	Requirements	8
5	Feasibility Analysis	15
6	Objectives	16
7	Implementation	19
8	Design	29
9	Diagrams	30
10	Result, Conclusion & Future scope	43

Abstract

The process of searching jobs is one of the most problematic issue freshers face, this process is used by various scamsters to lure freshers into scam and profit from the students. – To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification technique is proposed.

While browsing for jobs online we saw that many scamsters demanded money for booking slots to interviews that did not exist and also extort money from students with promise of giving them jobs in return, this served as motivation for this proposal. The objectives that are to be considered are: Prediction of real or fake job. And a front-end page to allow non-technical user to see the model.

Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts. Two major types of classifiers, such as single classifier and ensemble classifiers are considered for fraudulent job posts detection. However, experimental results indicate that ensemble classifiers are the best classification to detect scams over the single classifiers.

The model is trend to be as efficient as possible by making the dataset to be a part of double-blind study and also considering the various formats of posting jobs in professional websites and other site too.

Introduction

The main aim of this project is to reduce the fake jobs which is introduced by the different types of applications like LinkedIn, Facebook, Instagram etc....in recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs.

For this purpose, we are using machine learning approach which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. The outcomes of those tests are identified by decision node. Starting from the beginning at the root this tree is going through it until a leaf node is reached.

Decision tree learning is an approach that has been applied to spam filtering. This can be useful for forecasting the goal based on some criterion by implementing and training this model.

Ensemble Approach based Classifiers

Ensemble approach facilitates several machine learning algorithms to perform together to obtain higher accuracy of the entire system. Random forest exploits the concept of ensemble learning approach and regression technique applicable for classification-based problems.

This classifier assimilates several tree-like classifiers which are applied on various sub-samples of the dataset and each tree casts its vote to the most appropriate class for the input.

Boosting is an efficient technique where several unstable learners are assimilated into a single learner in order to improve accuracy of classification. Boosting technique applies classification algorithm to the reweighted versions of the training data and chooses the weighted majority vote of the sequence of classifiers.

AdaBoost is a good example of boosting technique that produces improved output even when the performance of the weak learners is inadequate. Boosting algorithms are quite efficient in solving spam filtration problems. Gradient boosting] algorithm is another boosting technique-based classifier that exploits the concept of decision tree. It also minimizes the prediction loss.

REQUIREMENTS

Coding platform: Jupyter Notebook, Streamlite

Language Used: Python

Concept Used: Machine Learning

FEATURES OF THE PROJECT

- It is Feasible.
- Free of Cost
- Python-based Project
- Easy to use
- Works Smoothly.

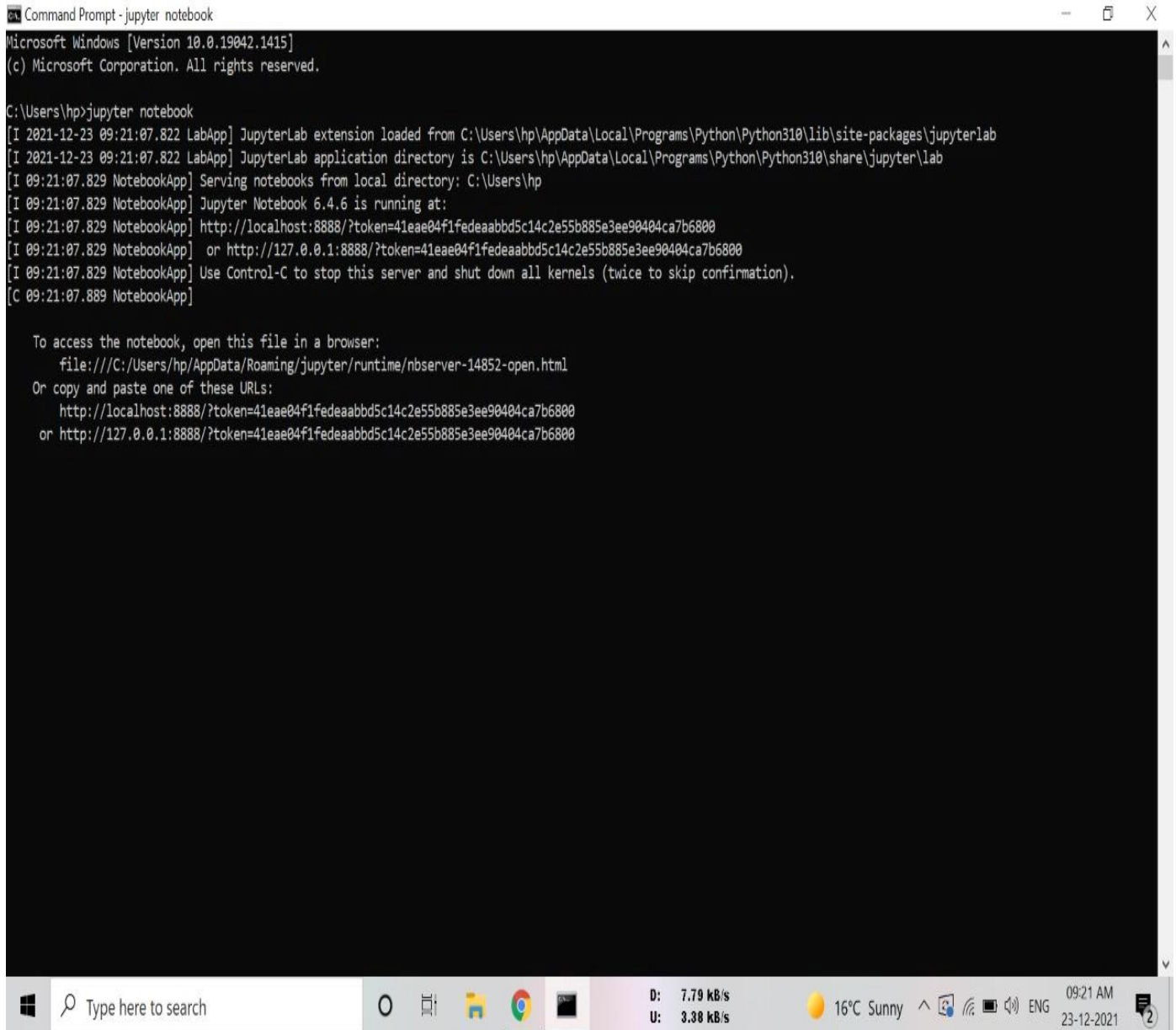
ADVANTAGES

- It provides better results in classification problems.
- It is strong in preprocessing outliers, irrelevant variables, a mix of continuous.
- It is categorical and discrete variables.
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

DISADVANTAGES

- The classification of the data from the dataset is very difficult.
- The accuracy of the job prediction is less.
- The data classification from the actual dataset and given dataset has to be performed manually.

IMPLEMENTATION



```
Command Prompt - jupyter notebook
Microsoft Windows [Version 10.0.19042.1415]
(c) Microsoft Corporation. All rights reserved.

C:\Users\hp>jupyter notebook
[I 2021-12-23 09:21:07.822 LabApp] JupyterLab extension loaded from C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\jupyterlab
[I 2021-12-23 09:21:07.822 LabApp] JupyterLab application directory is C:\Users\hp\AppData\Local\Programs\Python\Python310\share\jupyter\lab
[I 09:21:07.829 NotebookApp] Serving notebooks from local directory: C:\Users\hp
[I 09:21:07.829 NotebookApp] Jupyter Notebook 6.4.6 is running at:
[I 09:21:07.829 NotebookApp] http://localhost:8888/?token=41eae04f1fedeaaabbd5c14c2e55b885e3ee90404ca7b6800
[I 09:21:07.829 NotebookApp] or http://127.0.0.1:8888/?token=41eae04f1fedeaaabbd5c14c2e55b885e3ee90404ca7b6800
[I 09:21:07.829 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 09:21:07.889 NotebookApp]

To access the notebook, open this file in a browser:
    file:///C:/Users/hp/AppData/Roaming/jupyter/runtime/nbserver-14852-open.html
Or copy and paste one of these URLs:
    http://localhost:8888/?token=41eae04f1fedeaaabbd5c14c2e55b885e3ee90404ca7b6800
    or http://127.0.0.1:8888/?token=41eae04f1fedeaaabbd5c14c2e55b885e3ee90404ca7b6800
```

FIG – 01

- ❖ Figure 01 shows the command prompt screen.
- ❖ Here we have to go to the folder where our folder of codes are saved.

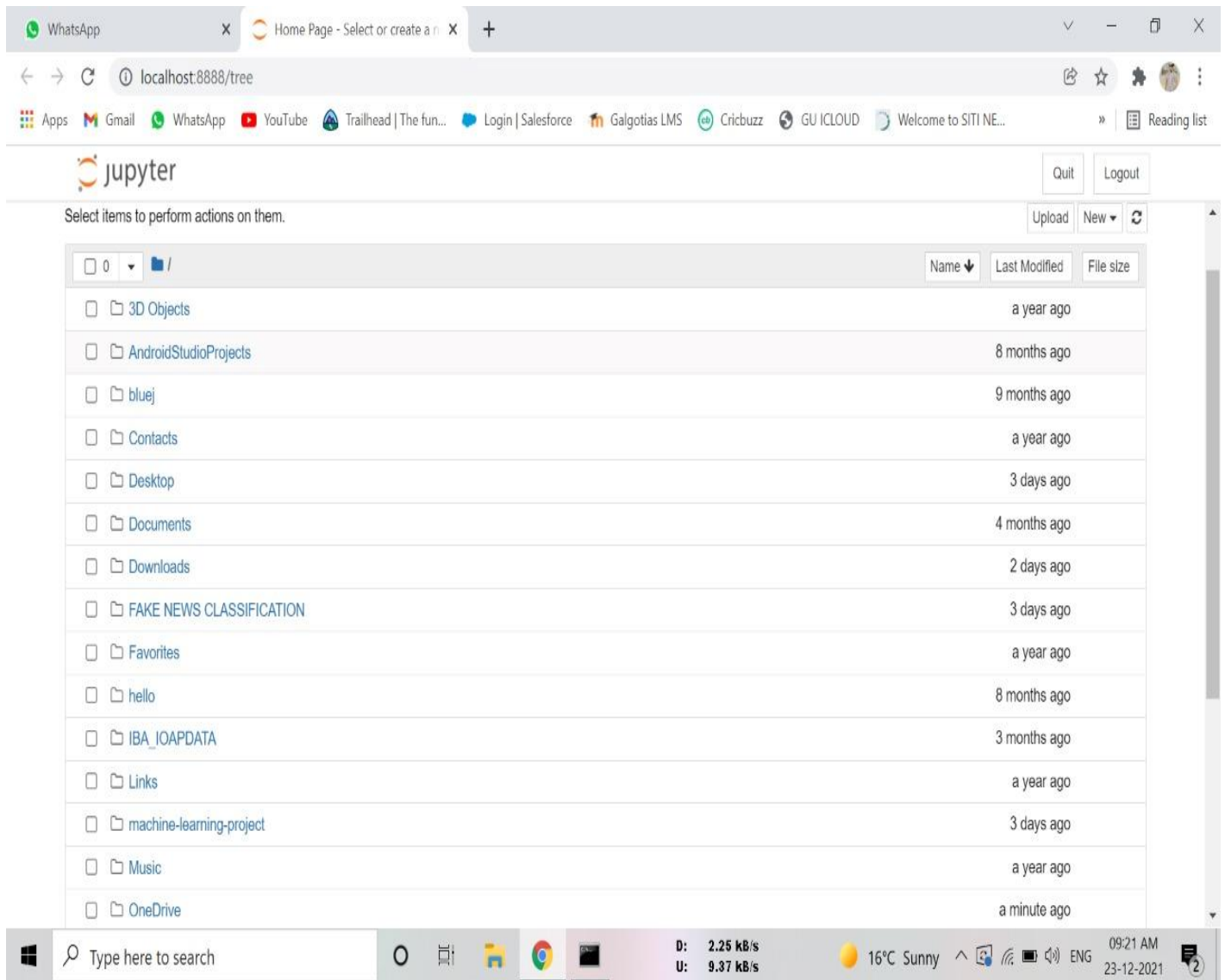


FIG – 02

- ❖ After opening command prompt, it will redirect to the local host window shown in figure 02.
- ❖ Here we can see our Jupyter window, where every file is saved.

The screenshot shows a Jupyter Notebook window titled 'Untitled' with the following code and output:

```
In [2]: import pandas as pd
import numpy as np

In [3]: data=pd.read_csv("news.csv", encoding="latin-1")

In [4]: data.sample()

Out[4]:
```

Unnamed: 0		title	text	label
2917	8276	Iraq and US confident to retake Mosul, Daesh c...	39 UTC Â© Hikmet Durgun The Daesh terrorists'...	FAKE

```
In [5]: data.head(3)

Out[5]:
```

Unnamed: 0		title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL

```
In [6]: data.columns

Out[6]: Index(['Unnamed: 0', 'title', 'text', 'label'], dtype='object')
```

The interface also shows a Windows taskbar at the bottom with the search bar, taskbar icons, and system tray information including network speed (0.31 kB/s), temperature (16°C Sunny), and time (09:22 AM 23-12-2021).

FIG – 03

- ❖ As shown in figure 03 we will start coding on jupyter window.
- ❖ Here we have to import pandas and numpy.
- ❖ Then open our csv file in read mode.
- ❖ It will show us the sample data.

The screenshot shows a Jupyter Notebook with the following content:

```
In [8]: data.head()
```

	title	text	label
0	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow...	FAKE
1	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumbleu...	FAKE
2	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	Bernie supporters on Twitter erupt in anger ag...	â Kaydee King (@KaydeeKing) November 9, 2016...	FAKE
4	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

```
In [9]: data["label"] = data["label"].map({'FAKE':0, 'REAL':1})
```

```
In [10]: data.head()
```

	title	text	label
0	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow...	0
1	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg LinkedIn Reddit Stumbleu...	0
2	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	1
3	Bernie supporters on Twitter erupt in anger ag...	â Kaydee King (@KaydeeKing) November 9, 2016...	0

FIG – 04

❖ We can fetch the desire data by slicing it as shown in figure 04.

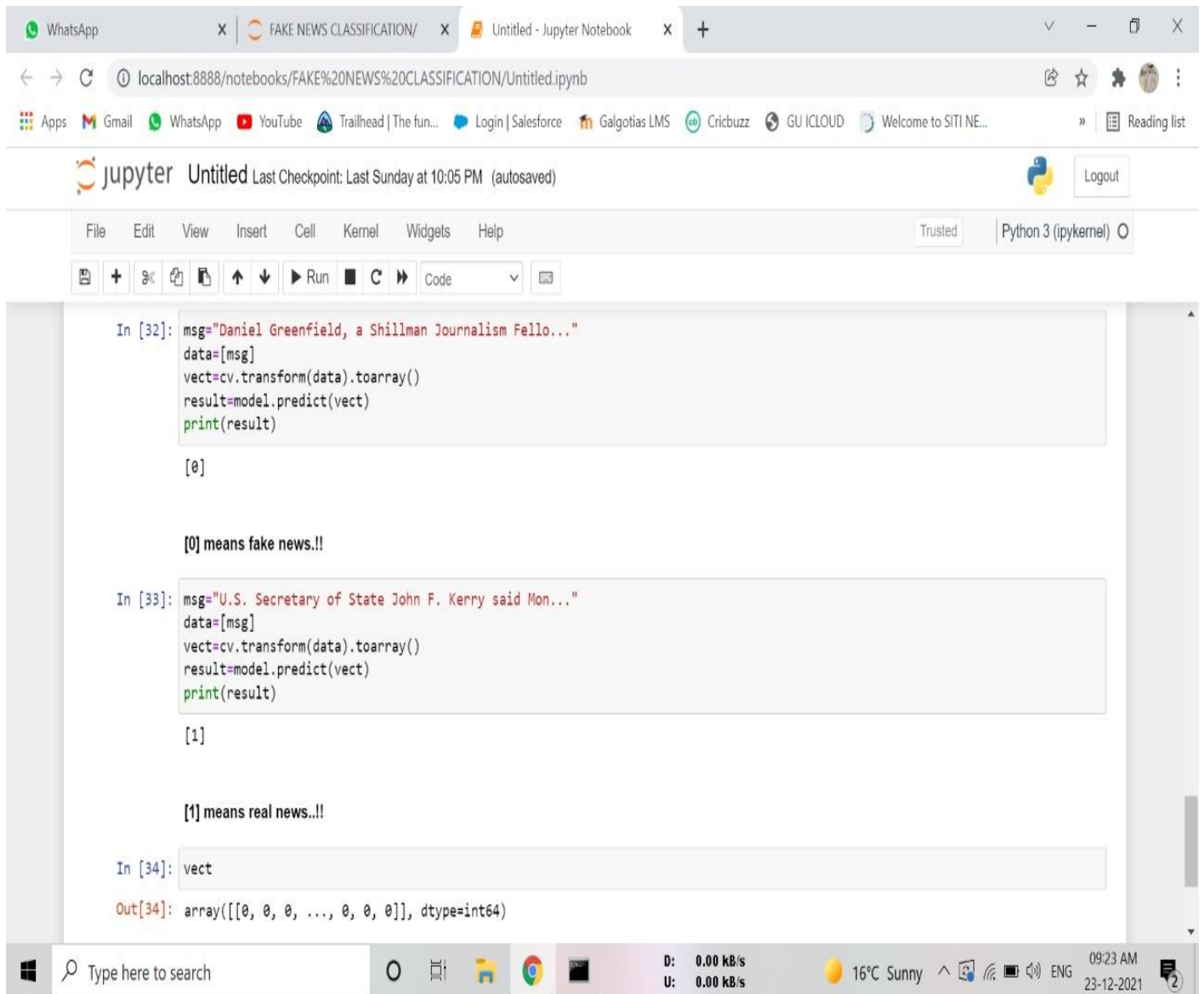


FIG – 05

- ❖ Figure 05 shows the model prediction.
- ❖ We will type our message.
- ❖ And saved is as array then, predict the model and saved it in result.
- ❖ It print the predicted results on the screen.

```
C:\Users\hp\FAKE NEWS CLASSIFICATION\ShadabfakeNews.py (hp) - Sublime Text (UNREGISTERED)
File Edit Selection Find View Goto Tools Project Preferences Help

FOLDERS
hp
  .android
  .gradle
  .idlerc
  .ipynb_checkpoints
  .ipython
  .jupyter
  .streamlit
  .vscode
  3D Objects
  AndroidStudio
  AppData
  Application Data
  bluej
  Contacts
  Cookies
  Desktop
  Documents
  Downloads
  FAKE NEWS CLASSIFICATION
    .ipynb_checkpoints
    news.csv
    news.pkl
  /* Shadabfake
  /* Untitled.ipynb
  vectorizer.py
  Favorites
  hello
  IDEA_10ADDDAT

ShadabfakeNews.py x
1 import pickle
2 import streamlit as st
3
4
5 model=pickle.load(open("news.pkl","rb"))
6 cv=pickle.load(open("vectorizer.pkl","rb"))
7
8
9 def main():
10     st.title("FAKE News Classifier Website")
11     st.subheader(":Made By Shadab With Python & Streamlit")
12     msg=st.text_input("Enter the Text : ")
13     if st.button("Predict"):
14         data=[msg]
15         vect=cv.transform(data).toarray()
16         prediction=model.predict(vect)
17         result=prediction[0]
18         if result==0:
19             st.error("This is Fake News ")
20         else:
21             st.success("This a Real News ")
22
23     main()

Line 1, Column 1
Tab Size: 4 Python
Type here to search
D: 0.15 kB/s
U: 0.00 kB/s
16°C Sunny
09:24 AM
23-12-2021
```

FIG – 06

- ❖ The above figure 06 show the backend program of the algorithm and machine learning.
- ❖ We have to run the above program.

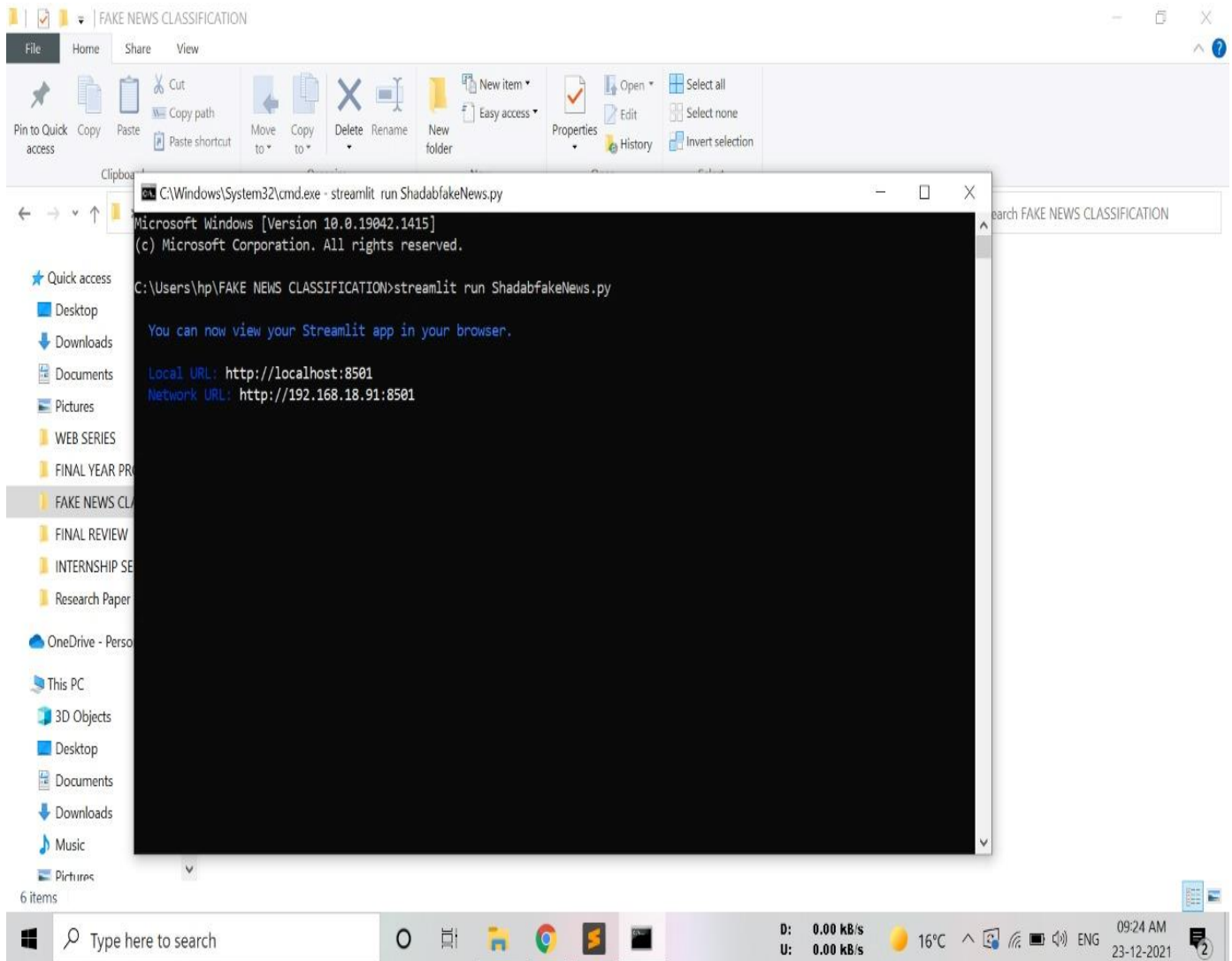


FIG – 07

- ❖ After running the streamlite program it will redirect to the command prompt as shown in figure 07.
- ❖ Here we can see the name of local host url and network url.

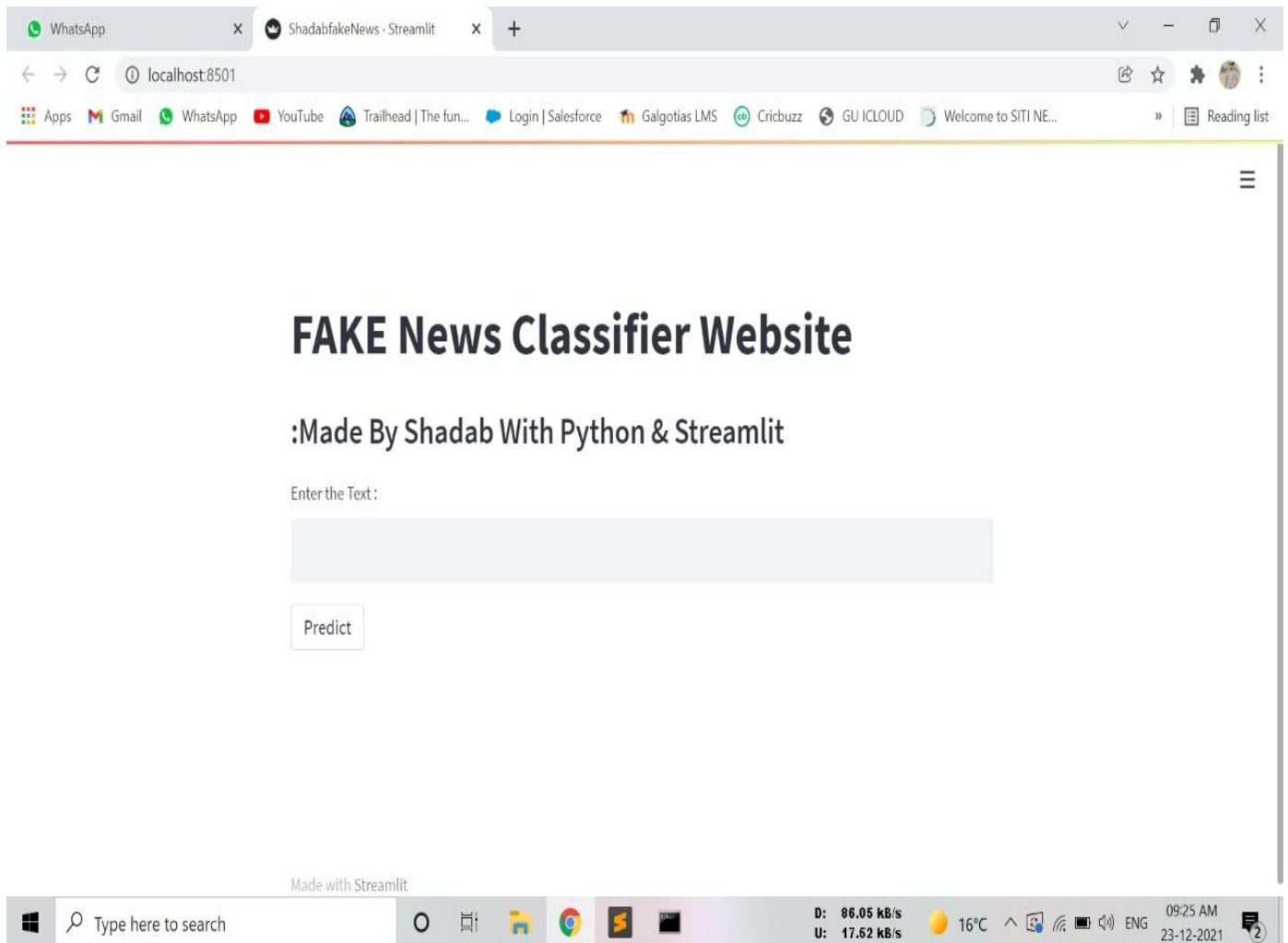


FIG – 08

- ❖ On hitting the url in previous image it will redirect to our website as shown in figure 08.
- ❖ Here we can paste our message.
- ❖ And on clicking on the predict button it will predict the result.

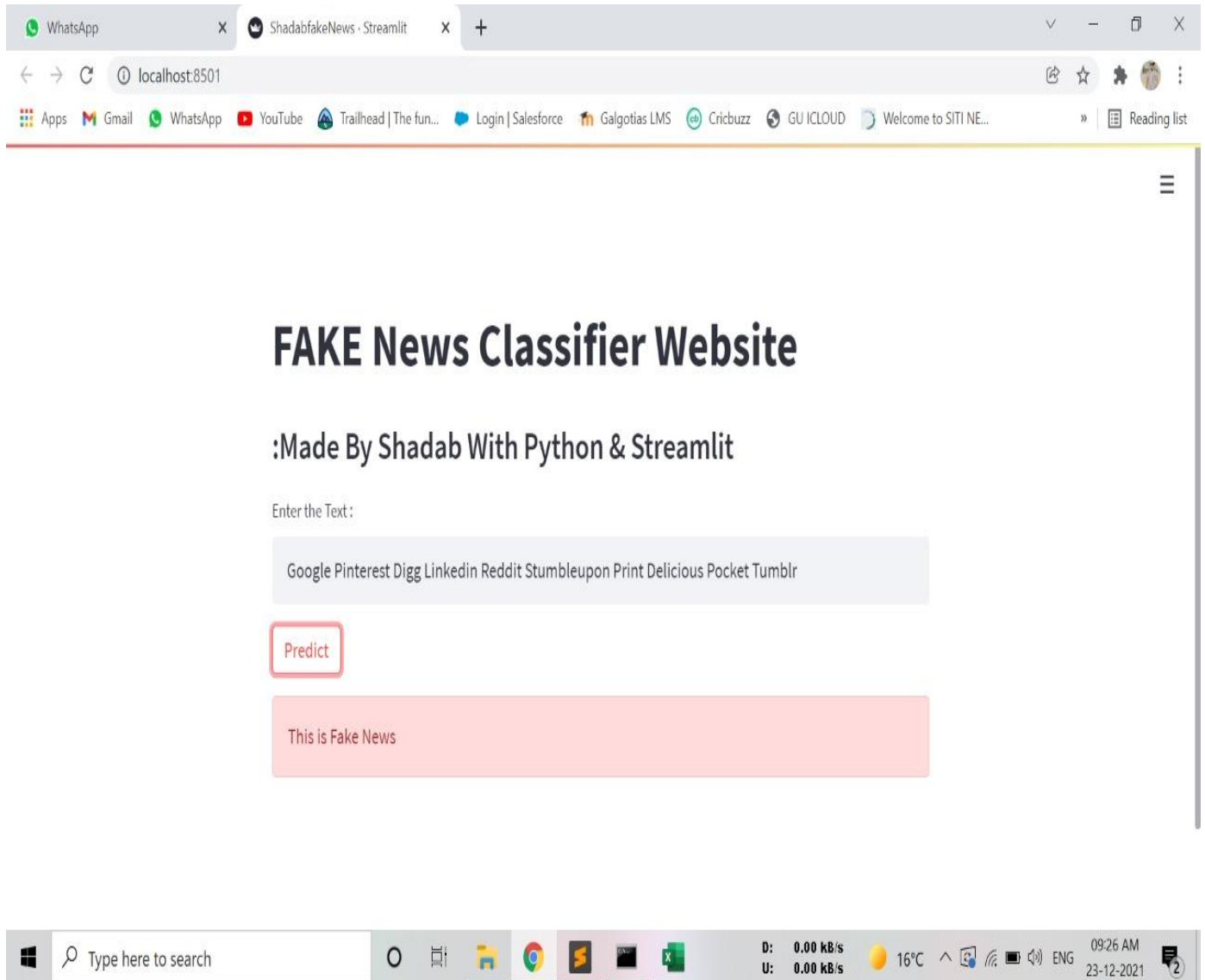


FIG – 09

❖ Figure 09 shows the message provide by user is fake.

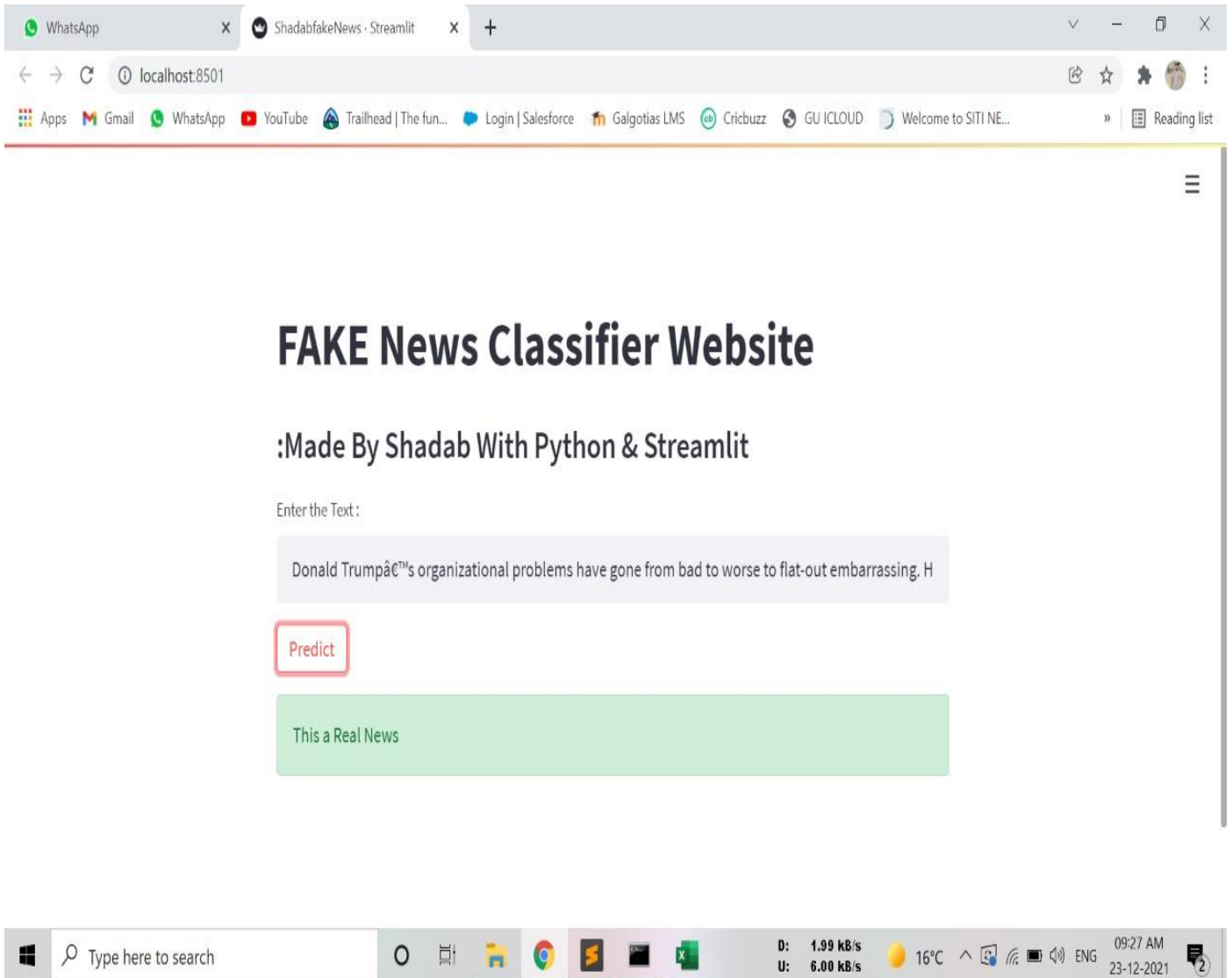


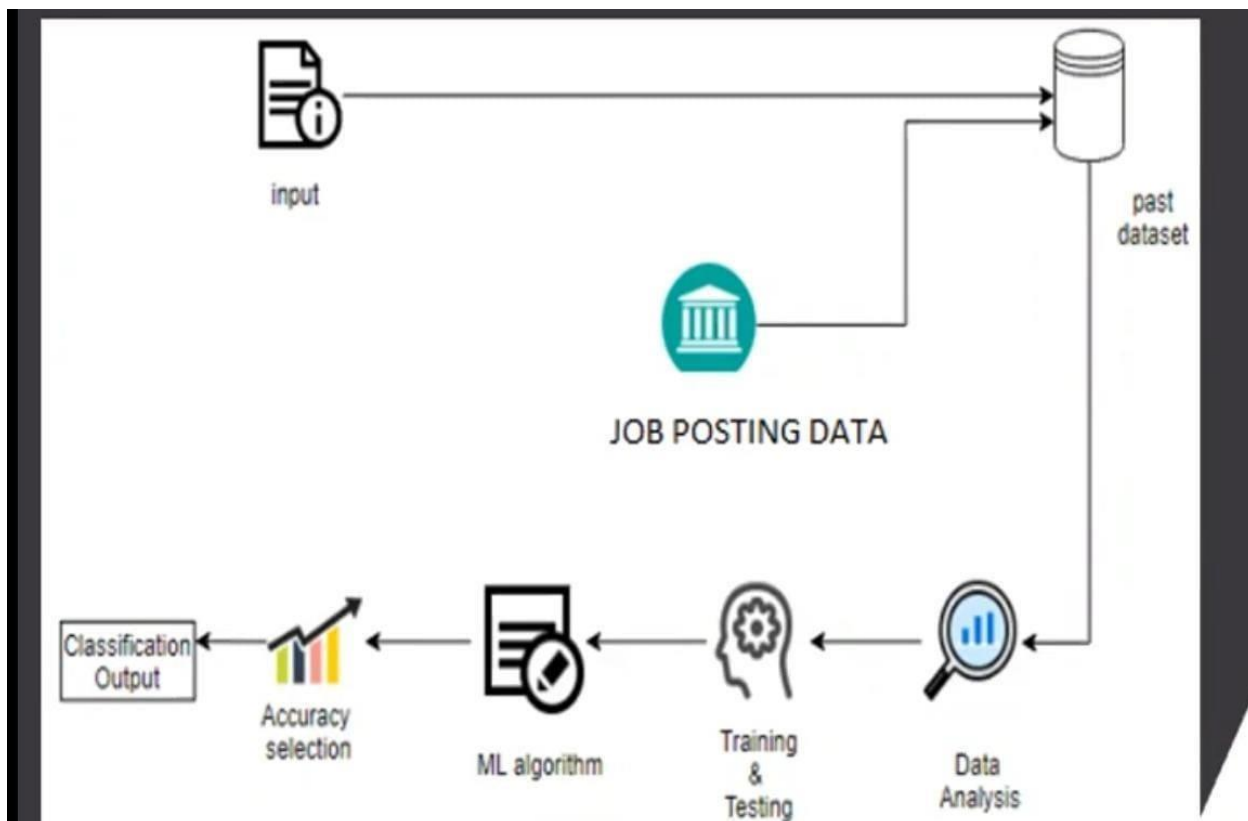
FIG – 10

- ❖ Figure 10 shows the provided data is real.
- ❖ That’s how the whole program works.

Sample data

	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

FLOW DIAGRAM



RELATED WORK

Some spams in recent times like Review spam detection, Email Spam detection, Fake news detection have drawn special attention in the domain of Online Fraud Detection in different types of social apps like LinkedIn, Instagram, Facebook etc....

(1) Review Spam Detection

People often post their reviews online forum regarding the products they purchase. It may guide other purchaser while choosing their products. In this context, spammers can manipulate reviews for gaining profit and hence it is required to develop techniques that detects these spam reviews.

This can be implemented by extracting features from the reviews by extracting features using Natural Language Processing (NLP). machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that uses dictionary or corpus to eliminate spam reviews.

(2) Email Spam Detection

Unwanted bulk mails, belong to the category of spam emails, often arrive to user mailbox. This may lead to unavoidable storage crisis as well as bandwidth consumption. To eradicate this problem, Gmail, Yahoo mail and Outlook service providers incorporate spam filters using Neural Networks.

(3) Fake News Detection

Fake news in social media characterizes malicious user accounts, echo chamber effects. The fundamental study of fake news detection relies on three perspectives- how fake news is written, how fake news spreads, how a user is related to fake news. Features related to news content and social context are extracted and a machine learning model are imposed to recognize fake news

PERFORMANCE COMPARISON
CHART FOR SINGLE CLASSIFIER
BASED PREDICTION

Performance Measure Metric	Naïve Bayes Classifier	Multi-Layer Perceptron Classifier	K-Nearest Neighbor Classifier	Decision Tree Classifier
Accuracy	72.06%	96.14%	95.95%	97.2%
F1-Score	0.72	0.96	0.96	0.97
Cohen-Kappa Score	0.12	0.3	0.38	0.67
MSE	0.52	0.05	0.04	0.03

DESCRIPTION OF REQUIREMENTS

MACHINE LEARNING

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

But, using the classic algorithms of machine learning, text is considered as a sequence of keywords; instead, an approach based on semantic analysis mimics the human ability to understand the meaning of a text.

Some Machine Learning Methods

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn't figure out

the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabelled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviour within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

Streamlit

Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

Installing Streamlit

1. Make sure you have python installed in your system.
2. Use the following command to install streamlit.

Development flow

If the source code of the streamlit's python script changes the app shows whether to rerun the application or not in the top-right corner. You can also select the 'Always rerun' option to rerun always when the source script changes.

This makes our development flow much easier, every time you make some changes it'll reflect immediately in your web app. This loop between coding and viewing results live makes you work seamlessly with streamlit.

Data flow

Streamlit allows you to write an app the same way you write a python code. The streamlit has a distinctive data flow, any time something changes in your code or anything needs to be updated on the screen, streamlit reruns your python script entirely from the top to the bottom. This happens when the user interacts with the widgets like a select box or drop-down box or when the source code is changed.

If you have some costly operations while rerunning your web app, like loading data from databases, you can use streamlit's `st.cache` method to cache those datasets, so that it loads faster.

Displaying the data

Streamlit provides you with many methods to display various types of data like arrays, tables, and data frames.

To write a string simply use, `st.write("Your string")`

To display a data frame use, `st.dataframe` method

Widgets

There are several widgets available in streamlit, like `st.selectbox`, `st.checkbox`, `st.slider`, and etc.

Layout

You can easily arrange your widgets or data seamlessly using the `st.sidebar` method. This method helps you to align data in the left panel sidebar. All you have to do is simply use `st.sidebar.selectbox` to display a selectbox in the left panel.

Working of streamlit

- Streamlit runs the python script from top to bottom
- Each time the user interacts the script is a rerun from top to bottom
- Streamlit allows you to use caching for costly operations like loading large datasets.

PyCharm Community

PyCharm is a hybrid-platform developed by JetBrains as an IDE for Python. It is commonly used for Python application development. Some of the unicorn organizations such as Twitter, Facebook, Amazon, and Pinterest use PyCharm as their Python IDE!

It supports two versions: v2.x and v3.x.

We can run PyCharm on Windows, Linux, or Mac OS. Additionally, it contains modules and packages that help programmers develop software using Python in less time and with minimal effort. Further, it can also be customized according to the requirements of developers.

Features of PyCharm:

1. Intelligent Code Editor:

- ❖ It helps us write high-quality codes!
- ❖ It consists of colour schemes for keywords, classes, and functions. This helps increase the readability and understanding of the code.
- ❖ It helps identify errors easily.
- ❖ It provides the autocomplete feature and instructions for the completion of the code.

2. Code Navigation:

- ❖ It helps developers in editing and enhancing the code with less effort and time.
- ❖ With code navigation, a developer can easily navigate to a function, class, or file.
- ❖ A programmer can locate an element, a symbol, or a variable in the source code within no time.
- ❖ Using the lens mode, further, a developer can thoroughly inspect and debug the entire source code.

3. Refactoring:

- ❖ It has the advantage of making efficient and quick changes to both local and global variables.
- ❖ Refactoring in PyCharm enables developers to improve the internal structure without changing the external performance of the code.
- ❖ It also helps split up more extended classes and functions with the help of the extract method.

4. Assistance for Many Other Web Technologies:

- ❖ It helps developers create web applications in Python.
- ❖ It supports popular web technologies such as HTML, CSS, and JavaScript.
- ❖ Developers have the choice of live editing with this IDE. At the same time, they can preview the created/updated web page.
- ❖ The developers can follow the changes directly on a web browser.
- ❖ PyCharm also supports AngularJS and NodeJS for developing web applications.

5. Support for Popular Python Web Frameworks:

- ❖ PyCharm supports web frameworks such as Django.
- ❖ It provides the autocomplete feature and suggestions for the parameters of Django.
- ❖ It helps in debugging the codes of Django.
- ❖ It also assists web2py and Pyramid, the other popular web frameworks.

6. Assistance for Python Scientific Libraries:

- ❖ PyCharm supports Python's scientific libraries such as Matplotlib, NumPy, and Anaconda.
- ❖ These scientific libraries help in building projects of Data Science and

- ❖ Machine Learning.
- ❖ It consists of interactive graphs that help developers understand data.
- ❖ It is capable of integrating with various tools such as IPython, Django, and Pytest. This integration helps innovate unique solutions.

Python

Nowadays, Python is in great demand. It is widely used in the software development industry. There are 'n' number of reasons for this.

High-level object-oriented programming language: Python includes effective symbolism.

Rapid application development: Because of its concise code and literal syntax, the development of applications gets accelerated. The reason for its wide usability is its simple and easy-to-master syntax. The simplicity of the code helps reduce the time and cost of development.

Dynamic typescript: Python has high-level incorporated data structures blended with dynamic typescript and powerful binding.

Features of Python:

- ❖ Python supports code reusability and modularity.
- ❖ It has a quick edit-inspect-debug cycle.
- ❖ Debugging is straightforward in Python programs.
- ❖ It has its own debugger written in Python itself, declaring to Python's reflective power.

- ❖ Python includes a plethora of third-party components present in the Python Package Index (PyPI).

Heroku

Heroku is a cloud service platform whose popularity has grown in recent years. Heroku is so easy to use that it's a top choice for many development projects.

With a special focus on supporting customer-focused apps, it enables simple application development and deployment. Since the Heroku platform manages

hardware and servers, businesses that use Heroku are able to focus on perfecting their apps. And not the infrastructure that supports them.

Heroku, a Platform-as-a-Service solution, is generally easy-to-use. But it's most beneficial to businesses in specific situations. Heroku has a free service model for small projects. Also, tiered service packages exist for cases where more complex business needs must be addressed.

The Heroku cloud service platform is based on a managed container (called dynos within the Heroku paradigm) system. It has integrated data services and a powerful ecosystem for deploying and running modern applications.

Features of Heroku:

1. Heroku Accommodates Many Development Languages:

Heroku supports several programming languages that are used as a web application deployment model. As one of the first cloud platforms, Heroku has been in development since June 2007. Back then, it supported only the Ruby programming language.

But now it also supports Java, Node.js, Scala, Clojure, Python, PHP, and Go. This

means a variety of developers can look to Heroku for an inexpensive way to scale their application, no matter their preferred development language.

2. Heroku Supports Diverse Solutions:

Heroku also provides custom buildpacks, where developers can deploy apps in any other programming language. For this reason, Heroku is a polyglot platform. It lets the developer build, run, and scale applications in a similar manner across all programming languages.

Polymorphism and scalability are reasons why Heroku is often seen as a preferred platform amongst developers.

3. Heroku Dynos Enable Easy Development and Better Usability:

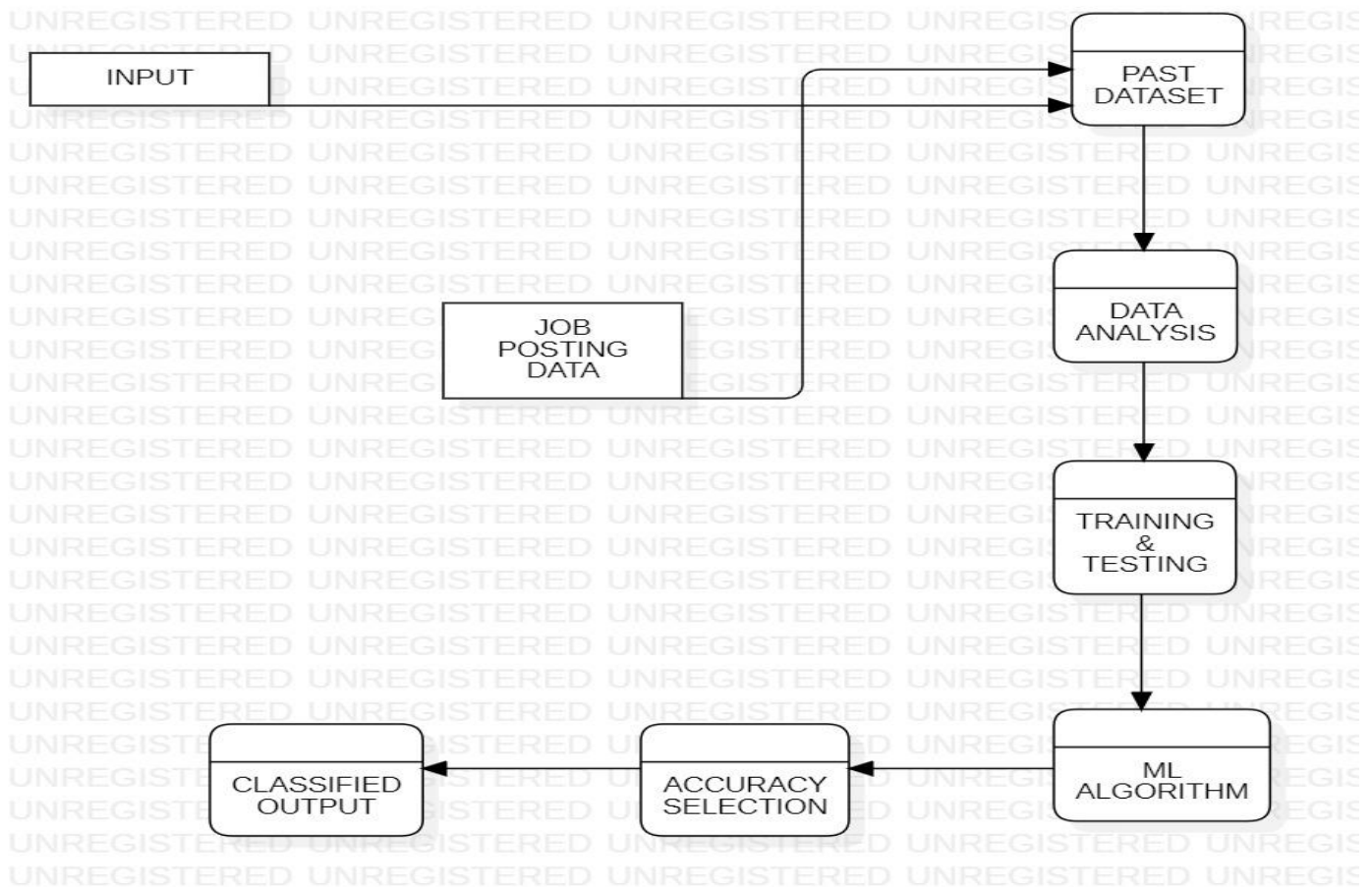
Applications that are run on Heroku typically have unique domain names, which are used to route HTTP requests to the correct container. Applications as services use application containers. Containers are designed to package and run services. Each of the application containers is a smart container on a reliable, fully-managed runtime environment.

4. Heroku Lets Developers Scale Applications Instantly:

This is accomplished either by increasing the number of dynos or by changing the type of dyno in which the app runs. When the application can scale so easily, the user can always expect more speed when using that application.

Data Flow Diagram

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyse an existing system or model a new one. Like all the best diagrams and charts, a DFD can often visually “say” things that would be hard to explain in words, and they work for both technical and nontechnical audiences, from developer to CEO. That’s why DFDs remain so popular after all these years. While they work well for data flow software and systems, they are less applicable nowadays to visualizing interactive, real-time or database-oriented software or systems.



Rules for creating DFD

- ❖ The name of the entity should be easy and understandable without any extra assistance (like comments).
- ❖ The processes should be numbered or put in ordered list to be referred easily.
- ❖ The DFD should maintain consistency across all the DFD levels.
- ❖ A single DFD can have maximum processes up to 9 and minimum 3 processes.

Advantages of DFD

- ❖ It helps us to understand the functioning and the limits of a system.
- ❖ It is a graphical representation which is very easy to understand as it helps visualize contents.
- ❖ Data Flow Diagram represent detailed and well explained diagram of system components.
- ❖ It is used as the part of system documentation file.
- ❖ Data Flow Diagrams can be understood by both technical or nontechnical person because they are very easy to understand.

Disadvantages of DFD

- ❖ At times DFD can confuse the programmers regarding the system.
- ❖ Data Flow Diagram takes long time to be generated, and many times due to this reasons analysts are denied permission to work on it.

Activity Diagram:

Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system.

The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all type of flow control by using different elements such as fork, join, etc

Purpose of Activity Diagrams

The basic purposes of activity diagrams is similar to other four diagrams. It captures the dynamic behaviour of the system. Other four diagrams are used to show the message flow from one object to another but activity diagram is used to show message flow from one activity to another.

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system, but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part.

It does not show any message flow from one activity to another. Activity diagram is sometimes considered as the flowchart. Although the diagrams look like a flowchart, they are not. It shows different flows such as parallel, branched, concurrent, and single.

The purpose of an activity diagram can be described as –

- ❖ Draw the activity flow of a system.
- ❖ Describe the sequence from one activity to another.
- ❖ Describe the parallel, branched and concurrent flow of the system.

How to Draw an Activity Diagram?

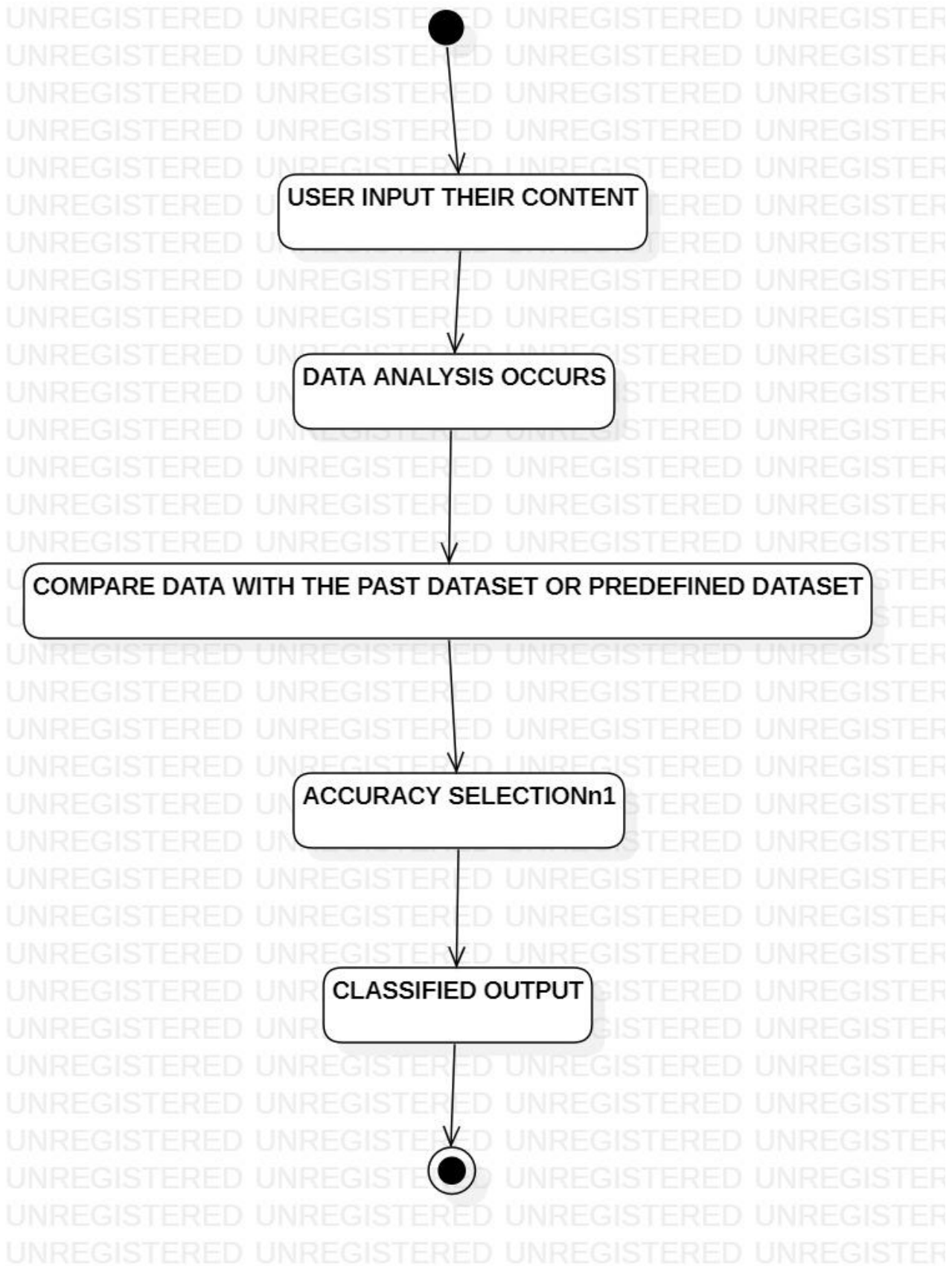
Activity diagrams are mainly used as a flowchart that consists of activities performed by the system. Activity diagrams are not exactly flowcharts as they have some additional capabilities. These additional capabilities include branching, parallel flow, swimlane, etc.

Before drawing an activity diagram, we must have a clear understanding about the elements used in activity diagram. The main element of an activity diagram is the activity itself. An activity is a function performed by the system. After identifying the activities, we need to understand how they are associated with constraints and conditions.

Before drawing an activity diagram, we should identify the following elements –

- ❖ Activities
- ❖ Association
- ❖ Conditions
- ❖ Constraints

Once the above-mentioned parameters are identified, we need to make a mental layout of the entire flow. This mental layout is then transformed into an activity diagram.



Where to Use Activity Diagrams?

The basic usage of activity diagram is similar to other four UML diagrams. The specific usage is to model the control flow from one activity to another. This control flow does not include messages.

Activity diagram is suitable for modelling the activity flow of the system. An application can have multiple systems. Activity diagram also captures these systems and describes the flow from one system to another. This specific usage is not available in other diagrams. These systems can be database, external queues, or any other system.

Activity diagram can be used for –

- ❖ Modelling work flow by using activities.
- ❖ Modelling business requirements.
- ❖ High level understanding of the system's functionalities.
- ❖ Investigating business requirements at a later stage.

Class Diagram:

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application.

Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modelling of object-oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages.

Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

Purpose of Class Diagrams

The purpose of class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction.

UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application, however class diagram is a bit different. It is the most popular UML diagram in the coder community.

The purpose of the class diagram can be summarized as –

- ❖ Analysis and design of the static view of an application.
- ❖ Describe responsibilities of a system.
- ❖ Base for component and deployment diagrams.

- ❖ Forward and reverse engineering.

How to Draw a Class Diagram?

Class diagrams are the most popular UML diagrams used for construction of software applications. It is very important to learn the drawing procedure of class diagram.

Class diagrams have a lot of properties to consider while drawing but here the diagram will be considered from a top-level view.

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. A collection of class diagrams represents the whole system.

The following points should be remembered while drawing a class diagram –

- ❖ The name of the class diagram should be meaningful to describe the aspect of the system.
- ❖ Each element and their relationships should be identified in advance.
- ❖ Responsibility (attributes and methods) of each class should be clearly identified
- ❖ For each class, minimum number of properties should be specified, as unnecessary properties will make the diagram complicated.
- ❖ Use notes whenever required to describe some aspect of the diagram. At the end of the drawing, it should be understandable to the developer/coder.
- ❖ Finally, before making the final version, the diagram should be drawn on plain paper and reworked as many times as possible to make it correct.

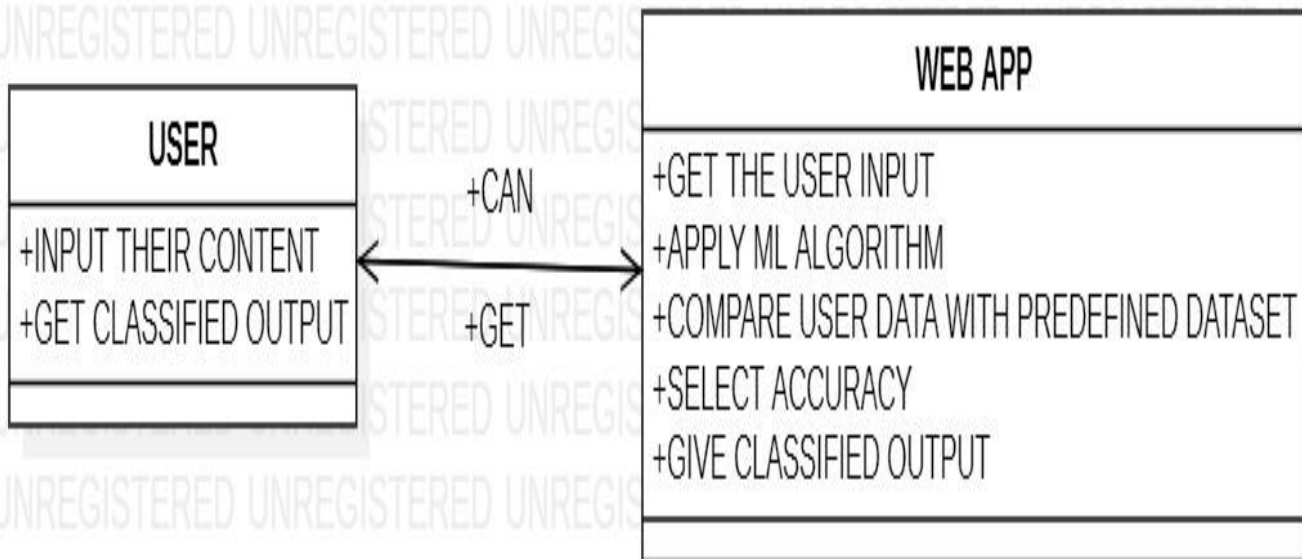
Where to Use Class Diagrams?

Class diagram is a static diagram and it is used to model the static view of a system. The static view describes the vocabulary of the system.

Class diagram is also considered as the foundation for component and deployment diagrams. Class diagrams are not only used to visualize the static view of the system but they are also used to construct the executable code for forward and reverse engineering of any system.

Generally, UML diagrams are not directly mapped with any object-oriented programming languages but the class diagram is an exception.

Class diagram clearly shows the mapping with object-oriented languages such as Java, C++, etc. From practical experience, class diagram is generally used for construction purpose.



In a nutshell it can be said, class diagrams are used for –

- ❖ Describing the static view of the system.
- ❖ Showing the collaboration among the elements of the static view.
- ❖ Describing the functionalities performed by the system.
- ❖ Construction of software applications using object-oriented languages.

Use Case Diagram:

To model a system, the most important aspect is to capture the dynamic behaviour. Dynamic behaviour means the behaviour of the system when it is running/operating.

Only static behaviour is not sufficient to model a system rather dynamic behaviour is more important than static behaviour. In UML, there are five diagrams available to model the dynamic nature and use case diagram is one of them. Now as we have to discuss that the use case diagram is dynamic in nature, there should be some internal or external factors for making the interaction.

These internal and external agents are known as actors. Use case diagrams consists of actors, use cases and their relationships. The diagram is used to model the system/subsystem of an application. A single use case diagram captures a particular functionality of a system.

Hence to model the entire system, a number of use case diagrams are used.

Purpose of Use Case Diagrams

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence,

collaboration, and Statechart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams.

Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analysed to gather its functionalities, use cases are prepared and actors are identified.

When the initial task is complete, use case diagrams are modelled to present the outside view.

In brief, the purposes of use case diagrams can be said to be as follows –

- ❖ Used to gather the requirements of a system.
- ❖ Used to get an outside view of a system.

- ❖ Identify the external and internal factors influencing the system.
- ❖ Show the interaction among the requirements are actors.

How to Draw a Use Case Diagram?

Use case diagrams are considered for high level requirement analysis of a system. When the requirements of a system are analysed, the functionalities are captured in use cases.

We can say that use cases are nothing but the system functionalities written in an organized manner. The second thing which is relevant to use cases are the actors. Actors can be defined as something that interacts with the system.

Actors can be a human user, some internal applications, or may be some external applications. When we are planning to draw a use case diagram, we should have the following items identified.

Functionalities to be represented as use case

- ❖ Actors
- ❖ Relationships among the use cases and actors.

Use case diagrams are drawn to capture the functional requirements of a system.

Where to Use a Use Case Diagram?

There are five diagrams in UML to model the dynamic view of a system. Now each and every model has some specific purpose to use. Actually, these specific purposes are different angles of a running system.

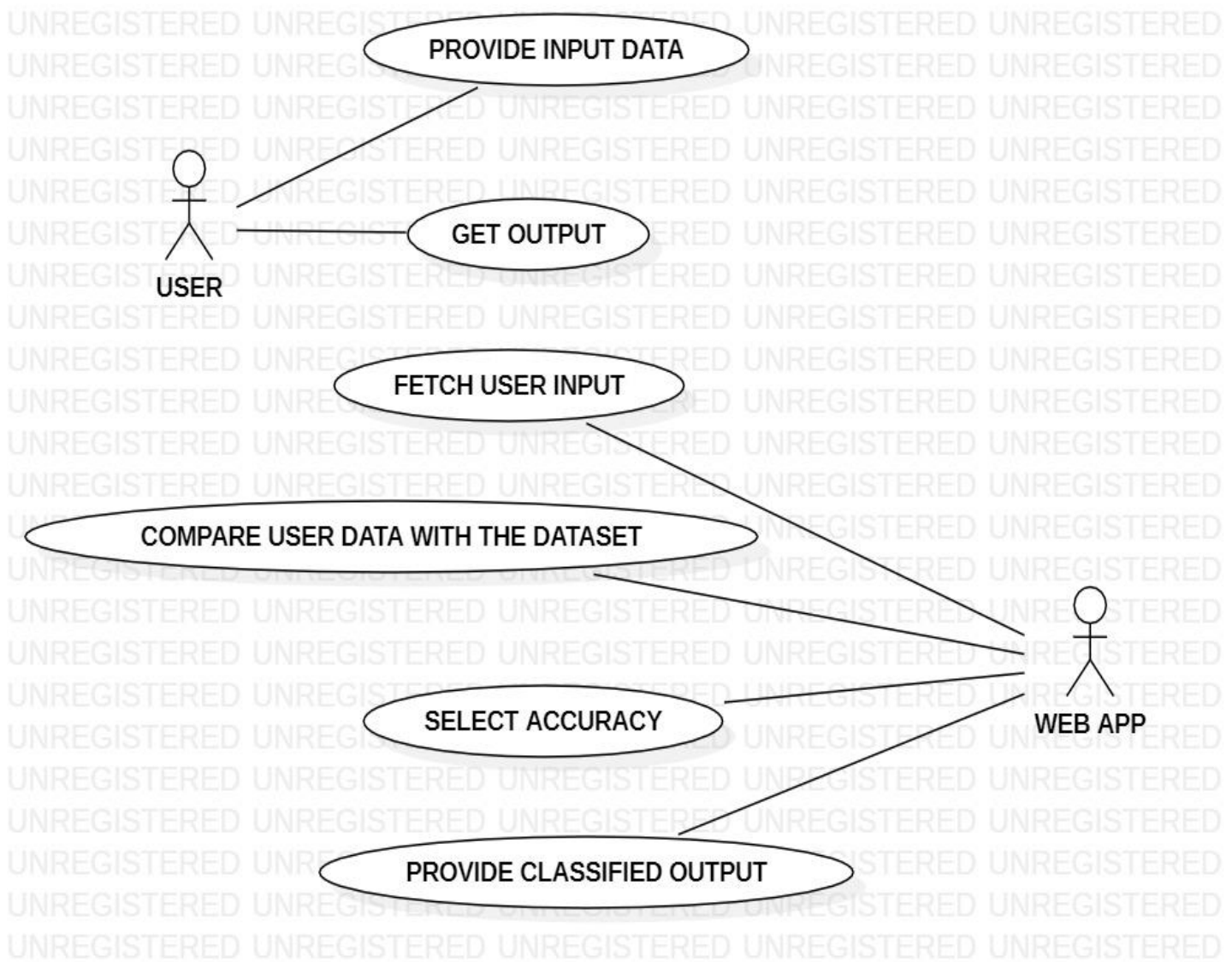
To understand the dynamics of a system, we need to use different types of diagrams. Use case diagram is one of them and its specific purpose is to gather system requirements and actors.

Use case diagrams specify the events of a system and their flows. But use case diagram never describes how they are implemented. Use case diagram can be imagined as a black box where only the input, output, and the function of the black box is known.

These diagrams are used at a very high level of design. This high-level design is refined again and again to get a complete and practical picture of the system. A well-structured use case also describes the pre-condition, post condition, and exceptions. These extra elements are used to make test cases when performing the testing.

Although use case is not a good candidate for forward and reverse engineering, still they are used in a slightly different way to make forward and reverse engineering. The same is true for reverse engineering. Use case diagram is used differently to make it suitable for reverse engineering.

In forward engineering, use case diagrams are used to make test cases and in reverse engineering use cases are used to prepare the requirement details from the existing application.



Use case diagrams can be used for –

- ❖ Requirement analysis and high-level design.
- ❖ Model the context of a system.
- ❖ Reverse engineering.
- ❖ Forward engineering.

CONCLUSION & FUTURE SCOPE

Employment scam detection will guide job-seekers to get only legitimate offers from companies. For tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this paper. Supervised mechanism is used to exemplify the use of several classifiers for employment scam detection. Experimental results indicate that Random Forest classifier outperforms over its peer classification tool. The proposed approach achieved accuracy 98.27% which is much higher than the existing methods.

In future it can reduce the fake job notification, with the help of this project people can differentiate between fake and real job notification. So people will be able to find the real jobs from different types of social media platform.