

## **Internship Report On**

### ***“Artificial Intelligence and Machine Learning”***

A Dissertation submitted in partial fulfillment of the requirement  
for the award of degree of

**MASTER OF COMPUTER APPLICATIONS**  
of  
**Visvesvaraya Technological University, Belagavi**



By

**BHUDEVI VATTIKUTI**  
**1RN19MCA08**

**Carried out at**  
**New Age Solutions Technologies (NASTECH)**

Under the Guidance of

**Internal guide:**  
**DR.N.P Kavya**  
**Professor & Head**  
**Dept. of MCA**

**External Guide:**  
**Azib Hasan**  
**Subject Matter Expert**  
**Nastech**



*ESTD:2001*  
*An Institute with a Difference*

**Department of Master of Computer Applications**  
**RNS Institute of Technology**  
**Dr. Vishnuvardhan Road, Channasandra,**  
**Bengaluru-560 098**  
**APRIL 2022**

# ***“Artificial Intelligence and Machine Learning”***

A Dissertation submitted in partial fulfillment of the requirement  
for the award of degree of

**MASTER OF COMPUTER APPLICATIONS**  
of  
**Visvesvaraya Technological University, Belagavi**



By

**BHUDEVI VATTIKUTI**  
**1RN19MCA08**

**Carried out at**  
**New Age Solutions Technologies (NASTECH)**

Under the Guidance of

**Internal guide:**  
**DR. N P Kavya**  
**Professor & Head**  
**Dept. of MCA**

**External Guide:**  
**Azib Hasan**  
**Subject Matter Expert**  
**Nastech**



*ESTD:2001*  
*An Institute with a Difference*

**Department of Master of Computer Applications**  
**RNS Institute of Technology**  
**Dr. Vishnuvardhan Road, Channasandra, Bengaluru-560 098**  
**APRIL 2022**



ESTD:2001  
An Institute with a Difference

## Department of Master of Computer Applications

### **RNS Institute of Technology**

**Dr. Vishnuvardhan Road, Channasandra, Bengaluru-560 098**

## **CERTIFICATE**

*This is to certify that Ms. Bhudevi Vattikuti, student of 6<sup>th</sup> semester MCA, bearing the USN: 1RN19MCA08 has completed his/her final semester internship/project work entitled "Artificial Intelligence and Machine Learning" as a partial fulfillment for the award of Master of Computer Applications degree, during the academic year 2022 under our joint supervision.*

#### **Internal Guide**

**Dr. N P Kavya**  
Professor & Head  
Department of MCA  
RNS Institute of Technology  
Bengaluru - 98

#### **External Guide**

**Mr. Azib Hasan**  
Subject Matter Expert  
New Age Solutions  
Technologies  
Thane (W) - 400608  
Mumbai, Maharashtra

#### **Head of the Department**

**Dr. N P Kavya**  
Professor & Head  
Department of MCA  
RNS Institute of Technology  
Bengaluru - 98

#### **Principal**

**Dr. M K Venkatesha**  
Principal  
RNS Institute of Technology  
Bengaluru - 98

## DECLARATION

I, **Ms. Bhudevi Vattikuti**, student of 6<sup>th</sup> MCA, RNS Institute of Technology, bearing USN: **1RN19MCA08** hereby declare that the project entitled “ **Artificial Intelligence and Machine Learning**” has been carried out by me under the supervision of External Guide **Mr. Azib Hasan**, Subject Matter expert, and Internal Guide **Dr. N P Kavya**, Professor & HoD and submitted in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications by the Visvesvaraya Technological University during the academic year 2022. This report has not been submitted to any other Organization / University for any award of degree or Certificate.

Signature

**Bhudevi Vattikuti**

## ACKNOWLEDGEMENT

The successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant support and encouragement has crowned my efforts.

I take this opportunity to acknowledge the guidance I have received from different individuals and place on record my appreciation and thanks.

I express my sincere gratitude to **Dr. R N Shetty**, Founder and **Sri. Satish R Shetty**, Managing Director, RNSIT for providing us wonderful academic environment.

My deep sense of gratitude to our Principal **Dr. M K Venkatesha**, for his kind support.

I am grateful to **Dr. N P Kavya**, Head of the Department of MCA, RNSIT for nurturing our technical skills and contributing towards the success of this project.

I would also express my heartfelt thanks to my internal guide **Dr. N P Kavya**, Professor & HoD, Department of MCA, RNSIT for her continuous guidance and valuable suggestions for this internship work.

It's my pleasure to thank **Nastech** for providing me the best platform to complete the internship work and glad to thank the external guide **Mr. Azib Hasan**, Subject Matter expert.

I also express my heartfelt thanks to all the teaching and non-teaching staff members of MCA Department for their encouragement and support throughout this work.

**BHUDEVI VATTIKUTI**

**1RN19MCA08**

## ABSTRACT

Artificial Intelligence is a technique for building systems that mimic human behavior or decision-making. Machine Learning is a subset of AI that uses data to solve tasks. These solvers are trained models of data that learn based on the information provided to them. AI is used in different domains to give insights into user behavior and give recommendations based on the data.

Machine Learning is used anywhere from automating mundane tasks to offering intelligent insights, industries in every sector try to benefit from it. You may already be using a device that utilizes it. For example, a wearable fitness tracker like Fitbit, or an intelligent home assistant like Google Home. But there are much more examples of ML in use.

Prediction: Machine learning can also be used in the prediction systems. Image recognition: Machine learning can be used for face detection in an image as well. There is a separate category for each person in a database of several people, Speech Recognition: It is the translation of spoken words into the text. It is used in voice searches and more. Voice user interfaces include voice dialing, call routing, and appliance control. It can also be used a simple data entry and the preparation of structured documents, Medical diagnoses: ML is trained to recognize cancerous tissues, and other harmful cells. Financial industry and trading: companies use ML in fraud investigations and credit checks.

# TABLE OF CONTENTS

	Page No.
<b>DECLARATION</b>	<b>i</b>
<b>ACKNOWLEDGEMENT</b>	<b>ii</b>
<b>ABSTRACT</b>	<b>iii</b>
<b>TABLE OF CONTENTS</b>	<b>iv</b>
<b>LIST OF FIGURES</b>	<b>v</b>
<b>LIST OF TABLES</b>	<b>vi</b>
<b>1. INTRODUCTION</b>	<b>1</b>
1.1. Aim	<b>1</b>
1.2. Project description	<b>1</b>
1.3. Scope	<b>3</b>
<b>2. COMPANY PROFILE</b>	<b>4</b>
2.1. Organization structure	<b>4</b>
2.2. Different departments and functions	<b>4</b>
2.3. Job process / Services / Facilities	<b>4</b>
<b>3. TOOLS AND TECHNOLOGY</b>	<b>5</b>
3.1. Tools/technology used by company	<b>5</b>
3.2. Tools learned in training	<b>6</b>
<b>4. INTERNSHIP WORK</b>	<b>7</b>
4.1. Task assigned	<b>7</b>
4.2. Application developed using modern tools	<b>10</b>
4.3. Professional learning (Discipline, attitude, planning, groupwork, self-assessment, etc)	<b>10</b>
<b>5. IMPLEMENTATION</b>	<b>12</b>
5.1. Screen shots	<b>12</b>
<b>6. SOFTWARE TESTING</b>	<b>18</b>
6.1. Sorts of investigations / Test cases	<b>18</b>
<b>7. CONCLUSION AND FUTURE WORK</b>	<b>20</b>
<b>REFERENCES</b>	

## List of Figures

<b>Figure No.</b>	<b>Figure Caption</b>	<b>Page No.</b>
5.1	BMI dataset analysis using k-nearest neighbours (KNN) Algorithm.	12
5.2	Forecasting using FbProphet.	13
5.3	Sentiment Analysis using tweets.	14
5.4	Implementation of different computer vision techniques.	15
5.5	Face and Eye detection.	16
5.6	Mall customers dataset analysis using K-means Clustering Algorithm.	17



## List of Tables

<b>Table No.</b>	<b>Table Caption</b>	<b>Page No.</b>
4.1.1	Top 5 rows of titanic dataset.	7
4.1.2	Top 5 rows of BMI dataset.	8
4.1.6	Top 5 rows of Mall Customers dataset.	9

## **Chapter – 1**

### **INTRODUCTION**

#### **1.1 Aim**

To provide the ability to automatically obtain deep insights, recognize unknown patterns, and create high performing predictive models from data, all without requiring explicit programming instructions.

#### **1.2 Description**

Machine Learning focuses on the development of computer programs that can access data and use it to learn 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.

##### **1.2.1 Machine Learning**

Machine Learning is the science of getting computers to learn without being explicitly programmed. It is closely related to computational statistics, which focuses on making prediction using computer. In its application across business problems, machine learning is also referred as predictive analysis.

Machine Learning is closely related to computational statistics. 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.

- **Supervised Learning**

Supervised Learning is a type of learning in which we are given a data set and we already know what a correct output should look like, having the idea that there is a relationship between the input and output. Basically, it is learning task of learning a function that maps an input to an output based on example input-output pairs.

- **Unsupervised Learning**

It is a type of learning that allows us to approach problems with little or no idea what our problem should look like. We can derive the structure by clustering the data based on a relationship among the variables in data. With unsupervised learning there is no feedback based on prediction result.

- **Semi-Supervised Learning:**

In semi supervised learning, it contains the combination of both supervised and unsupervised learning. Model is trained with labelled datasets but in this case, algorithm is flexible to choose their own decisions and predictions will be based on its own understandings.

- **Reinforcement learning**

It 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 behavior within a specific context in order to maximize its performance.

### **1.2.2 Computer Vision**

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.[1]

### 1.2.3 Natural language Processing

Natural language processing strives to build machines that understand and respond to text or voice data and respond with text or speech of their own in much the same way humans do.

NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

#### NLP tasks

- Speech recognition also called speech-to-text, is the task of reliably converting voice data into text data. Part of speech tagging also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context.
- Word sense disambiguation is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context.
- Named entity recognition or NEM, identifies words or phrases as useful entities. NEM identifies ‘Kentucky’ as a location or ‘Fred’ as a man's name.
- Sentiment analysis attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion, suspicion—from text.[2]

### 1.3 Scope

Some of the activities computers with artificial intelligence are designed for, include Learning, Planning, Problem solving, Speech/Face recognition. Artificial Intelligence has immense potential to change each sector of the economy for the benefit of society. There is not just one technology under AI, but there are various useful technologies such as self-improving algorithms, machine learning, big data, pattern recognition.

## Chapter – 2

### COMPANY PROFILE

#### 2.1. Organization structure

Nastech is formed with the purpose of bridging the gap between Academia and Industry. Nastech is one of the leading Global Certification and Training service providers for technical and management programs for educational institutions. We collaborate with educational institutes to understand their requirements and form a strategy in consultation with all stakeholders to fulfill those by skilling, reskilling and upskilling the students and faculties on new age skills and technologies.

#### 2.2. Different departments and functions

- We offer industry and project-oriented training programs which not only expose students to hands-on training experience but also make them practical oriented towards the industry-readiness expected in today's time.
- We take pride that all our programs are mapped to a certain Global Certification Exams i.e., after the students are done with their training, they will prove themselves on a global level via a global certification exam.
- We lead from the front in terms of costing of our overall global certification and training programs.

#### 2.3. Services

- Industry and project-oriented student training programs.
- Certification programs mapped to Global Certification Exams from Microsoft/EC Council/Google/AWS/ Adobe).
- Placement training for pre-final and final year students, LMS and Online assessment solutions for future ready campuses.

## Chapter – 3

# TOOLS AND TECHNOLOGY

### 3.1 Tools/technology used by company

- **Google Colaboratory**

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser. With Colab you can import an image dataset, train an image classifier on it, and evaluate the model, all in just a few lines of code. Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including GPUs and TPUs.

- **GitHub**

GitHub is an online software development platform used for storing, tracking, and collaborating on software projects. It enables developers to upload their own code files and to collaborate with fellow developers on open-source projects. GitHub also serves as a social networking site in which developers can openly network, collaborate, and pitch their work.

- **Kaggle**

Kaggle is platform to compete with others in competitions which are based on machine learning tasks. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

- **Azure**

Microsoft Azure is a cloud computing service that offers a range of software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) option for deploying applications and services on Microsoft-managed data center infrastructure.

### 3.2. Tools learned in training.

- Fundamentals of Python and data structures.
- Libraries and dependencies.
- NumPy: NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. The array object in NumPy is called ndarray.
- Pandas: Pandas makes it simple to do many of the time consuming, repetitive tasks associated with working with data, including Data cleansing, Data fill, Data normalization, Merges and joins, Data visualization, Statistical analysis, Data inspection, Loading and saving data,
- Matplotlib: Visualization with Python Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- Scikit-learn (Sklearn): It is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistency interface in Python.
- Popular Machine Learning algorithms such as Linear regression, Logistic regression, Decision tree, KNN (K-Nearest Neighbors) algorithm, K-means Clustering.
- Computer vision libraries such as OpenCV, tesseract-ocr and its different features to create and deploy data science solutions.
- Natural Language Processing techniques and performing sentimental analysis Twitter API.
- Web scraping using BeautifulSoup a python package for parsing HTML and XML documents.
- Using GitHub as a Version control software to keep track of every modification to the code.
- Using Heroku as a container-based cloud Platform as a service (PaaS) to deploy, manage and scale modern apps.

## Chapter – 4

### INTERNSHIP WORK

#### 4.1 Task assigned:

The tasks are assigned to achieve the results of various experiments using different methodologies.

##### 4.1.1 Exploratory Data Analysis on dataset

The philosophy behind this approach was to examine the data before building a model. Exploratory Data Analysis or EDA is used to take insights from the data. Data Scientists and Analysts try to find different patterns, relations, and anomalies in the data using some statistical graphs and other visualization techniques.

- Import libraries and load dataset Uncover underlying structure.
- Check for missing values.
- Visualizing the missing values.
- Replacing the missing values.
- Asking Analytical Questions and Visualizations.

**Table 4.1.1: Top 5 rows of titanic dataset**

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S



### 4.1.2 KNN Algorithm

This is a supervised learning algorithm that considers different centroids and uses a usually Euclidean function to compare distance. Then, it analyses the results and classifies each point to the group to optimize it to place with all closest points to it. It classifies new cases using a majority vote of  $k$  of its neighbours. The case it assigns to a class is the one most common among its  $K$  nearest neighbours.

KNN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification

**Table 4.1.2: Top 5 rows of BMI dataset**

	Weight(x2)	Height(y2)	Class
0	51	167	Underweight
1	66	177	Normal
2	75	169	Overweight
3	69	176	Normal
4	50	173	Underweight

### 4.1.3 Forecasting using FbProphet

Forecasting methods are implemented to make decisions regarding buying, provisioning, replenishment, and financial planning. Prophet is open source software released by Facebook's Core Data Science team. The library is so powerful that it has the capability of handling stationarity within the data and also seasonality related components. By stationarity, we mean that there should be constant mean, variance, and covariance in the data if we divide the data into segments with respect to time and seasonality means the same type of trend the data is following if segregated based on time intervals.[3]

### 4.1.4 Computer Vision Techniques.

Image clarification comprises of a variety of challenges, including viewpoint variation, scale variation, intra-class variation, image deformation, image occlusion, illumination conditions, and background clutter. Computer vision researchers have come up with a data-driven approach to classify images into distinct categories. They first accumulate a training dataset of labelled images and computer process the data.

#### 4.1.5 Sentiment Analysis using tweets.

Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in the text source. Tweets are often useful in generating a vast amount of sentiment data upon analysis. These data are useful in understanding the opinion of the people about a variety of topics.

The dataset provided is the Sentiment140 Dataset which consists of number of tweets that have been extracted using the Twitter API. The various columns present in the dataset are:

- Target: the polarity of the tweet (positive or negative).
- Ids: Unique id of the tweet.
- Date: the date of the tweet.
- Flag: It refers to the query. If no such query exists then it is NO QUERY.
- User: It refers to the name of the user that tweeted.
- Text: It refers to the text of the tweet.

#### 4.1.6 K-Means Algorithm

K-Means is an unsupervised algorithm that solves the problem of clustering. It classifies data using a number of clusters. The data points inside a class are homogeneous and heterogeneous to peer groups. k -means clustering is a method of vector quantization, originally from signal processing.

K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. K-means clustering is rather easy to apply to even large data sets.

**Table 4.1.6: Top 5 rows of Mall Customers dataset.**

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

#### **4.1.7 Spam or Ham classification using Natural Language Processing**

Classifying spam and ham messages is one of the most common natural language processing tasks for emails and chat engines. With the advancements in machine learning and natural language processing techniques, it is now possible to separate spam messages from ham messages with a high degree of accuracy. Where the task was to detect the specified word is spam or ham using the created model.

#### **4.1.8 Optical Character Recognition using Tesseract**

Tesseract is an open-source optical character recognition (OCR) platform. OCR extracts text from images and documents without a text layer and outputs the document into a new searchable text file, PDF, or most other popular formats. Tesseract is highly customizable and can operate using most languages, including multilingual documents and vertical text [4].

### **4.2 Application developed using modern tools:**

Some of the application developed and deployed using modern tools are:

- Data analysis on different datasets using various algorithms to recognize the patterns.
- BMI data set analysis using k-nearest neighbors (KNN) Algorithm.
- Forecasting using FbProphet.
- Face and eye recognise using Cascade Classifier technique.
- Sentiment Analysis using tweets by tweepy library.
- Implementation of different computer vision techniques.
- Optical character recognizer using pytesseract technique.

### 4.3 Professional learning

Professional learning involves the skills we need to be successful in our job or career.that the focus is not just on individuals' learning but on professional learning within the context of a cohesive group, that focuses on collective knowledge and growth, and occurs within an ethic of interpersonal caring that permeates the life of the community.

It is a cultural way of working oriented toward and building capacity for continuous and sustainable learning.

Professional learning takes place at several levels: the individual, the workplace, and the organization. Learn more in: Educational Leadership Sustainability: Maintaining Wellness, Coping with Stress, and Preventing Burnout.

- Apply a "Growth" Mindset  
Your current skills. Are they adequate for your present role? Do they align adequately with your team or organization's priorities, mission and vision?
- Focus on Objectives  
Drawing up a clear plan of action will help you to organize your learning time more effectively, strengthen your self-discipline, and boost your motivation. It will also give you a way to measure your progress.
- Manage Obstacles and Distractions  
Identify the obstacles that might make it difficult for you to stick to your learning schedule. Then, go through each obstacle and brainstorm strategies that will help you to overcome it.

## Chapter – 5

### IMPLEMENTATION

#### 5.1 Screen shot

- KNN Algorithm

```

1 plt.figure(figsize=(9,7))
2 plt.scatter(df1['Weight(x2)'],df1['Height(y2)'],color='m',label="Normal",marker="*",s=60)
3 plt.scatter(df2['Weight(x2)'],df2['Height(y2)'],color='k',label="Overweight",marker="*")
4 plt.scatter(df3['Weight(x2)'],df3['Height(y2)'],color='r',label="underweight",marker="*")
5 plt.scatter(pred[0][0],pred[0][1],color='yellow',label='Predicted value',marker="*",s=100)
6 plt.legend(loc="upper left")
7 plt.xticks(np.arange(47,85,2),rotation=60)
8 plt.yticks(np.arange(160,185,2))
9 plt.grid()
10 plt.show()

```

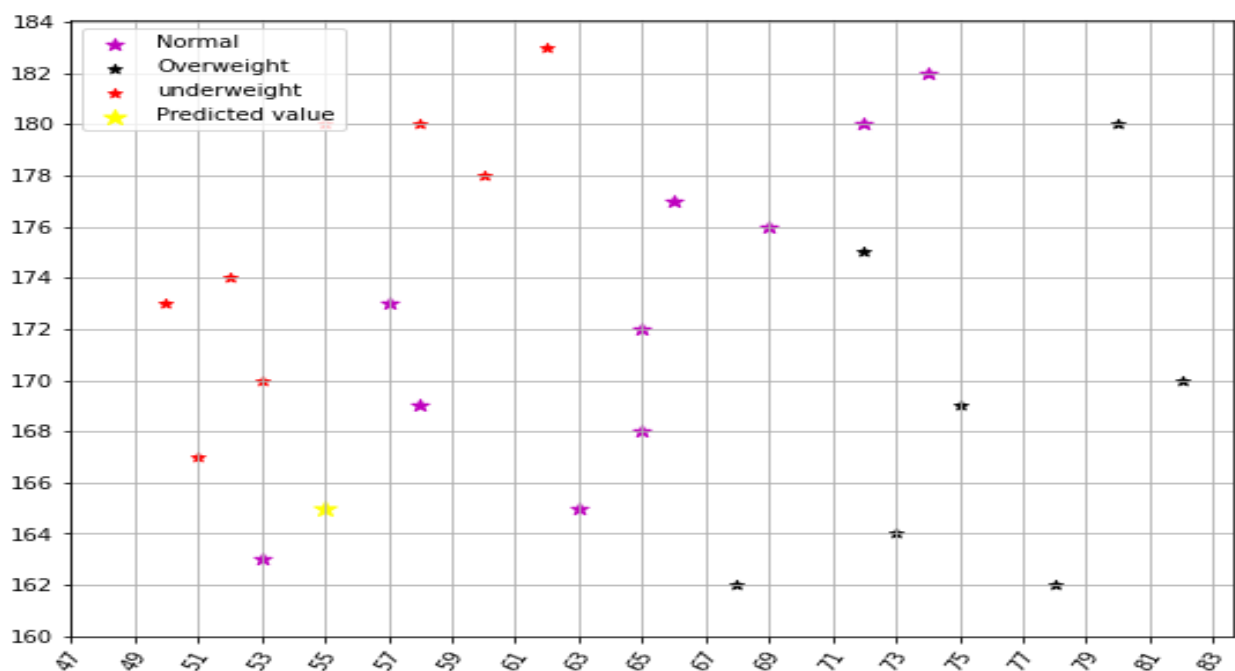


Figure 5.1: BMI dataset analysis using k-nearest neighbors (KNN) Algorithm.

- **Forecasting using FbProphet**

```

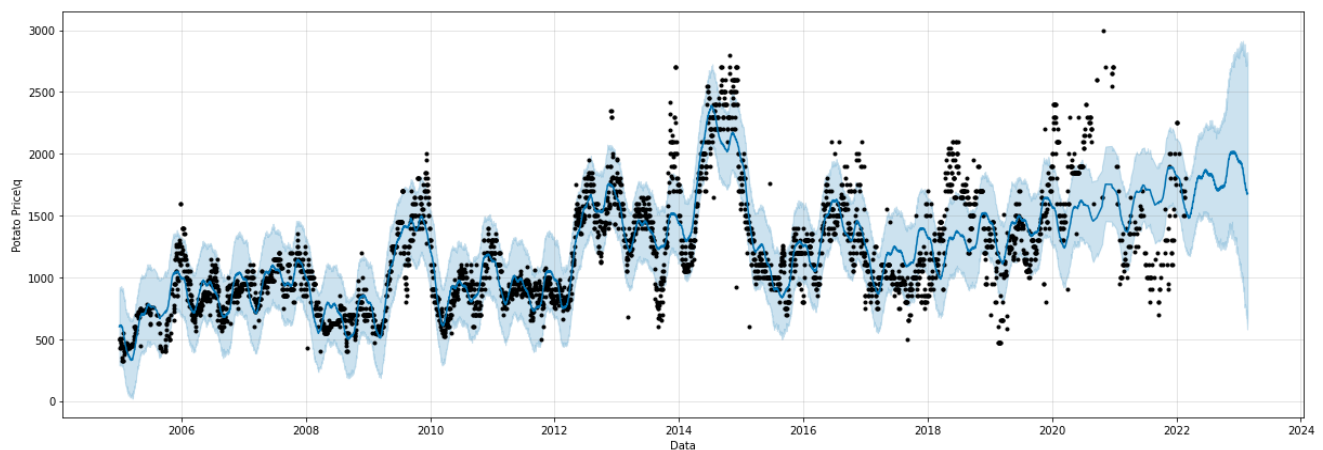
1 import pandas as pd
2 df=pd.read_html("/content/drive/MyDrive/6th sem internship/DailyWiseMarketArrivals.html")[3]
3 df.isnull().sum()
4 df=df.dropna()
5 df.to_csv("Potato.csv",index=False)
6 df.rename(columns={'Modal Price (Rs/q)': 'y', 'Date': 'ds'},inplace=True)
7 df_new=df[['ds', 'y']]
8 df_new['y']=df_new['y'].astype(int)
9 df_new['ds']=pd.to_datetime(df_new['ds'])
10
11 import matplotlib.pyplot as plt
12 plt.figure(figsize=(17,6))
13 plt.scatter(df_new['ds'],df_new['y'])
14 plt.show()
15

```

```

1 import fbprophet
2 op=fbprophet.Prophet(changepoint_prior_scale=1)
3 op.fit(df_new)
4 forecast=op.make_future_dataframe(periods=365,freq="D")
5 forecast=op.predict(forecast)
6 forecast
7 op.plot(forecast,xlabel="Data",ylabel="Potato Price\q",figsize=(17,6))
8 plt.figure(figsize=(17,6))
9 plt.show()

```

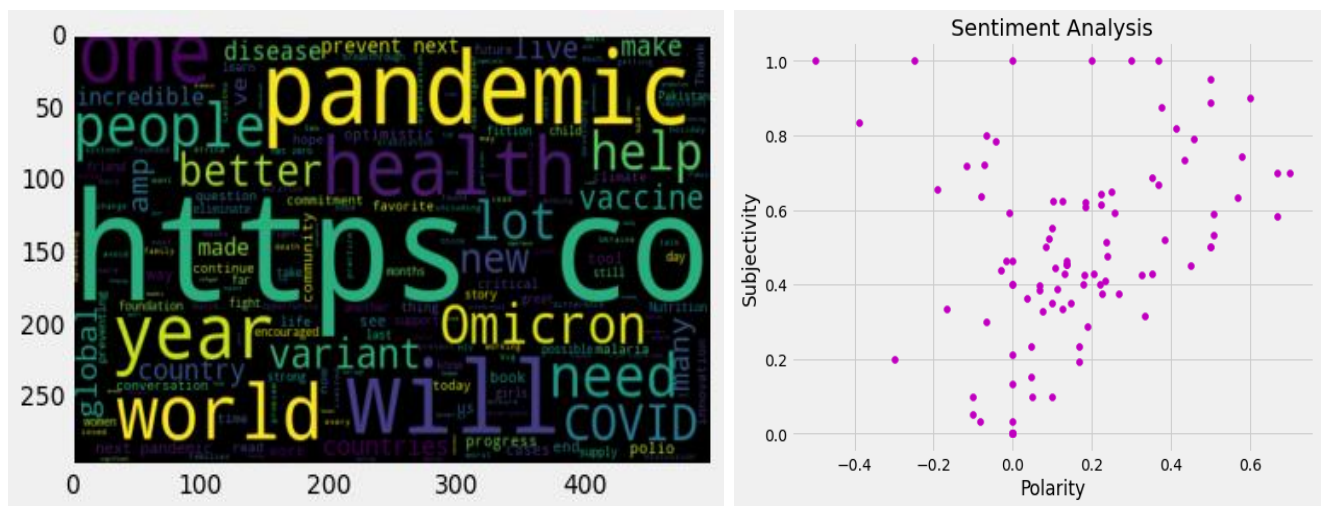


**Figure 5.2: Forecasting using FbProphet.**

```

1 import tweepy
2 from textblob import TextBlob
3 from wordcloud import WordCloud
4 import numpy as np
5 import pandas as pd
6 import re
7 import matplotlib.pyplot as plt
8 plt.style.use("fivethirtyeight")
9 authenticate=tweepy.OAuthHandler(consumerKey,consumerSecret)#create an authentication object
10 authenticate.set_access_token(accessToken,accessTokenSecret)
11 api=tweepy.API(authenticate,wait_on_rate_limit=True)
12 posts=api.user_timeline(screen_name="BillGates",count=100,lang="en",tweet_mode="extended")
13 #clean the text
14 #create a function to clean text
15 def cleanTxt(text):
16     text=re.sub('@[a-zA-Z0-9]+',' ',text)
17     text=re.sub(r'#',' ',text)
18     text=re.sub(r'RT[\s]+',' ',text)
19     text=re.sub(r'https?:\/\/\s+',' ',text)
20     return text
21
22 def getSubjectivity(text):#creat a function to get the subjectivity
23     return TextBlob(text).sentiment.subjectivity
24
25 def getPolarity(text):#create a function to get the polarity
26     return TextBlob(text).sentiment.polarity
27
28 df['Subjectivity']=df['Tweets'].apply(getSubjectivity)#create two columns
29 df['Polarity']=df['Tweets'].apply(getPolarity)
30 df
31 #Plot the WordCloud
32 allWords= ' '.join([twts for twts in df['Tweets']])
33 wordCloud=WordCloud(width=500,height=300,random_state=21,
34                     max_font_size=119).generate(allWords)
35 plt.imshow(wordCloud,interpolation="bilinear")
36 plt.grid()
37 plt.show()

```



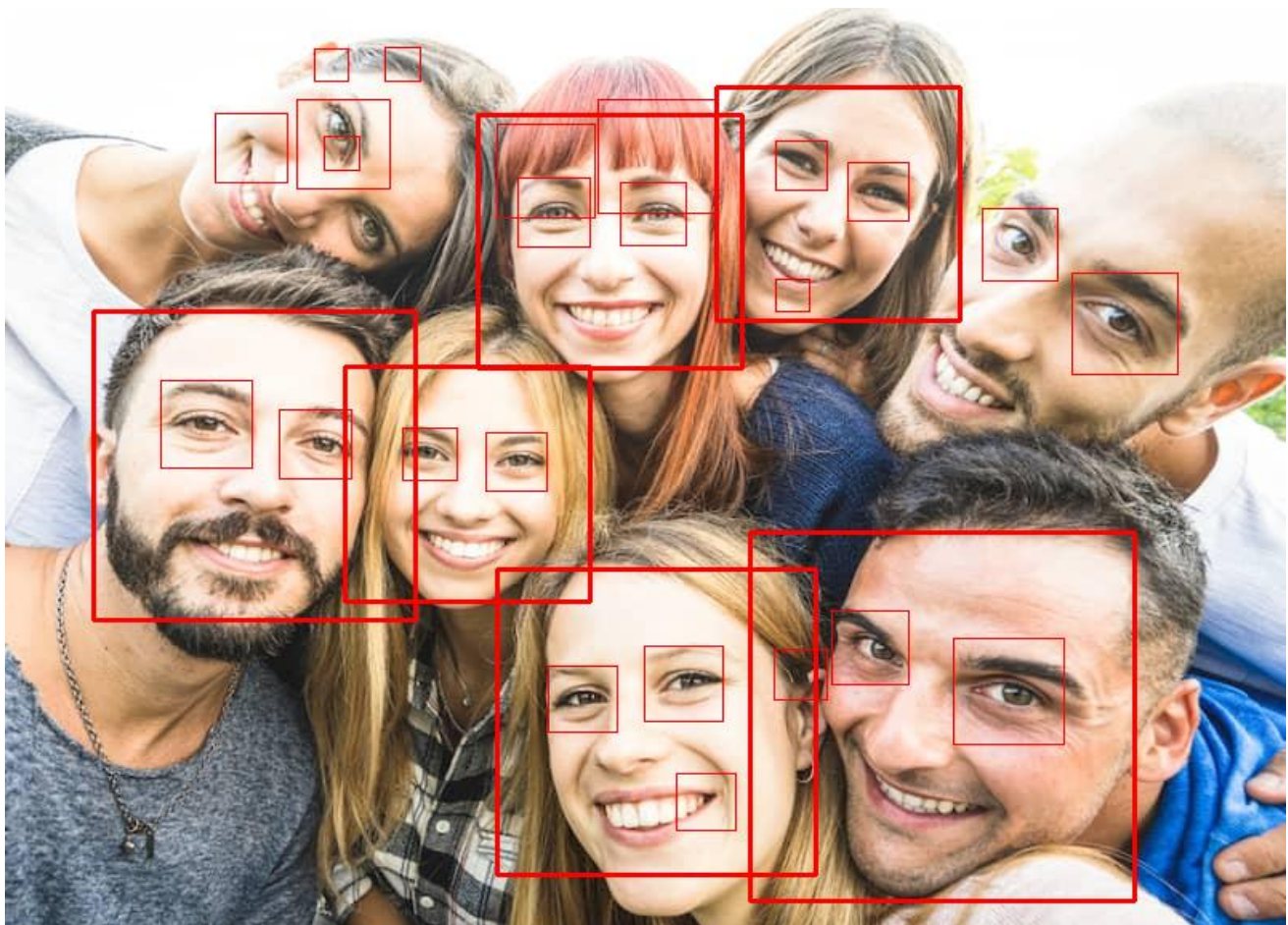
**Figure 5.3: Sentiment Analysis using tweets.**





- Face and Eye detection

```
1 import cv2
2 from google.colab.patches import cv2_imshow
3 !wget https://www.quickanddirtytips.com/sites/default/files/images/2332/people-persons-peoples.jpg -O p1.jpg
4 img=cv2.imread("/content/p1.jpg")
5 img=cv2.imread("/content/p1.jpg")
6 mode=cv2.CascadeClassifier("/content/haarcascade_frontalface_default.xml")
7 face=mode.detectMultiScale(img,1.1,2)
8 for(x,y,w,h) in face:
9     cv2.rectangle(img,(x,y),(x+w,y+h),(0,0,255),2)
10    model1=cv2.CascadeClassifier("/content/haarcascade_eye.xml")
11    eye=model1.detectMultiScale(img,1.1,2)
12    for(x,y,w,h) in eye:
13        cv2.rectangle(img,(x,y),(x+w,y+h),(0,0,255),1)
14    cv2_imshow(img)
```



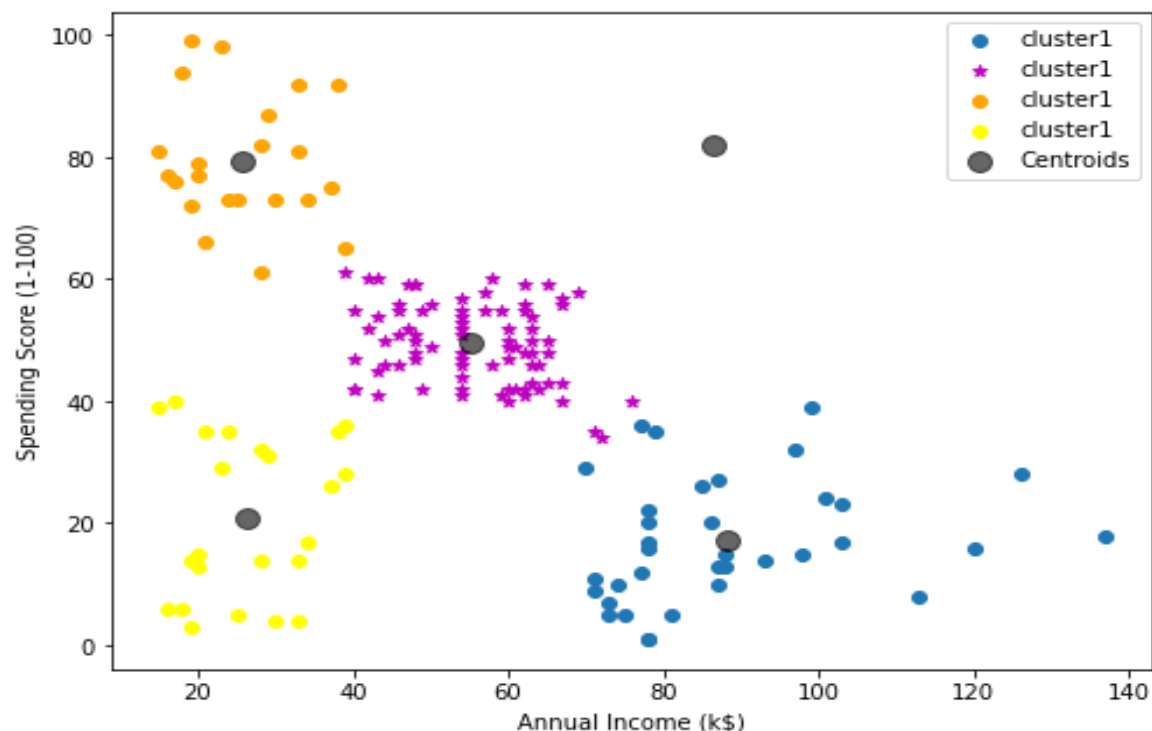
**Figure 5.5: Face and Eye detection.**

## • K-means Clustering Algorithm

```

1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 df=pd.read_csv("/content/drive/MyDrive/6th sem internship/Mall_Customers.csv")
5 x=df.iloc[:,3:5].values
6 x[y==0,0] #annual income value for elements in 0th cluster
7 x[y==0,1] #Spending score value for elements in 0th cluster
8 x[y==1,0] #annual income value for elements in 1st cluster
9 x[y==1,1] #Spending score value for elements in 1st cluster
10 plt.figure(figsize=(8,6))
11 plt.scatter(x[y==0,0],x[y==0,1],label='cluster1')
12 plt.scatter(x[y==1,0],x[y==1,1],color='m',label='cluster1',marker='*')
13 plt.scatter(x[y==3,0],x[y==3,1],color='orange',label='cluster1')
14 plt.scatter(x[y==4,0],x[y==4,1],color='yellow',label='cluster1')
15 plt.scatter(model.cluster_centers_[0,0],model.cluster_centers_[0,1],
16             alpha=0.6,s=100,color="k",label="Centroids")
17 plt.xlabel("Annual Income (k$)")
18 plt.ylabel("Spending Score (1-100)")
19 plt.legend()
20 plt.show()

```



**Figure 5.6: Mall customers dataset analysis using K-means Clustering Algorithm.**

## Chapter – 6

# SOFTWARE TESTING

### 6.1 Test cases

Testing forms an integral part of any software development project. Testing helps in ensuring that the final product is by and large, free of defects and it meets the desired requirements. Proper testing in the development phase helps in identifying the critical errors in the design and implementation of various functionalities thereby ensuring product reliability. Even though it is a bit time-consuming and a costly process at first, it helps in the long run of software development.

Although machine learning systems are not traditional software systems, not testing them properly for their intended purposes can lead to a huge impact in the real world. This is because machine learning systems reflect the biases of the real world. Not accounting or testing for them will inevitably have lasting and sometimes irreversible impacts.

There are two different classes of tests for Machine Learning systems:

- Pre-train tests
- Post-train tests

**Pre-train tests:** The intention is to write such tests which can be run without trained parameters so that we can catch implementation errors early on. This helps in avoiding the extra time and effort spent in a wasted training job. We can test the following in the pre-train test:

- The model predicted output shape is proper or not.
- Test dataset leakage i.e., checking whether the data in training and testing datasets have no duplication
- Temporal data leakage which involves checking whether the dependencies between training and test data do not lead to unrealistic situations in the time domain like training on a future data point and testing on a past data point.

- Check for the output ranges. In the cases where we are predicting outputs in a certain range (for example when predicting probabilities), we need to ensure the final prediction is not outside the expected range of values.
- Ensuring a gradient step training on a batch of data leads to a decrease in the loss.
- Data profiling assertions.

**Post-train tests:** post-train tests are aimed at testing the model's behavior. We want to test the learned logic and it could be tested on the following points and more:

- Invariance tests which involve testing the model by tweaking only one feature in a data point and checking for consistency in model predictions. For example, if we are working with a loan prediction dataset then change in sex should not affect an individual's eligibility for the loan given all other features are the same or in the case of titanic survivor probability prediction data, change in the passenger's name should not affect their chances of survival.
- Directional expectations wherein we test for a direct relation between feature values and predictions. For example, in the case of a loan prediction problem, having a higher credit score should definitely increase a person's eligibility for a loan.
- Apart from this, you can also write tests for any other failure modes identified for your model

## **Chapter – 7**

### **CONCLUSION AND FUTURE WORK**

The major industries using Artificial Intelligence (AI) and Machine Learning (ML) include Agriculture, Education and Infrastructure, Healthcare, Transport, Banking, Cyber Security, Manufacturing, Entertainment, Hospitality, and others. AI has a great potential to bring a change in the economy of the country. Also, the best part about AI is that, there is not just one Technology under it but several other technologies including machine learning, self-improving algorithms, pattern recognition, big data, and others.

The emerging scope of ML will enhance the performance of machines with less intervention of human beings. The main motive of the integration of machine learning into various domains is to reduce error function and improve the real time results at minimum cost and time

## REFERENCES

- [1] <https://www.ibm.com/in-en/topics/computer-vision>
- [2] <https://www.ibm.com/cloud/learn/natural-language-processing>
- [3] <https://www.analyticsvidhya.com/blog/2020/10/time-series-forecasting-using-facebook-prophet-library-in-python/>
- [4] <https://guides.nyu.edu/tesseract>