A

Project Report

On

MACHINE LEARNING MODEL DEPLOYED ON WEB PAGE

Submitted in partial fulfillment of the requirement for the VI semester

Bachelors Of Computer Science Engineering

By

Ashwani Kaira(2061783)

Gautam Singh

Adhikari(2061815)

Anugrah George(2061778)

Under the Guidance of – Mr. Ravindra Koranga

Name of Faculty – Mr. Ravindra Koranga

DEPARTMENT OF SCHOOL OF COMPUTING
GRAPHIC ERA HILL UNIVERSITY, BHIMTAL CAMPUS
SATTAL ROAD, P.O. BHOWALI,
DISTRICT- NAINITAL-263132
2022 - 2023

STUDENT'S DECLARATION

We Ashwani Kaira(2061783),Gautam Singh Adhikari(2061815) and Anugrah George(2061778) , here by declare the work, which is being presented in the project, entitled "Machine learning model deployed on web page " in partial fulfillment of the requirement for **B.Tech [CSE] VI sem** as a mini project is an authentic record of my own work carried out under the supervision of GEHU Bhimtal college /company guide name Mr. Ravindra Koranga The matter embodied in this project has not been submitted by me for the award of any other degree.

Date: 10/06/2023   ……………….
                                                   (Full signature of student)

I

<u>CERTIFICATE</u>

The project report entitled "Machine learning model deployed on web page" being submitted by Ashwani Kaira(2061783), Gautam Singh Adhikari(2061815), Anugrah George(2061778). To Graphic Era Hill University Bhimtal Campus for the award of bonafide work carried out by him. He has worked under my guidance and supervision and fulfilled the requirement for the submission of report.

(Mr. Ravindra Koranga)
Project Guide (CSE Dept.)

II

<u>ACKNOWLEDGEMENT</u>

III

<u>TABLE OF CONTENTS</u>

IV

CHAPTER 1:-INTRODUCTION

1.1 Prologue

As students are going through their academics and pursuing their interested courses, it is veryimportant for them to assess their capabilities and identify their interests so that they will get toknow in which career area their interests and capabilities are going to put them in. This will helpthem in improving their performance and motivating their interests so that they will be directedtowards their targeted career and get settled in that. Also recruiters while recruiting thecandidates after assessing them in all different aspects, these kind of career recommendersystems help them in deciding in which job role the candidate should be kept in based on his/her performance and other evaluations. This paper mainly concentrates on the career area predictionof computer science domain candidates.

1.1 In this project we will create a model which is able to predict the chances of getting job or not. If any student is able to get job then his model is also capable to predict the chances that in which field he/she get the job. This model takes student's skills as input and analysis based on input and dataset.After doing analysis this model will return prediction as output.

1.2 Background and Motivations

Solve a Problem:

Do you have a problem that you think machine learning can solve?Maybe it's an open business problem or a problem at work. Maybe it's an opportunity you can see in the market. Nevertheless, you are thinking about machine learning as a tool for you to learn and apply to a problem.In this case you may be interested to learn tools that provide implementations of algorithms you could use quickly. You will also very likely be interested in the creative ways to use these tools, such as case studies on problems like the problem you want to solve.

1.3 **Technical Achievement**:

Is learning machine learning a mark of achievement?

Maybe machine learning is a popular technical field and you get great pride from learning new and difficult technologies and tools. Maybe you see machine learning as your next big challenge and opportunity for growth and a chance to demonstrate your abilities to learn and master technical materials.If this sounds like you, you may be interested in books of algorithms where you can fast track getting an understanding of a method and how to use it without having to get down into the latest research. You will also very likely be interested in completing courses, entering competitions and implementing algorithms yourself.Certain tasks are extremely difficult to program by hand:

Spam Filtering: The upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust antispam filters. Machine learning methods of recent are being used to successfully detect and filter spam emails. We present a systematic review of some of the popular machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and the research trend in spam filtering. The preliminary discussion in the study background examines the applications of machine learning techniques to the email spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use machine learning techniques was done.

Face Recognition: Facial recognition is a technology that is capable of recognizing a person based on their face. It employs machine learning algorithms which find, capture, store and analyse facial features in order to match them with images of individuals in a pre-existing database.

Machine Translation: Machine translation is the task of automatically converting source text in one language to text in another language. In a machine translation task, the input already consists of a sequence of symbols in some language, and the computer program must convert this into a sequence of symbols in another language

Speech Recognition: Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice.

Data Mining: Data mining is a subset of business analytics and refers to exploring an existing large dataset to unearth previously unknown patterns, relationships and anomalies that are present in the data. It gives us the ability to find completely new insights that we weren't necessarily looking for – unknown unknowns, if you like.

CHAPTER 2:- PROPOSED SYSTEM

2.1 History of Machine Learning:

It's all well and good to ask if androids dream of electric sheep, but science fact has evolved to a point where it's beginning to coincide with science fiction. No, we don't have autonomous androids struggling with existential crises — yet — but we are getting ever closer to what people tend to call "artificial intelligence."Machine Learning is a sub-set of artificial intelligence where computer algorithms are used to autonomously learn from data and information. In machine learning computers don't have to be explicitly programmed but can change and improve their algorithms by themselves.Today, machine learning algorithms enable computers to communicate with humans, autonomously drive cars, write and publish sport match reports, and find terrorist suspects. I firmly believe machine learning will severely impact most industries and the jobs within them, which is why every manager should have at least some grasp of what machine learning is and how it is evolving.

Python library used in Machine Learning

Machine Learning, as the name suggests, is the science of programming a computer by which theare able to learn from different kinds of data. A more general definition given by Arthur Samuel is – "Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed." They are typically used to solve various types of life problems.In the older days, people used to perform Machine Learning tasks by manually coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in Machine Learning are:

Numpy
Scikit-learn
Pandas
Pickle
Flask

1. Numpy:
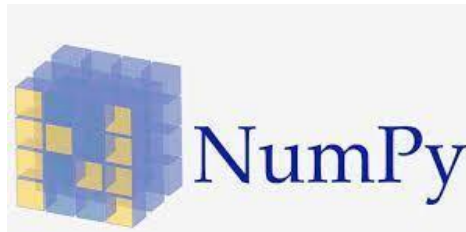
Numpy is not a programming language, it is a Python Extension whose purpose is to provide functions and capability to transform arrays.Following are the purpose of Working with Numpy:

Smaller Memory Consumption than List

Implementation of Multi-Dimensional Arrays

NumPy arrays are faster than Python List

NumPy can be used to transform the Arrays

Python does not have inbuilt support for Arrays

Offers function like Reshape, Sort, Reverse, etc

How do I get NumPy?

In order to install Numpy Package, we need to use the following command:

pip install numpy

This command will install this Python Module to your local system. For mere detailed installations steps on the other platform.

Numpy Functions for Machine Learning

In the article below, we will list down the common features and functions that can be used in machine learning for transforming the input data.

Creating a Numpy Array

Numpy can be created using a list or a tuple. Usually in Machine Learning, we would import data is some form and then convert the data into NumPy array.

Reshaping Array Dimensions

This is one of the most common requirements in Machine Learning, we need to reshape the construct of an array. For example, we can convert a 1 Dimensional Array into 2 Dimensional Array. Ofter, the machine learning algorithms take a 2 Dimensional array as input, but if we have a single row as an input, we need to convert this single dimension to multiple dimensions. You will see the requirement of this function a lot.

Finding out the Length of the Numpy Array

Finding out the length of the array is another commonly used function. In order to find the length of the array, we need to use "len" function of Python.

Slicing NumPy Arrays

Slicing Array represents finding the subset of the Array. We can define the subset of NumPy array by specifying the starting and ending index. Given below are the code blocks for different operations that can be applied to the slicing of an Array.

Joining NumPy Arrays

Joining multiple Array s together is called Concatenation. This can be achieved using the "concatenate()" function.

Sorting Arrays in NumPy

Sorting involves putting the elements in Sequence. The elements can be arranged in ascending order (if elements are numeric) and can be arranged in Alphabetical Format (if the elements are Strings).

2. Scikit-learn:

Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

Please note that sklearn is used to build machine learning models. It should not be used for reading the data, manipulating and summarizing it. There are better libraries for that (e.g. NumPy, Pandas etc.)

Components of scikit-learn:

Scikit-learn comes loaded with a lot of features. Here are a few of them to help you understand the spread:

Supervised learning algorithms: Think of any supervised machine learning algorithm you might have heard about and there is a very high chance that it is part of scikit-learn. Starting from Generalized linear models (e.g Linear Regression), Support Vector Machines (SVM), Decision Trees to Bayesian methods – all of them are part of scikit-learn toolbox. The spread of machine learning algorithms is one of the big reasons for the high usage of scikit-learn. I started using scikit to solve supervised learning problems and would recommend that to people new to scikit / machine learning as well.

Cross-validation: There are various methods to check the accuracy of supervised models on unseen data using sklearn.

Unsupervised learning algorithms: Again there is a large spread of machine learning algorithms in the offering – starting from clustering, factor analysis, principal component analysis to unsupervised neural networks.

Various toy datasets: This came in handy while learning scikit-learn. I had learned SAS using various academic datasets (e.g. IRIS dataset, Boston House prices dataset). Having them handy while learning a new library helped a lot.

5

Feature extraction: Scikit-learn for extracting features from images and text (e.g. Bag of words).

Community / Organizations using scikit-learn:

One of the main reasons behind using open source tools is the huge community it has. Same is true for sklearn as well. There are about 35 contributors to scikit-learn till date, the most notable being Andreas Mueller (P.S. Andy's machine learning cheat sheet is one of the best visualizations to understand the spectrum of machine learning algorithms).

There are various Organizations of the likes of Evernote, Inria and AWeber which are being displayed on scikit learn home page as users. But I truly believe that the actual usage is far more.
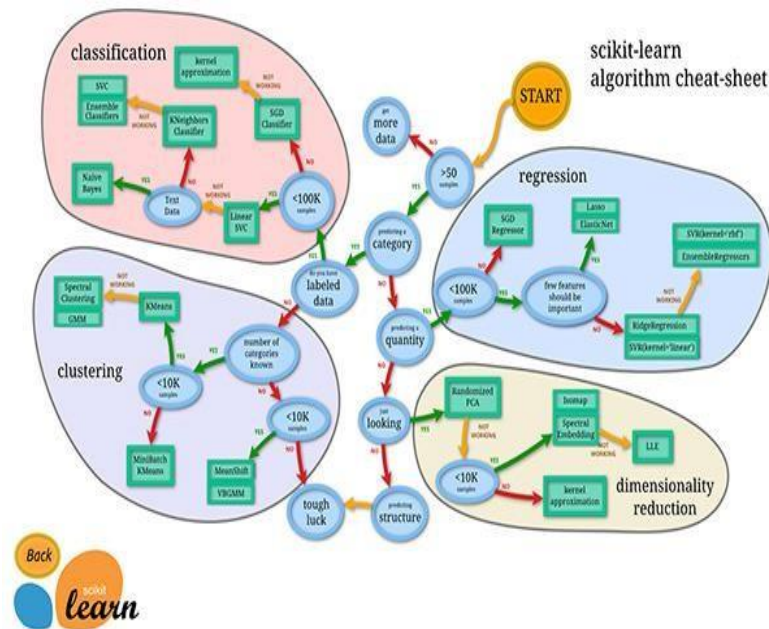
In addition to these communities, there are various meetups across the globe. There was also a Kaggle knowledge contest, which finished recently but might still be one of the best places to start playing around with the library.

3. **Pandas**:

The pandas package is the most important tool at the disposal of Data Scientists and Analysts working in Python today. The powerful machine learning and glamorous visualization tools may get all the attention, but pandas is the backbone of most data projects.

If you're thinking about data science as a career, then it is imperative that one of the first things you do is learn pandas. In this report, we will go over the essential bits of information about pandas, including how to install it, its uses, and how it works with other common Python data analysis packages such as matplotlib and scikit-learn.

What's Pandas for?

Pandas has so many uses that it might make sense to list the things it can't do instead of what it can do.
This tool is essentially your data's home. Through pandas, you get acquainted with your data by cleaning, transforming, and analyzing it.
For example, say you want to explore a dataset stored in a CSV on your computer. Pandas will extract the data from that CSV into a DataFrame — a table, basically — then let you do things like:
Calculate statistics and answer questions about the data, like
What's the average, median, max, or min of each column?
Does column A co-relate with column B?
What does the distribution of data in column C look like?
Clean the data by doing things like removing missing values and filtering rows or columns by some criteria
Visualize the data with help from Matplotlib. Plot bars, lines, histograms, bubbles, and more.
Store the cleaned, transformed data back into a CSV, other file or database
Before you jump into the modeling or the complex visualizations you need to have a good understanding of the nature of your dataset and pandas is the best avenue through which to do that.

How does pandas fit into the data science toolkit?Not only is the pandas library a central component of the data science toolkit but it is used in conjunction with other libraries in that collection.Pandas is built on top of the NumPy package, meaning a lot of the structure of NumPy is used or replicated in Pandas.

7

4. Pickle:

Pickle is a module in Python used for serializing and de-serializing Python objects. This converts Python objects like lists, dictionaries, etc. into byte streams (zeroes and ones). You can convert the byte streams back into Python objects through a process called unpickling. Pickling is also known as serialization, flattening, or marshalling.

How to Pickle

The pickle module has two methods.

pickle.dump()

The pickle.dump() method dumps the Python object in the pickle file. This creates a .pickle file in your current working directory (pickle.dump(what_are_we_dumping, where_are_we_dumping_it)):

In the code snippet above, we are creating an example_dict.pickle file from an example_dict dictionary. Statements 1 and 2 perform the same task of converting the dictionary into a pickle file. Using the with statement ensures that open file descriptors are closed automatically after the program execution leaves the context of the with statement. The 'wb' in the open statement means we are writing bytes to file.

pickle.load()

The pickle.load() method lets you use the .pickle file (pickle.load(what_do_we_want_to_load)) by loading it in the memory:

When to Pickle

- To save a program's state.
- To send Python data over a TCP connection.
- To save Python objects.

When Not to Pickle

When working with multiple Python versions: Unpickling objects pickled in different Python versions can be a hassle.

When working across multiple languages: The data format used by pickle is Python-specific, which means that non-Python programs may not be able to reconstruct pickled Python objects.

When working with a recursive data structure: Trying to pickle a highly recursive data structure may exceed the maximum recursion depth. A RuntimeError will be raised in this case. You can raise this limit with sys.setrecursionlimit().

FLASK:-

What is Flask?

Flask is a web framework that provides libraries to build lightweight web applications in python. It is developed by **Armin Ronacher** who leads an international group of python enthusiasts (POCCO). It is based on WSGI toolkit and jinja2 template engine. Flask is considered as a micro framework.

What is WSGI?

It is an acronym for web server gateway interface which is a standard for python web application development. It is considered as the specification for the universal interface between the web server and web application.

.What is Jinja2?

Jinja2 is a web template engine which combines a template with a certain data source to render the dynamic web pages.

Flask Environment Setup

To install flask on the system, we need to have python 2.7 or higher installed on our system. However, we suggest using python 3 for the development in the flask.

9

CHAPTER 3:- S/W AND H/W REQUIREMENTS

3.1 S/W AND H/W REQUIREMENTS

Processor – Intel Xeon E2630 v4 – 10 core processor, 2.2 GHz with Turboboost upto 3.1 GHz. 25 MB Cache

Motherboard – ASRock EPC612D8A

RAM – 128 GB DDR4 2133 MHz

2 TB Hard Disk (7200 RPM) + 512 GB SSD

GPU – NVidia TitanX Pascal (12 GB VRAM)

Intel Heatsink to keep temperature under control

The product is mainly a text summarizing Google Chrome extension using Deep

Learning concepts. The main purpose is to provide reliable summaries of web pages or

uploaded files depends on the user's choice. The unnecessary sentences will be discarded to

obtain the most important sentences.

The product includes the following components:

• Text Parser: It will divide the texts into paragraphs, sentences and words.

• Feature Vector Creator: This component will calculate and get the feature

representations of sentences.

• Autoencoder: The root part of the Deep Learning. Autoencoder offers a compressed

representation of a given sentence.

• Classifier: The classifier determines if a sentence is a summary sentence or not.nstall python.

3.1.1 <u>Security</u>:

We are working on host machine only so security is not a problem.AI and ML require more data, and more complex data, than other technologies. The algorithms developed by mathematicians and data scientists come out of research projects. "We're only recently as a scientific community coming to understand that there are security issues with AI," says Raff.

CHAPTER 4:- DFD (DATA FLOW DIAGRAM)
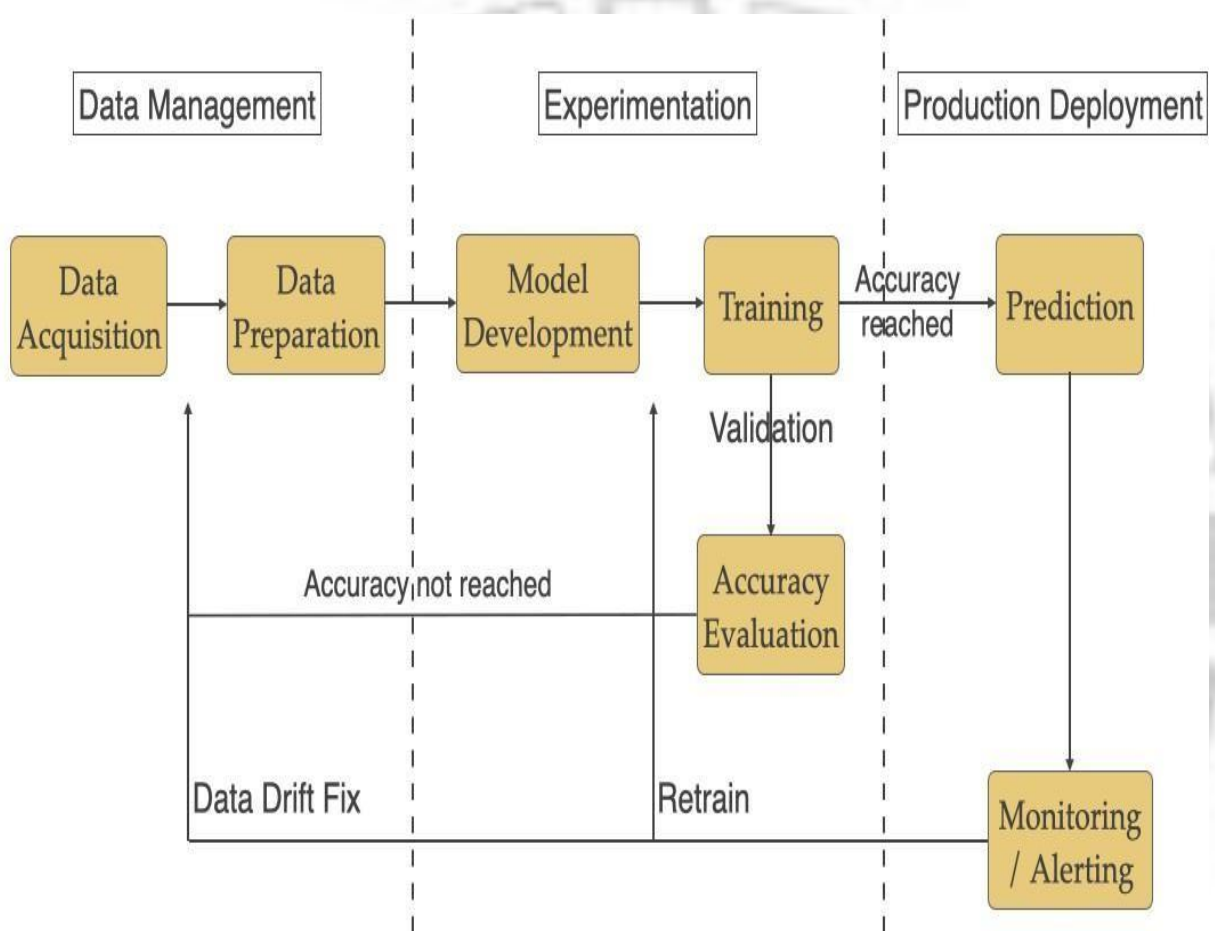
CHAPTER 5:- DATABASE DESIGN (APPLICABLE IN CASE THE PROJECT HAS DBMS AT BACKEND)

Project does not has dbms at the backend…. We have datset image of which is given below:-

Cell reference: A1 — Formula bar: `mid`

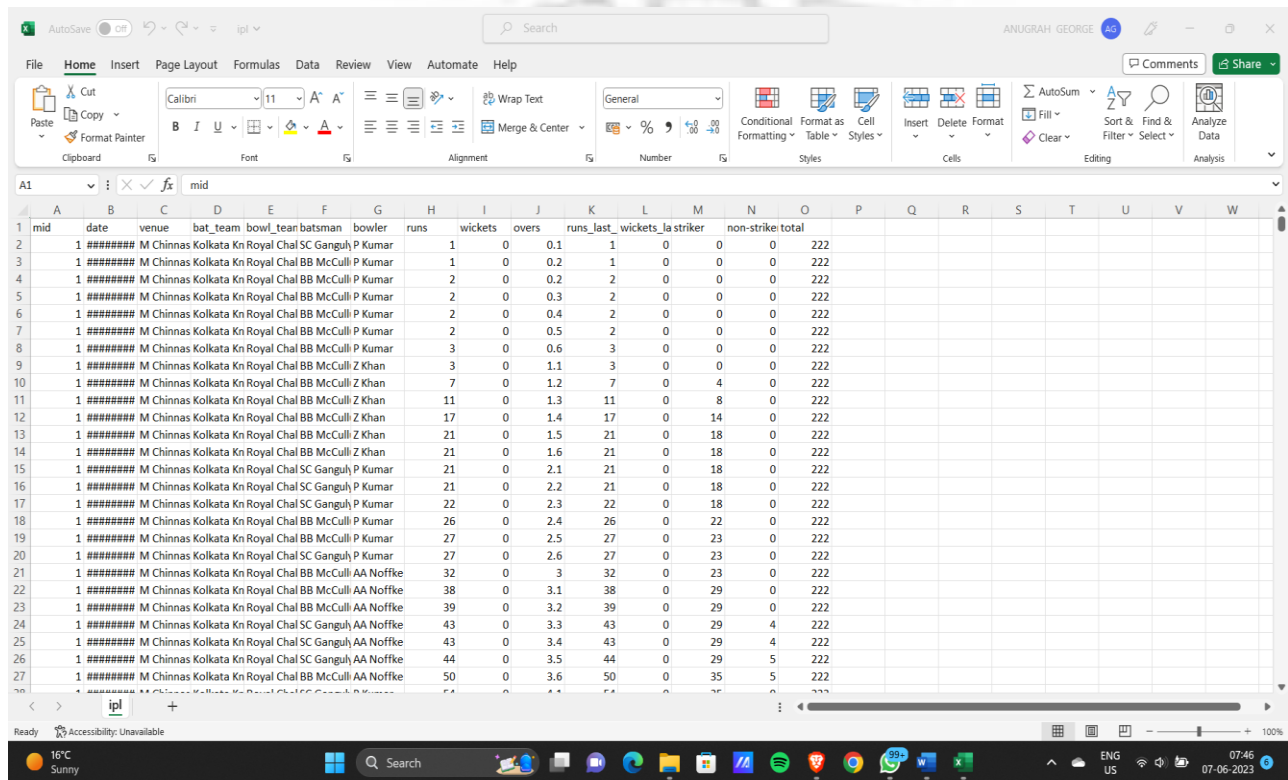|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
21388	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	VS Malik	42	1	4.3	42	1	17	5	179
21389	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	VS Malik	42	1	4.4	42	1	17	5	179
21390	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	VS Malik	42	1	4.5	42	1	17	5	179
21391	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	VS Malik	46	1	4.6	45	1	17	9	179
21392	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	KK Cooper	48	1	5.1	47	1	19	9	179
21393	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	KK Cooper	52	1	5.2	51	1	23	9	179
21394	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	KK Cooper	53	1	5.3	48	1	24	9	179
21395	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	KK Cooper	53	1	5.4	44	1	24	9	179
21396	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	KK Cooper	53	1	5.5	38	1	24	9	179
21397	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	KK Cooper	53	1	5.6	38	1	24	9	179
21398	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	R Tewatia	53	1	6.1	34	1	24	9	179
21399	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	R Tewatia	53	1	6.2	33	1	24	9	179
21400	440	########	Punjab Cri	Kings XI Pu	Rajasthan	M Vohra	R Tewatia	54	2	6.3	34	2	25	9	179
21401	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	R Tewatia	55	2	6.4	35	2	9	1	179
21402	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	55	2	6.5	35	2	9	1	179
21403	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	56	2	6.6	36	2	10	1	179
21404	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	PV Tambe	57	2	7.1	33	2	11	1	179
21405	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	PV Tambe	57	2	7.2	33	1	11	1	179
21406	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	PV Tambe	61	2	7.3	37	1	11	5	179
21407	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	PV Tambe	65	2	7.4	41	1	11	9	179
21408	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	PV Tambe	66	2	7.5	41	1	11	10	179
21409	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	PV Tambe	67	2	7.6	41	1	12	10	179
21410	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	68	2	8.1	38	1	13	10	179
21411	440	########	Punjab Cri	Kings XI Pu	Rajasthan	WP Saha	R Tewatia	69	2	8.2	33	1	13	11	179
21412	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	69	2	8.3	33	1	13	11	179
21413	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	73	2	8.4	37	1	17	11	179
21414	440	########	Punjab Cri	Kings XI Pu	Rajasthan	SE Marsh	R Tewatia	74	2	8.5	38	1	18	11	179

14

CHAPTER 6:- ER DIAGRAM

Since database is not used so ER diagram is not made…

CHAPTER 7:- CODING OF FUNCTION

CODE:- 1. ML model training code.
           2. python code for deploying ml code to webpage using flask.
           3. html file

ML model training code:-

```
# Importing essential libraries
import pandas as pd
import pickle

# Loading the dataset
df = pd.read_csv('ipl.csv')

# --- Data Cleaning ---
# Removing unwanted columns
columns_to_remove = ['mid', 'venue', 'batsman', 'bowler', 'striker', 'non-striker']
df.drop(labels=columns_to_remove, axis=1, inplace=True)

# Keeping only consistent teams
consistent_teams = ['Kolkata Knight Riders', 'Chennai Super Kings', 'Rajasthan Royals',
            'Mumbai Indians', 'Kings XI Punjab', 'Royal Challengers Bangalore',
            'Delhi Daredevils', 'Sunrisers Hyderabad']
df = df[(df['bat_team'].isin(consistent_teams)) & (df['bowl_team'].isin(consistent_teams))]

# Removing the first 5 overs data in every match
df = df[df['overs']>=5.0]

# Converting the column 'date' from string into datetime object
from datetime import datetime
df['date'] = df['date'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d'))

# --- Data Preprocessing ---
# Converting categorical features using OneHotEncoding method
encoded_df = pd.get_dummies(data=df, columns=['bat_team', 'bowl_team'])
# Rearranging the columns
encoded_df = encoded_df[['date', 'bat_team_Chennai Super Kings', 'bat_team_Delhi Daredevils', 'bat_team_Kings
XI Punjab',
        'bat_team_Kolkata Knight Riders', 'bat_team_Mumbai Indians', 'bat_team_Rajasthan Royals',
        'bat_team_Royal Challengers Bangalore', 'bat_team_Sunrisers Hyderabad',
        'bowl_team_Chennai Super Kings', 'bowl_team_Delhi Daredevils', 'bowl_team_Kings XI Punjab',
        'bowl_team_Kolkata Knight Riders', 'bowl_team_Mumbai Indians', 'bowl_team_Rajasthan Royals',
        'bowl_team_Royal Challengers Bangalore', 'bowl_team_Sunrisers Hyderabad',
        'overs', 'runs', 'wickets', 'runs_last_5', 'wickets_last_5', 'total']]
```

 # Splitting the data into train and test set
```python
X_train = encoded_df.drop(labels='total', axis=1)[encoded_df['date'].dt.year <= 2016]
X_test = encoded_df.drop(labels='total', axis=1)[encoded_df['date'].dt.year >= 2017]

y_train = encoded_df[encoded_df['date'].dt.year <= 2016]['total'].values
y_test = encoded_df[encoded_df['date'].dt.year >= 2017]['total'].values

# Removing the 'date' column
X_train.drop(labels='date', axis=True, inplace=True)
X_test.drop(labels='date', axis=True, inplace=True)
```

 # --- Model Building ---
 # Linear Regression Model
```python
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train)
```

 # Creating a pickle file for the classifier
```python
filename = 'first-innings-score-lr-model.pkl'
pickle.dump(regressor, open(filename, 'wb'))
```

CREATING WEB PAGE WITH FLASK

App.py
 # Importing essential libraries
```python
from flask import Flask, render_template, request
import pickle
import numpy as np
```

 # Load the Random Forest CLassifier model
```python
filename = 'first-innings-score-lr-model.pkl'
regressor = pickle.load(open(filename, 'rb'))
app = Flask(_name_)
@app.route('/')
def home():
  return render_template('index.html')
@app.route('/predict', methods=['POST'])
def predict():
   temp_array = list()
  if request.method == 'POST':
   batting_team = request.form['batting-team']
if batting_team == 'Chennai Super Kings':
       temp_array = temp_array + [1,0,0,0,0,0,0,0]
     elif batting_team == 'Delhi Daredevils':
       temp_array = temp_array + [0,1,0,0,0,0,0,0]
     elif batting_team == 'Kings XI Punjab':
```

17

```python
            temp_array = temp_array + [0,0,1,0,0,0,0,0]
        elif batting_team == 'Kolkata Knight Riders':
            temp_array = temp_array + [0,0,0,1,0,0,0,0]
        elif batting_team == 'Mumbai Indians':
            temp_array = temp_array + [0,0,0,0,1,0,0,0]
        elif batting_team == 'Rajasthan Royals':
            temp_array = temp_array + [0,0,0,0,0,1,0,0]
        elif batting_team == 'Royal Challengers Bangalore':
            temp_array = temp_array + [0,0,0,0,0,0,1,0]
        elif batting_team == 'Sunrisers Hyderabad':
            temp_array = temp_array + [0,0,0,0,0,0,0,1]

        bowling_team = request.form['bowling-team']
        if bowling_team == 'Chennai Super Kings':
            temp_array = temp_array + [1,0,0,0,0,0,0,0]
        elif bowling_team == 'Delhi Daredevils':
            temp_array = temp_array + [0,1,0,0,0,0,0,0]
        elif bowling_team == 'Kings XI Punjab':
            temp_array = temp_array + [0,0,1,0,0,0,0,0]
        elif bowling_team == 'Kolkata Knight Riders':
            temp_array = temp_array + [0,0,0,1,0,0,0,0]
        elif bowling_team == 'Mumbai Indians':
            temp_array = temp_array + [0,0,0,0,1,0,0,0]
        elif bowling_team == 'Rajasthan Royals':
            temp_array = temp_array + [0,0,0,0,0,1,0,0]
        elif bowling_team == 'Royal Challengers Bangalore':
            temp_array = temp_array + [0,0,0,0,0,0,1,0]
        elif bowling_team == 'Sunrisers Hyderabad':
            temp_array = temp_array + [0,0,0,0,0,0,0,1]


        overs = float(request.form['overs'])
        runs = int(request.form['runs'])
        wickets = int(request.form['wickets'])
        runs_in_prev_5 = int(request.form['runs_in_prev_5'])
        wickets_in_prev_5 = int(request.form['wickets_in_prev_5'])

        temp_array = temp_array + [overs, runs, wickets, runs_in_prev_5, wickets_in_prev_5]

        data = np.array([temp_array])
        my_prediction = int(regressor.predict(data)[0])

        return render_template('result.html', lower_limit = my_prediction-10, upper_limit =
my_prediction+5)
```

```
if__name__== '_main_':
  app.run(debug=True)
```

HTML FILE:-

Index.html

```
<!DOCTYPE html>
<html lang="en" dir="ltr">
   <head>
     <meta charset="utf-8">
     <title>First Innings Score Predictor</title>
     <link rel="shortcut icon" href="{{ url_for('static', filename='ipl-favicon.ico') }}">
     <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='styles.css') }}">
     <script src="https://kit.fontawesome.com/5f3f547070.js" crossorigin="anonymous"></script>
     <link href="https://fonts.googleapis.com/css2?family=Open+Sans:wght@300&display=swap"
rel="stylesheet">
   </head>

   <body style="background-color:black">


     <!-- Website Title -->
     <div class="container">
       <h1 class='container-heading'> <span class='heading_first'><b><font color="white" size="9">
Innings Score Predictor for <!--</span><span class="heading_second">--!>Indian Premier League
(IPL)</font>font </b></span></h1>
         <div class='description'>
           <br>
           <font color="white" size="6"> A Machine Learning Web App, Built with Flask.</font>
       </div>
     </div>

     <!-- Prediction Area -->
     <div class="prediction-area">

       <!-- IPL Banner - 1 -->
       <div class='ipl-banner-left'>

         <div class="slider-frame">
            <div class="slide-images-up">
<!-- Image-1 -->
              <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='mi.jpg') }}" alt="Mumbai
```
19

```
 Indians">
                </div>
                <!-- Image-2 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='srh.png') }}" alt="Sunrisers
Hyderabad">
                </div>
                <!-- Image-3 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='rcb.png') }}" alt="Royal
Challengers Bangalore">
                </div>
                <!-- Image-4 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='dc.png') }}" alt="Delhi Capitals">
                </div>
                <!-- Image-5 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='rr.png') }}" alt="Rajasthan
Royals">
                </div><!-- Image-6 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='kkr.jpg') }}" alt="Kolkata Knight
Riders">
                </div> <!-- Image-7 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='kxip.png') }}" alt="Kings XI
Punjab">
                </div><!-- Image-8 -->
                <div class="image-container">
                  <img class="image-width" src="{{ url_for('static', filename='csk.png') }}" alt="Chennai Super
Kings">
                </div> </div>  /div>
<!-- Prediction Form -->
        <div class='div2'>
          <form action="{{ url_for('predict') }}" method="POST">
            <!-- Batting Team Dropdown -->
            <select class="form-input align-center" name="batting-team">
              <option value="none">--- Select a Batting team ---</option>
                <option value="Mumbai Indians">Mumbai Indians</option>
                <option value="Kolkata Knight Riders">Kolkata Knight Riders</option>
                <option value="Chennai Super Kings">Chennai Super Kings</option>
                <option value="Rajasthan Royals">Rajasthan Royals</option>
                <option value="Kings XI Punjab">Kings XI Punjab</option>
                <option value="Royal Challengers Bangalore">Royal Challengers Bangalore</option>
                <option value="Delhi Daredevils">Delhi Daredevils</option>
                <option value="Sunrisers Hyderabad">Sunrisers Hyderabad</option>
              </select><br>
```

20

```
<!-- Bowling Team Dropdown -->
            <select class="form-input align-center" name="bowling-team">
                <option value="none">--- Select a Bowling team ---</option>
                <option value="Mumbai Indians">Mumbai Indians</option>
                <option value="Kolkata Knight Riders">Kolkata Knight Riders</option>
                <option value="Chennai Super Kings">Chennai Super Kings</option>
                <option value="Rajasthan Royals">Rajasthan Royals</option>
                <option value="Kings XI Punjab">Kings XI Punjab</option>
                <option value="Royal Challengers Bangalore">Royal Challengers Bangalore</option>
                <option value="Delhi Daredevils">Delhi Daredevils</option>
                <option value="Sunrisers Hyderabad">Sunrisers Hyderabad</option>
            </select><br>
            <input class="form-input" type="text" name="overs" placeholder="Overs (>= 5.0) eg.
7.2"><br>
            <input class="form-input" type="text" name="runs" placeholder="Runs eg. 64"><br>
            <input class="form-input" type="text" name="wickets" placeholder="Wickets eg. 4"><br>
            <input class="form-input" type="text" name="runs_in_prev_5" placeholder="Runs scored
in previous 5 Overs eg. 42"><br>
            <input class="form-input" type="text" name="wickets_in_prev_5" placeholder="Wickets
taken in previous 5 Overs eg. 3"><br>

    <input type="submit" class="myButton" value="Predict Score">
    </form>
     </div>

     <!-- IPL Banner - 2 -->
     <div class='ipl-banner-right'>

        <div class="slider-frame">
           <div class="slide-images-down">

            <!-- Image-1 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='mi.jpg') }}" alt="Mumbai
Indians">
            </div>
            <!-- Image-2 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='srh.png') }}"
alt="Sunrisers Hyderabad">
            </div>
            <!-- Image-3 -->
```

21

```html
<div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='rcb.png') }}" alt="Royal
Challengers Bangalore">
                </div>
                <!-- Image-4 -->
                <div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='dc.png') }}" alt="Delhi
Capitals">
                </div>
                <!-- Image-5 -->
                <div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='kkr.jpg') }}"
alt="Kolkata Knight Riders">
                </div>
                <!-- Image-6 -->
                <div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='rr.png') }}"
alt="Rajasthan Royals">
                </div>
                <!-- Image-7 -->
                <div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='kxip.png') }}" alt="Kings
XI Punjab">
                </div>
                <!-- Image-8 -->
                <div class="image-container">
                    <img class="image-width" src="{{ url_for('static', filename='csk.png') }}"
alt="Chennai Super Kings">
                </div>
</div>
 </div>
 </div>
 </div>
  <!-- Footer -->
  <!--
 <div class='footer'>
 <div class="contact">
 <h2> <center> <b><font color="white"> Made by Anugrah George </font></b></center></h2></p>
 </div>
 </body>
</html>
```

22

Result.html

```html
<!DOCTYPE html>

<html lang="en" dir="ltr">
  <head>
    <meta charset="utf-8">
    <title>First Innings Score Predictor</title>
    <link rel="shortcut icon" href="{{ url_for('static', filename='ipl-favicon.ico') }}">
    <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='styles.css') }}">
    <script src="https://kit.fontawesome.com/5f3f547070.js" crossorigin="anonymous"></script>
    <link href="https://fonts.googleapis.com/css2?family=Open+Sans:wght@300&display=swap"
rel="stylesheet">
  </head>

  <body bgcolor='cyan'>

    <!-- Website Title -->
    <div class="container">
      <h1 class='container-heading'> <span class='heading_first'> <b> Innings Score Predictor for <!--
</span><span class="heading_second">--!>Indian Premier League (IPL)</b></span></h1>
      <div class='description'>
        <p> <h2>A Machine Learning Web App, Built with Flask.</h2></p>
      </div>
    </div>

    <!-- Prediction Area -->
    <div class="prediction-area">

      <!-- IPL Banner - 1 -->
      <div class='ipl-banner-left'>

        <div class="slider-frame">
          <div class="slide-images-up">

            <!-- Image-1 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='mi.jpg') }}" alt="Mumbai
Indians">
            </div>
            <!-- Image-2 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='srh.png') }}"
alt="Sunrisers Hyderabad">
            </div> <!-- Image-3 -->
```

23

```html
<div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='rcb.png') }}" alt="Royal
Challengers Bangalore">
            </div>
            <!-- Image-4 -->
            <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='dc.png') }}" alt="Delhi
Capitals">
            </div>
            <!-- Image-5 -->
            <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='rr.png') }}"
alt="Rajasthan Royals">
            </div>
            <!-- Image-6 -->
            <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='kkr.jpg') }}"
alt="Kolkata Knight Riders">
            </div>
            <!-- Image-7 -->
            <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='kxip.png') }}" alt="Kings
XI Punjab">
            </div>
            <!-- Image-8 -->
            <div class="image-container">
                <img class="image-width" src="{{ url_for('static', filename='csk.png') }}"
alt="Chennai Super Kings">
            </div>

        </div>
      </div>

    </div>

    <!-- Result -->
    <div class="results">
      <p class='result-p'>The final predicted score (range): {{ lower_limit }} to {{ upper_limit }}</p>
    </div><!-- IPL Banner - 2 -->
 <div class='ipl-banner-right'>
 <div class="slider-frame">
            <div class="slide-images-down">
```

24

```
<!-- Image-1 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='mi.jpg') }}" alt="Mumbai
Indians">
            </div>
            <!-- Image-2 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='srh.png') }}"
alt="Sunrisers Hyderabad">
            </div>
            <!-- Image-3 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='rcb.png') }}" alt="Royal
Challengers Bangalore">
            </div>
            <!-- Image-4 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='dc.png') }}" alt="Delhi
Capitals">
            </div>
            <!-- Image-5 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='kkr.jpg') }}"
alt="Kolkata Knight Riders">
            </div>
            <!-- Image-6 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='rr.png') }}"
alt="Rajasthan Royals">
            </div>
            <!-- Image-7 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='kxip.png') }}" alt="Kings
XI Punjab">
            </div>
            <!-- Image-8 -->
            <div class="image-container">
              <img class="image-width" src="{{ url_for('static', filename='csk.png') }}"
alt="Chennai Super Kings">
          </div>
         </div>
        </div></div>
    </div>
```

25

```
!-- Footer -->
  <!--
  <div class='footer'>
    <div class="contact">
            </div>  <--!>
  <p class='footer-description'><h1> <b> <center> Made by Anugrah George </b>
</center><h1></p></body>
</html>
```

CHAPTER 8:- TESTING STRATEGIES

What is the goal of ML testing?

First of all, what are we trying to achieve when performing ML testing, as well as any software testing whatsoever?

Quality assurance is required to make sure that the software system works according to the requirements. Were all the features implemented as agreed? Does the program behave as expected? All the parameters that you test the program against should be stated in the technical specification document.
Moreover, software testing has the power to point out all the defects and flaws during development. You don't want your clients to encounter bugs after the software is released and come to you waving their fists. Different kinds of testing allow us to catch bugs that are visible only during runtime.
However, in machine learning, a programmer usually inputs the data and the desired behavior, and the logic is elaborated by the machine. This is especially true for deep learning. Therefore, the purpose of machine learning testing is, first of all, to ensure that this learned logic will remain consistent, no matter how many times we call the program.

Model evaluation in machine learning testing:

Usually, software testing includes:

Unit tests. The program is broken down into blocks, and each element (unit) is tested separately.
Regression tests. They cover already tested software to see if it doesn't suddenly break.
Integration tests. This type of testing observes how multiple components of the program work together.
Moreover, there are certain rules that people follow: don't merge the code before it passes all the tests, always test newly introduced blocks of code, when fixing bugs, write a test that captures the bug.

Machine learning adds up more actions to your to-do list. You still need to follow ML's best practices. Moreover, every ML model needs not only to be tested but evaluated. Your model should generalize well. This is not what we usually understand by testing, but evaluation is needed to make sure that the performance is satisfactory.

Machine learning Models Used for Testing:-

Supervised learning:

Supervised learning is the simplest of the learning models to understand. Learning in the supervised model entails creating a function that can be trained by using a training data set, then applied to unseen data to meet some predictive performance. The goal is to build the function so that it generalizes well over data it has never seen.

You build and test a mapping function with supervised learning in two phases (see image below). In the first phase, you segment a data set into two types of samples: training data and test data. Both training and test data contain a test vector (the inputs) and one or more known desired output values. You train the mapping function with the training data set until it meets some level of performance (a metric for how accurately the mapping function maps the training data to the associated desired output). In the context of supervised learning, this occurs with each training sample, where you use the error (actual vs. desired output) to alter the mapping function. In the next phase, you test the trained mapping function against the test data. The test data represents data that has not been used for training and provides a good measure for how well the mapping function generalizes to unseen data.

Supervised learning classified into two categories of algorithms:

Classification: A classification problem is when the output variable is a category, such as "Red" or "blue" or "disease" and "no disease".

Regression: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

Supervised learning deals with or learns with "labeled" data. This implies that some data is already tagged with the correct answer.

Types:-

Regression
Logistic Regression
Classification
Naive Bayes Classifiers
K-NN (k nearest neighbors)
Decision Trees
Advantages:-

Supervised learning allows collecting data and produces data output from previous experiences.Helps to optimize performance criteria with the help of experience.

28

Supervised machine learning helps to solve various types of real-world computation problems.

Disadvantages:-Classifying big data can be challenging.Training for supervised learning needs a lot of computation time. So, it requires a lot of time.

Problems and Issues in Supervised learning: Before we get started, we must know about how to pick a good machine learning algorithm for the given dataset. To intelligently pick an algorithm to use for a supervised learning task, we must consider the following factors [4]:

1. **Heterogeneity of Data:** Many algorithms like neural networks and support vector machines like their feature vectors to be homogeneous numeric and normalized. The algorithms that employ distance metrics are very sensitive to this, and hence if the data is heterogeneous, these methods should be the afterthought. Decision Trees can handle heterogeneous data very easily.

2. **Redundancy of Data:** If the data contains redundant information, i.e. contain highly correlated values, then it's useless to use distance based methods because of numerical instability. In this case, some sort of Regularization can be employed to the data to prevent this situation.

3. **Dependent Features:** If there is some dependence between the feature vectors, then algorithms that monitor complex interactions like Neural Networks and Decision Trees fare better than other algorithms.

4. **Bias-Variance Tradeoff:** A learning algorithm is biased for a particular input x if, when trained on each of these data sets, it is systematically incorrect when predicting the correct output for x, whereas a learning algorithm has high variance for a particular input x if it predicts different output values when trained on different training sets. The prediction error of a learned classifier can be related to the sum of bias and variance of the learning algorithm, and neither can be high as they will make the prediction error to be high. A key feature of machine learning algorithms is that they are able to tune the balance between bias and variance automatically, or by manual tuning using bias parameters, and using such algorithms will resolve this situation.

5. **Curse of Dimensionality:** If the problem has an input space that has a large number of dimensions, and the problem only depends on a subspace of the input space with small dimensions, the machine learning algorithm can be confused by the huge number of dimensions and hence the variance of the algorithm can be high. In practice, if the data scientist can manually remove irrelevant features from the input data, this is likely to improve the accuracy of the learned function. In addition, there are many algorithms for feature selection that seek to identify the relevant features and discard the irrelevant ones, for instance Principle Component Analysis for unsupervised learning. This reduces the dimensionality.

6. **Overfitting:** The programmer should know that there is a possibility that the output values may constitute of an inherent noise which is the result of human or sensor errors. In thiscase, the algorithm must not attempt to infer the function that exactly matches all the data. Being too careful in fitting the data can cause overfitting, after which the model will answer perfectly for all training examples but will have a very high error for unseen samples. If the input space of the dataset we were working on had 1000 dimensions, then it's better to first perform PCA on the data before using a supervised learning algorithm on it.

29

Unsupervised learning:

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself. For instance, suppose it is given an image having both dogs and cats which it has never seen.

Thus the machine has no idea about the features of dogs and cats so we can't categorize it as 'dogs and cats '. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having dogs in it and the second part may contain all pics having cats in it. Here you didn't learn anything before, which means no training data or examples.

It allows the model to work on its own to discover patterns and information that was previously undetected. It mainly deals with unlabelled data.

Unsupervised learning is classified into two categories of algorithms:

Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

**Types of Unsupervised Learning:-**

Clustering

- Exclusive (partitioning)
- Agglomerative
- Overlapping
- Probabilistic

Clustering Types:-

- Hierarchical clustering
- K-means clustering
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysis

Flow Chart:

Flow Chart of the implementation of machine learning for predicting suitable solvers.
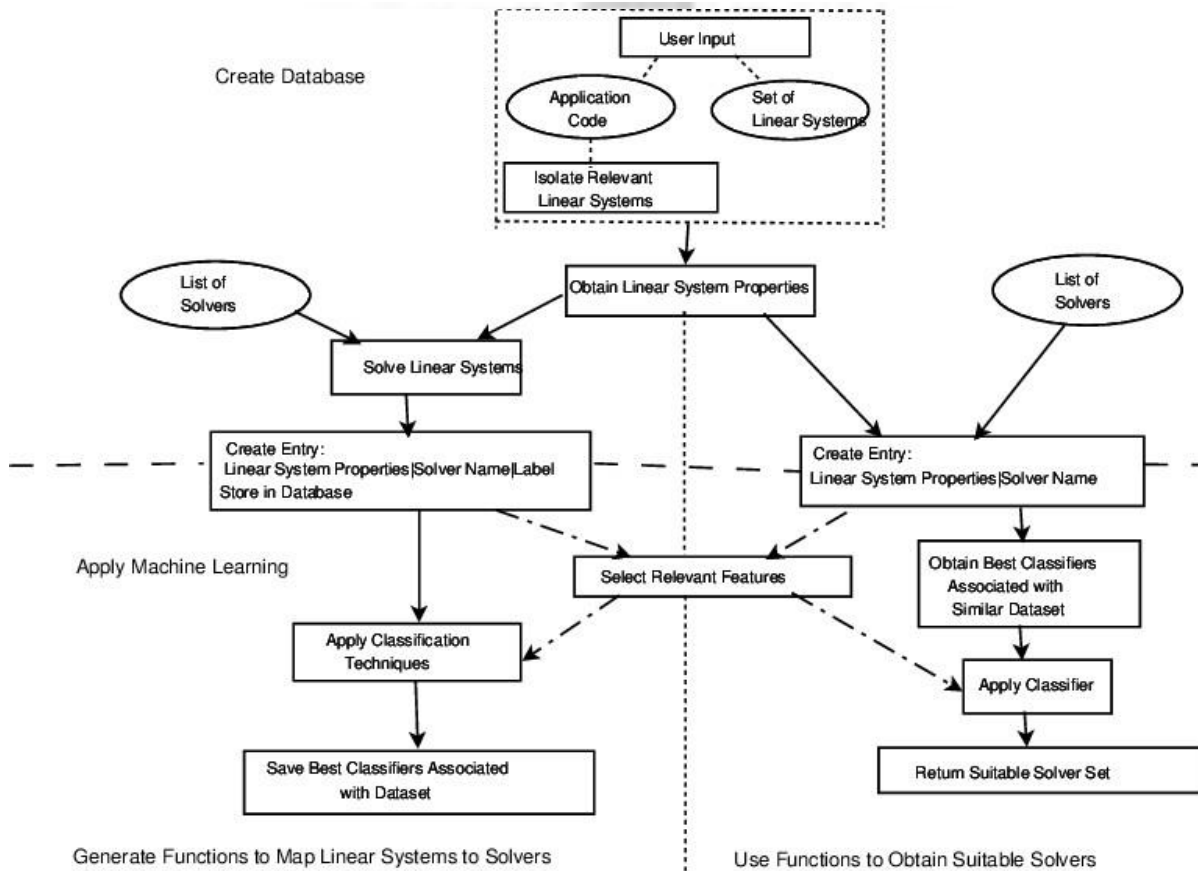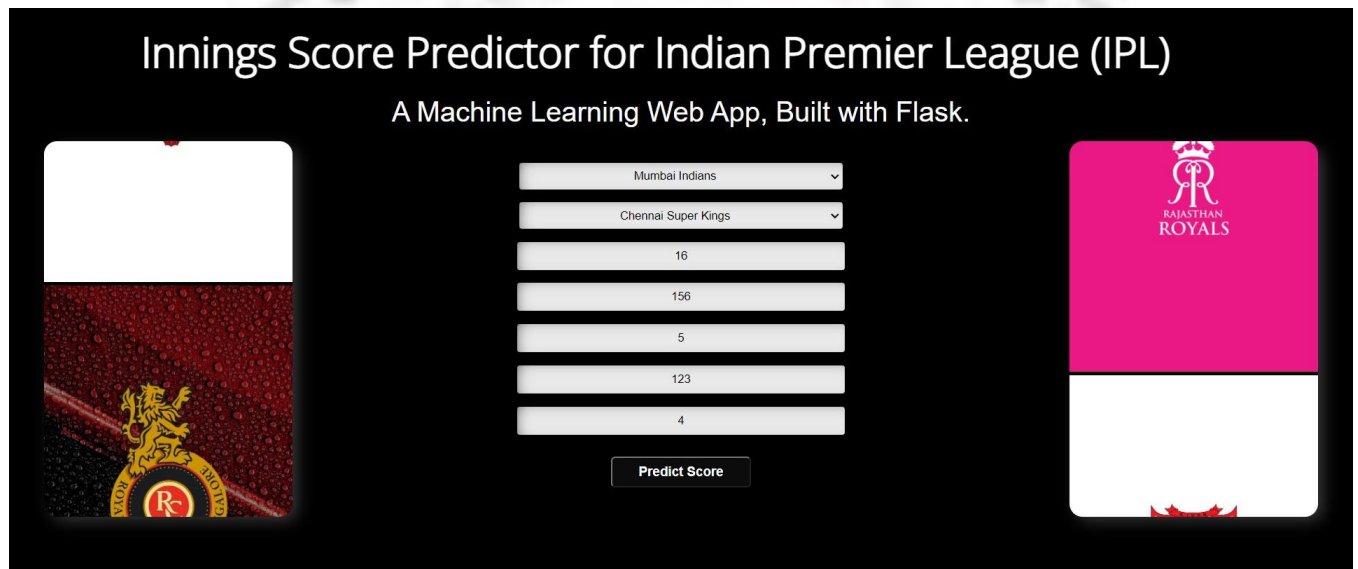
CHAPTER 9:- LIMITATIONS

1. No database is connected so data is temporary.

2. Machine must have flask and python libraies installed to run this model.

3. It is not hosted so you have to run the code again and again go to the interface.

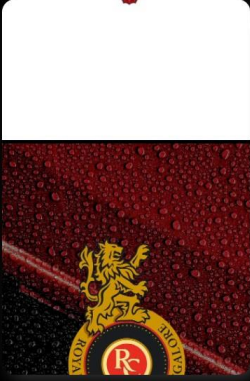4 .It is the project on the ground level more work is being done on the project to make it

even better.

CHAPTER 11:-CONCLUSION

So at the end we want to say that this project contains good knowledge of machine learning techniques and it can tell any student his/her job possibility with the help of web api it looks very attractive when it comes to use this model.Here are some pictorial represtation of this application-

Innings Score Predictor for Indian Premier League (IPL)

A Machine Learning Web App, Built with Flask.

The final predicted score (range): 272 to 287

34

REFERENCES

[1] "Intro to Machine Learning | Udacity." Intro to Machine Learning | Udacity. Accessed April 27, 2016. https://www.udacity.com/course/intro-to-machine-learning--ud120.

[2] "Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition. Mining, Inference, and Prediction. 2nd Edition. Accessed April 27, 2016. http://statweb.stanford.edu/~tibs/ElemStatLearn/.

[3] "No Free Lunch Theorems." No Free Lunch Theorems. Accessed April 27, 2016. http://www.no-free-lunch.org/.

[4] Hastie, Trevor, Robert Tibshirani, and J. H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-color Illustrations. New York: Springer, 2001.

[5] "Weka 3: Data Mining Software in Java." Weka 3. Accessed April 27, 2016. http://www.cs.waikato.ac.nz/ml/weka/.

[6] Bozhinova, Monika, Nikola Guid, and Damjan Strnad. Naivni Bayesov Klasifikator: Diplomsko Delo. Maribor: M. Bozhinova, 2015.

[7] Schölkopf, Bernhard, Christopher J. C. Burges, and Alexander J. Smola. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 1999.

[8] Norving, Peter, and Stuart Russel. Artificial Intelligence: A Modern Approach. S.l.: Pearson Education Limited, 2013.

[9] Witten, I. H., and Eibe Frank. Data Mining: Practical Machine Learning Tools andTechniques. Amsterdam: Morgan Kaufman, 2005.