**A Internship**

ON

**“ MACHINE LEARNING WITH PYTHON ”**

# OF

**TE (AI &DS Engineering)**

**Academic Year: 2022-2023**

SUBMITTED BY

**Tejas Sadashiv Sawant**

**Roll No.: T1411065**

GUIDED BY

**Prof. Dikshenrda Sarpate**



**Department of AI & DS Engineering**

**Zeal Education Society’s**

**Zeal College of Engineering & Research Narhe, Pune-411041**

**Zeal College Of Engineering & Research, Narhe, Pune**

**2022-2023**

**CERTIFICATE**



This is to certify that the internship report entitled “**MACHINE LEARNING WITH PYTHON**” has been carried out by **Tejas Sadashiv Sawant (T1411065)** under my guidance in partial fulfilment of the engineering of **AI & DS Engineering** of **Zeal College Of Engineering & Research, Narhe, Pune** during Academic year **2022-2023.** To the best of knowledge and belief this work has not seen submitted elsewhere for the award of any other Engineering.

|  |  |
| --- | --- |
| **Prof. Dikshendra Sarpate** | **Prof. Dikshendra Sarpate** |
| **Guide** | **H.O.D** |
| **AI & DS Engineering** | **AI & DS Engineering** |
| **Department** | **Department** |

**Place: Pune Date:**

**INTERNSHIP OFFER LETTER**

****

**INTERNSHIP CERTIFICATE**

A certificate of internship

Description automatically generated with medium confidence

**Industry Details :**

* **Name of the Industry** – TECHNICAL CODING RESEARCH INNOVATION.
* **Address of the Industry** - TECHNICAL CODING RESEARCH INNOVATION, 4B First floor, Morya Residency, Usarli-Khurd, Vichumbe, New Panvel, Maharashtra, India- 410206.
* **Website Address of the Industry** - <https://tcrinnovation.co.in>
* **Managing Director & Co-Founder -** Mrs. Neelam Bafna

**Industry IT Services :**

**Research and Innovation:**

Research and innovation plays an essential role in triggering smart and sustainable growth of the company. Working in a team on various technical projects and writing research papers along with blogs and get them published here.

**Consulting Services:**

Our management consulting services focus on our clients' most critical issues and opportunities: strategy, marketing, organization, operations, technology, transformation, digital, advanced analytics, corporate finance, mergers & acquisitions and sustainability across all industries and geographies. We bring deep, functional expertise, but are known for our holistic perspective: we capture value across boundaries and between the silos of any organization. We have proven a multiplier effect from optimizing the sum of the parts, not just the individual pieces.

**Internship Programs:**

Our Internship Programs are specially designed so Interns can learn new skills or enhance their skills in the most efficient way. Interns are First trained by the Mentors in their applied Domain through Live Lectures. Comprehensive Notes & Study Materials are provided along with Intensive Training in the domain. After the Training Period is Complete, Interns are given the Opportunity to work on an Industrial-Level Project which will be the Project Phase and one of the most Important Phase of their Internship. The Training completion certificate  along with the Internships Certificate are awarded after the Successful Completion of your Project with Letter of Recommendation.

**Details of the External Mentor (Supervisor) from Industry:**

* **Name** – Imran Kabir
* **Contact Number** - +91 66926 28467
* **Email ID** – [imran.kabir@gmail.com](mailto:imran.kabir@gmail.com)

**LIST OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Contents** | **Page No.** |
|  | INTRODUCTION | 8 |
|  | OBJECTIVES | 10 |
|  | MOTIVATION | 11 |
|  | METHODOLOGICAL DETAILS | 12 |
|  | RESULT | 19 |
|  | ACKNOWLDEGEMENT | 31 |
|  | FUTURE SCOPE | 32 |
|  | CONCLUSION | 33 |
|  | REFERENCES | 34 |

Chapter 1

**INTRODUCTION**

A web development internship is an opportunity for individuals interested in learning how to build websites and web applications to gain practical, hands-on experience working on real-world projects. It typically involves working closely with experienced web developers, designers, and other professionals in a team environment to build, test, and deploy web-based solutions.

During a web development internship, interns may be responsible for a range of tasks, including designing user interfaces, writing code in various programming languages, testing and debugging web applications, and collaborating with team members to ensure project goals are met. The specific responsibilities and expectations of the internship may vary depending on the organization and the specific internship program.

Overall, a web development internship is a valuable opportunity for individuals looking to gain experience in the field, build their skills, and make connections within the industry. It can be a stepping stone to a career in web development or related fields, and it can also help interns to identify areas of interest and specialize their skills.

**About the Company**

Headquartered in Mumbai, TCR Engineering Services is an ISO 17025 and NABL accredited independent Material Testing and Quality Assurance Laboratory serving 2500+ customers globally.  Established in 1973, TCR has a trusted legacy with a strong presence in India and internationally in countries like Saudi Arabia, Kuwait and UAE. TCR enables organizations across the globe to develop and execute solutions for efficiently managing plant operations. TCR aims to innovate in a way that minimizes the gap between their offerings and their client needs.

For almost half a century, TCR has built an enterprise that is distinctly known for its honesty, reliability and transparency.  TCR’s team is distinguished by knowledge, imagination and experience gained across industries and that is reflected in every project they undertake. TCR, because of its global presence, can rapidly assemble the right team with the right experience to help clients anywhere in the world. TCR has worked with several industries and verticals that include Automotive, Oil Refineries, Petrochemical plants, Chemical Processing, Defense, Electronics, Nuclear Power, Capital Goods, and manufacturing industries to determine material properties, improve product performance, assist in developing new and better products/materials, evaluate remaining life of an industrial equipment, understand reasons for unmet expectation for a component’s performance and or to identify why a product may have failed.

TCR delivers unbiased results on time, every single time. The multi-disciplinary certified and experienced team of professionals at TCR is inclusive of metallurgical, mechanical, electrical, and chemical engineers; materials scientists; chemists; physicists; NDT inspectors and computer scientists who are skilled to meet rigorous standards in the testing field, whether serving the Private, Public Sector, Government or the Military.   In recent years, TCR is recognized as one of the fastest growing innovative and successful companies in India. The company won the prestigious award from NACE International for “Excellent Laboratory in Private Sector” in September 2007.

* 1. **Vision**

To be recognized as the leading global procurement and supply chain consulting and outsourced services firm

* 1. **Dedication**

Dedication is a consistent feeling of loyalty and support to a person, organization, or idea. Dedication in the workplace means an employee will have a strong work ethic and seek to improve both themselves and the organization around them because they feel strongly connected to their work environment.

* 1. **Strategy**

Information technology companies develop growth strategies to increase their revenue and profit. Depending on their stage of development, they may aim to grow organically or they may try to accelerate growth through acquisitions, mergers or partnerships. The consultancy firm Deloitte notes that any growth strategy involves risk and technology companies must manage the risk to succeed.

Chapter 2

**OBJECTIVES**

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it is expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24.

hours by a doctor are not available since it requires more sapience, time and expertise. Since we have a good amount of data in today’s world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

The main objectives of developing this project are:

* To develop machine learning model to predict future possibility of heart disease by implementing Logistic Regression, SVM, KNN like Machine Learning techniques.
* To determine significant risk factors based on medical dataset which may lead to heart disease.
* To analyse feature selection methods and understand their working principle.

Chapter 3

**MOTIVATION**

Machine learning techniques have been around us and has been compared and used for analysis for many kinds of data science applications. The major motivation behind this research-based project was to explore the feature selection methods, data preparation and processing behind the training models in the machine learning. With first hand models and libraries, the challenge we face today is data where beside their abundance, and our cooked models, the accuracy we see during training, testing and actual validation has a higher variance.

Hence this project is carried out with the motivation to explore behind the models, and further implement Logistic Regression model to train the obtained data. Furthermore, as the whole machine learning is motivated to develop an appropriate computer-based system and decision support that can aid to early detection of heart disease, in this project we have develop a model which classifies if patient will have heart disease in ten years or not based on various features (i.e., potential risk factors that can cause heart disease) using logistic regression. Hence, the early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine.

Chapter 4

**METODLOGICAL DETAILS**

**4.1 What is Machine Learning?**

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

A subset of machine learning is closely related to [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics), which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization) delivers methods, theory and application domains to the field of machine learning.

**4.2 Types of Machine Learning**

The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve. Broadly Machine Learning can be categorized into four categories.

1. **Supervised Learning**
2. **Unsupervised Learning**
3. **Reinforcement Learning**
4. **Semi-supervised Learning**

**4.2.1 Supervised Learning**

Supervised Learning is a type of learning in which we are given a data set and we already know what are 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. It infers a function from labeled training data consisting of a set of training examples. Supervised learning problems are categorized.

**4.2.2 Unsupervised Learning**

Unsupervised Learning 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. Basically, it is a type of self-organized learning that helps in finding previously unknown patterns in data set without pre-existing label.

**4.2.3 Reinforcement Learning**

Reinforcement learning 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. Simple reward feedback is required for the agent to learn which action is best.

**4.2.4 Semi-supervised Learning**

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

**4.3 Tools and Technologies used**

**4.3.1 Jupyter**

Jupyter Lab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality.

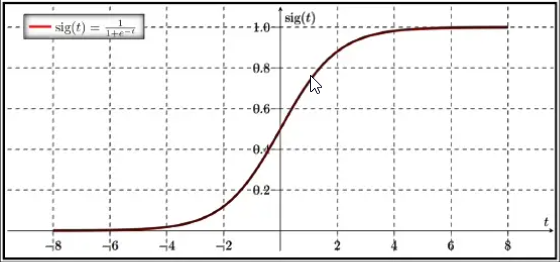
**4.3.2 Python**

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for an emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included language due to its comprehensive standard library.

**4.3.3 Logistic Regression**

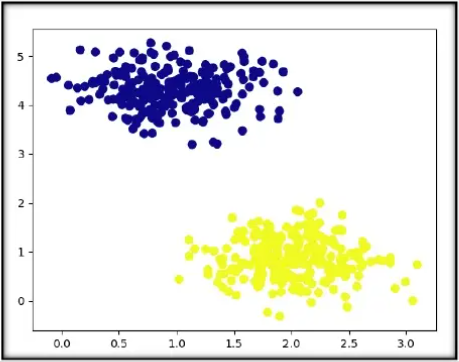
Logistic regression is a supervised classification is unique Machine Learning algorithms in Python that finds its use in estimating discrete values like 0/1, yes/no, and true/false. This is based on a given set of

independent variables. We use a logistic function to predict the probability of an event and this gives us an output between 0and 1. Although it says 'regression', this is actually a classification algorithm. Logistic regression fits data into a logit function and is also called logit regression.



**4.3.4 Support Vector Machine (SVM)**

SVM is a supervised classification is one of the most important Machine Learning algorithm in Python that plots a line that divides different categories of your data. In this ML algorithm, we calculate the vector to optimize the line. This is to ensure that the closest point in each group lies farthest from each other. While you will almost always find this to be a linear vector, it can be other than that. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMS can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. When data is unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups.



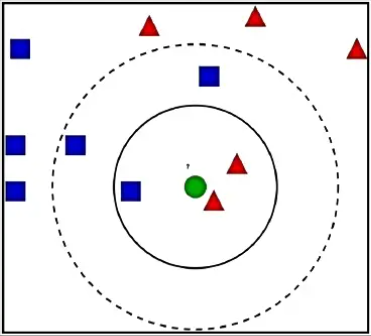
**4.3.5 K-Nearest Neighbor (KNN)**

This is a Python Machine Learning algorithm for classification and regression- mostly for classification.

This is a supervised learning algorithm that considers different centroids and uses a usually Euclidean

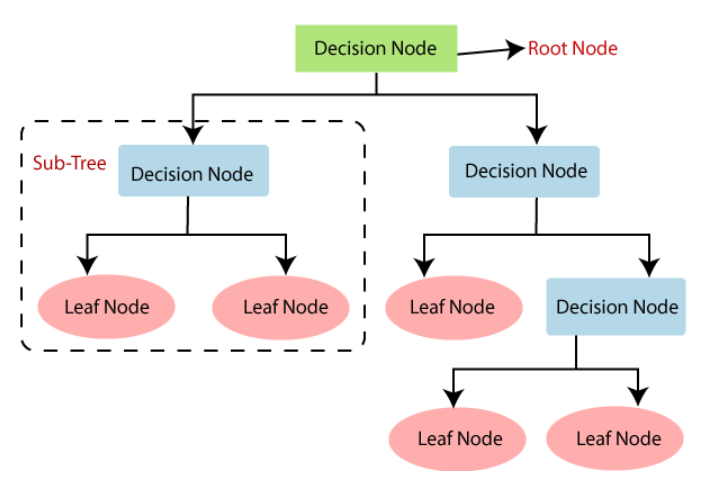
function to compare distance. Then, it analyzes the results and classifies each point to the group to optimize it to place with all closest points to it. It classifies new cases using a majority vote of k of its neighbors. The case it assigns to a class is the one most common among its K nearest neighbors. For this, it uses a distance function. k-NN is a type of instance-based learning, or lazy learning, where the

function is only approximated locally and all computation is deferred until classification. k-NN is a special case of a variable-bandwidth, kernel density "balloon" estimator with a uniform kernel.



**4.3.6 Decision Tree**

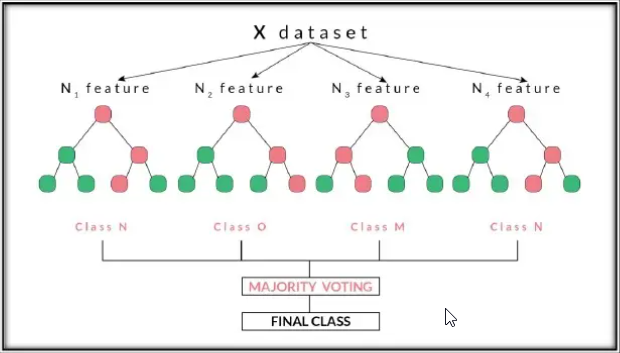
A decision tree falls under supervised Machine Learning Algorithms in Python and comes of use for both classification and regression- although mostly for classification. This model takes an instance, traverses the tree, and compares important features with a determined conditional statement. Whether it descends to the left child branch or the right depends on the result. Usually, more important features are closer to the root. Decision Tree, a Machine Learning algorithm in Python can work on both categorical and continuous dependent variables. Here, we split a population into two or more homogeneous sets. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.



**4.3.7 Random Forest**

A random forest is an ensemble of decision trees. In order to classify every new object based on its

attributes, trees vote for class- each tree provides a classification. The classification with the most votes wins in the forest. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.



**4.2.8 Hardware Requirement Specification**

Processor - Pentium III 630MHz or more

RAM - 128 MB or more

Hard Disc - 20 GB or more

Monitor - 15” colour monitor or advance

Keyboard - Any Keyboard

Mouse - Any mouse

Printer - In case of printing reports

**4.2.9 Software Requirement Specification**

Operating System - Windows 98, Windows XP, Windows7, Windows 10,

Linux

Language - Python

Browser - Any of Mozilla, Opera, Chrome etc

Chapter 5

**RESULTS**

**Code:**

\*\*Project name\*\* - Prediction of Heart disease detection

"""\*\*Import Required Libraries\*\*"""

# Commented out IPython magic to ensure Python compatibility.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# %matplotlib inline

import os

import warnings

warnings.filterwarnings('ignore')

"""\*\*Import dataset\*\*"""

dataset = pd.read\_csv("/content/drive/MyDrive/TCR Internship

Project/heart.csv")

dataset

"""\*\*Shape of dataset\*\*"""

dataset.shape

"""\*\*Some Operations on dataset\*\*"""

dataset.head()

dataset.tail()

type(dataset)

dataset.info()

dataset.describe()

dataset.columns

"""\*\*Checking total number of NA values\*\* """

dataset.isna().sum()

"""\*\*Checking total number of NULL values\*\* """

dataset.isnull().sum()

"""\*\*Exploratory Data Analysis (EDA)\*\*

\*\*Analysing the 'target' variable\*\*

"""

dataset.target.describe()

dataset.target.unique()

#Checking correlation between columns

dataset.corr()["target"].abs().sort\_values(ascending=False)

#This shows that most columns are moderately correlated with target, but

'fbs' is very weakly correlated.

dataset.target.value\_counts()

"""Patient without heart problems - labeled as 0

Patient with heart problems - labeled as 1

"""

print("Percentage of patients without heart problems:

"+str(round(138\*100/303,2)))

print("Percentage of patients with heart problems:

"+str(round(165\*100/303,2)))

y = dataset["target"]

sns.countplot(y)

sns.distplot(dataset['target'])

"""\*\*Analysing the 'sex' variable\*\*"""

dataset.sex.value\_counts()

sns.barplot(dataset["sex"],y)

"""We notice that the 'sex' feature has 2 unique features."""

sns.distplot(dataset['sex'])

"""\*\*Analysing the 'cp' variable\*\*"""

dataset.cp.value\_counts()

sns.barplot(dataset["cp"],y)

""" The CP feature has values from 0 to 3.We notice, that chest pain of '0', are much less likely to have heart problems""

sns.distplot(dataset['cp'])

"""\*\*Analysing the 'age' variable\*\*"""

dataset.age.value\_counts()

sns.barplot(dataset["age"],y)

"""Nothing special here."""

sns.distplot(dataset['age'])

"""\*\*Analysing the 'trestbps' variable\*\*"""

dataset.trestbps.value\_counts()

sns.barplot(dataset["trestbps"],y)

"""Nothing special here."""

sns.distplot(dataset['trestbps'])

"""\*\*Analysing the 'chol' variable\*\*"""

dataset.chol.value\_counts()

sns.barplot(dataset["chol"],y)

"""Nothing special here """

sns.distplot(dataset['chol'])

"""\*\*Analysing the 'fbs' variable\*\*"""

dataset.fbs.value\_counts()

sns.barplot(dataset["fbs"],y)

"""Not much difference here."""

sns.distplot(dataset['fbs'])

"""\*\*Analysing the 'restecg' variable\*\*"""

dataset.restecg.value\_counts()

sns.barplot(dataset["restecg"],y)

"""We realize that people with restecg '1' and '0' are much more likely to

have a heart disease than with restecg '2'"""

sns.distplot(dataset['restecg'])

"""\*\*Analysing the 'exang' variable\*\*"""

dataset.exang.value\_counts()

sns.barplot(dataset["exang"],y)

"""We notice here that people with exang=1, are much less likely to have

heart problems."""

sns.distplot(dataset['exang'])

"""\*\*Analysing the 'slope' variable\*\*"""

dataset.slope.value\_counts()

sns.barplot(dataset["slope"],y)

"""We observe, that Slope '2' causes heart pain much more than Slope '0'

and '1'"""

sns.distplot(dataset['slope'])

"""\*\*Analysing the 'ca' variable\*\*"""

dataset.ca.value\_counts()

sns.barplot(dataset["ca"],y)

"""We notice that ca=4 has large number of heart patients."""

sns.distplot(dataset['ca'])

"""\*\*Analysing the 'thal' variable\*\*"""

dataset.thal.value\_counts()

sns.barplot(dataset["thal"],y)

"""thal=2 has large number of heart patients."""

sns.distplot(dataset['thal'])

"""\*\*Get an overview distribution of each column\*\*"""

dataset.hist(figsize=(16, 20), xlabelsize=8, ylabelsize=8)

sns.pairplot(dataset, hue='target')

"""\*\*Correlation heatmap\*\*"""

dataset.corr()

f, ax = plt.subplots(figsize=(15, 10))

sns.heatmap(dataset.corr(),annot=True,cmap='PiYG',linewidths=.5)

"""\*\*Splitting the data -Train Test split\*\*"""

from sklearn.model\_selection import train\_test\_split

x = dataset.drop("target",axis=1)

y= dataset["target"]

X\_train,X\_test,Y\_train,Y\_test =

train\_test\_split(x,y,test\_size=0.20,random\_state=0)

X\_train.shape

X\_test.shape

Y\_train.shape

Y\_test.shape

from sklearn.metrics import accuracy\_score

"""\*\*Logistic Regression\*\*"""

from sklearn.linear\_model import LogisticRegression

model\_logistic\_reg = LogisticRegression()

model\_logistic\_reg.fit(X\_train,Y\_train)

Y\_pred\_logistic\_reg = model\_logistic\_reg.predict(X\_test)

Y\_pred\_logistic\_reg.shape

print("Predicted Values : ",Y\_pred\_logistic\_reg)

Y\_test[0:10] #You can check accuracy by observing predicted results and

test data.

accuracy\_score\_logistic\_reg =

round(accuracy\_score(Y\_pred\_logistic\_reg,Y\_test)\*100,2)

print("The accuracy score achieved using Logistic Regression is:

"+str(accuracy\_score\_logistic\_reg)+" %")

"""\*\*SVM\*\*"""

from sklearn import svm

model\_svm = svm.SVC(kernel='linear')

model\_svm.fit(X\_train, Y\_train)

Y\_pred\_svm = model\_svm.predict(X\_test)

Y\_pred\_svm.shape

print("Predicted Values : ",Y\_pred\_svm)

Y\_test[0:10] #You can check accuracy by observing predicted results and

test data.

accuracy\_score\_svm = round(accuracy\_score(Y\_pred\_svm,Y\_test)\*100,2)

print("The accuracy score achieved using Linear SVM is:

"+str(accuracy\_score\_svm)+" %")

"""\*\*K Nearest Neighbors\*\*"""

from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n\_neighbors=7)

knn.fit(X\_train,Y\_train)

Y\_pred\_knn=knn.predict(X\_test)

Y\_pred\_knn.shape

print("Predicted Values : ",Y\_pred\_knn)

Y\_test[0:10] #You can check accuracy by observing predicted results and

test data.

accuracy\_score\_knn = round(accuracy\_score(Y\_pred\_knn,Y\_test)\*100,2)

print("The accuracy score achieved using KNN is: "+str(accuracy\_score\_knn)+" %")

"""\*\*Decision Tree\*\*"""

from sklearn.tree import DecisionTreeClassifier

max\_accuracy = 0

for x in range(200):

dt = DecisionTreeClassifier(random\_state=x)

dt.fit(X\_train,Y\_train)

Y\_pred\_dt = dt.predict(X\_test)

current\_accuracy = round(accuracy\_score(Y\_pred\_dt,Y\_test)\*100,2)

if(current\_accuracy>max\_accuracy):

max\_accuracy = current\_accuracy

best\_x = x

dt = DecisionTreeClassifier(random\_state=best\_x)

dt.fit(X\_train,Y\_train)

Y\_pred\_dt = dt.predict(X\_test)

print(Y\_pred\_dt.shape)

print("Predicted Values : ",Y\_pred\_dt)

Y\_test[0:10] #You can check accuracy by observing predicted results and

test data.

accuracy\_score\_dt = round(accuracy\_score(Y\_pred\_dt,Y\_test)\*100,2)

print("The accuracy score achieved using Decision Tree is:

"+str(accuracy\_score\_dt)+" %")

"""\*\*Random Forest\*\*"""

from sklearn.ensemble import RandomForestClassifier

max\_accuracy = 0

for x in range(2000):

rf = RandomForestClassifier(random\_state=x)

rf.fit(X\_train,Y\_train)

Y\_pred\_rf = rf.predict(X\_test)

current\_accuracy = round(accuracy\_score(Y\_pred\_rf,Y\_test)\*100,2)

if(current\_accuracy>max\_accuracy):

max\_accuracy = current\_accuracy

best\_x = x

rf = RandomForestClassifier(random\_state=best\_x)

rf.fit(X\_train,Y\_train)

Y\_pred\_rf = rf.predict(X\_test)

Y\_pred\_rf.shape

print("Predicted Values : ",Y\_pred\_rf)

Y\_test[0:10] #You can check accuracy by observing predicted results and

test data.

accuracy\_score\_rf = round(accuracy\_score(Y\_pred\_rf,Y\_test)\*100,2)

print("The accuracy score achieved using Random Forest is:

"+str(accuracy\_score\_rf)+" %")

"""\*\*Summary of accuracy scores\*\*"""

all\_accuracy\_scores =

[accuracy\_score\_logistic\_reg,accuracy\_score\_svm,accuracy\_score\_knn,accuracy

\_score\_dt,accuracy\_score\_rf]

algorithms\_used = ["Logistic Regression","Support Vector Machine","KNearest Neighbors","Decision Tree","Random Forest"]

for i in range(len(algorithms\_used)):

print("\nThe accuracy score achieved using "+algorithms\_used[i]+" is:

"+str(all\_accuracy\_scores[i])+" %")

sns.set(rc={'figure.figsize':(15,8)})

plt.xlabel("Algorithms")

plt.ylabel("Accuracy score")

sns.barplot(algorithms\_used,all\_accuracy\_scores)

"""\*\*Here we can see that Random Forest is better than other algorithms.\*\*"""

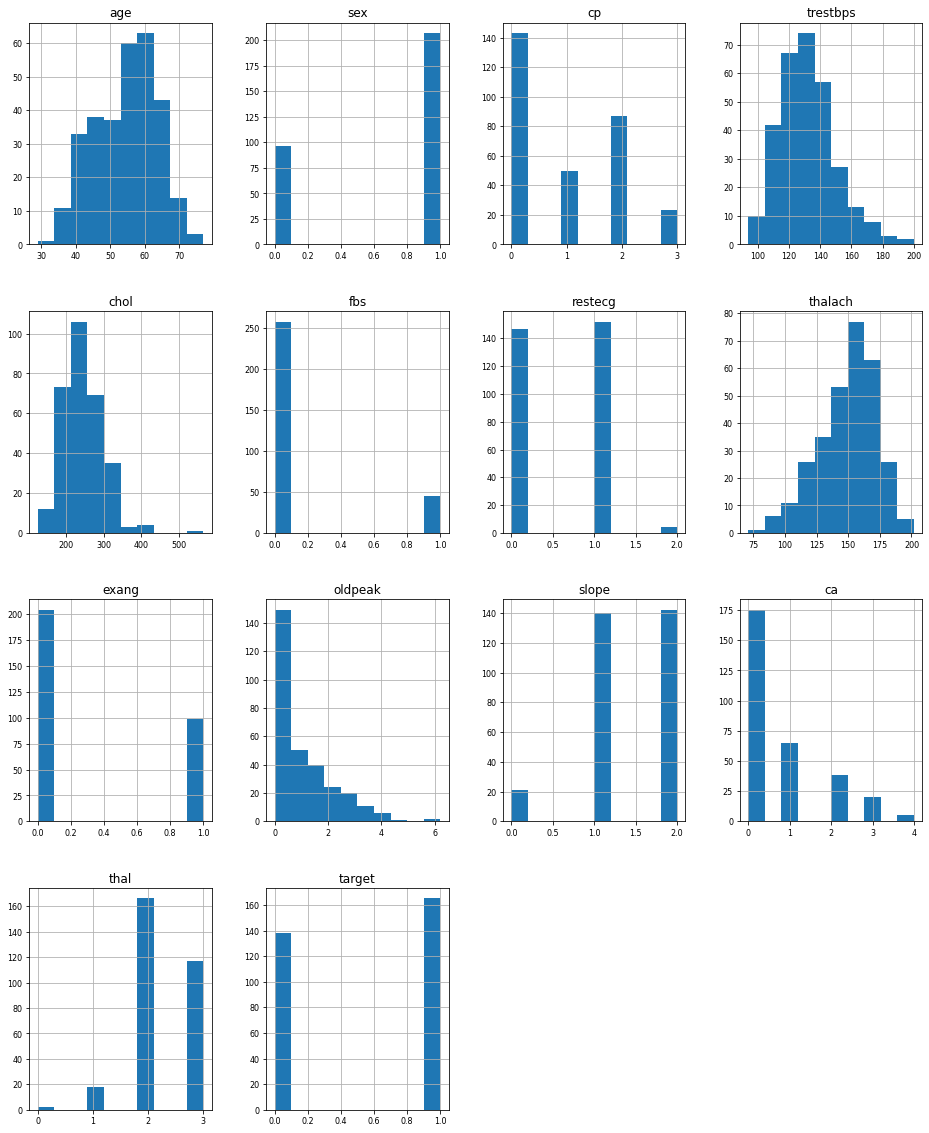


Fig.1: Overview distribution of each column

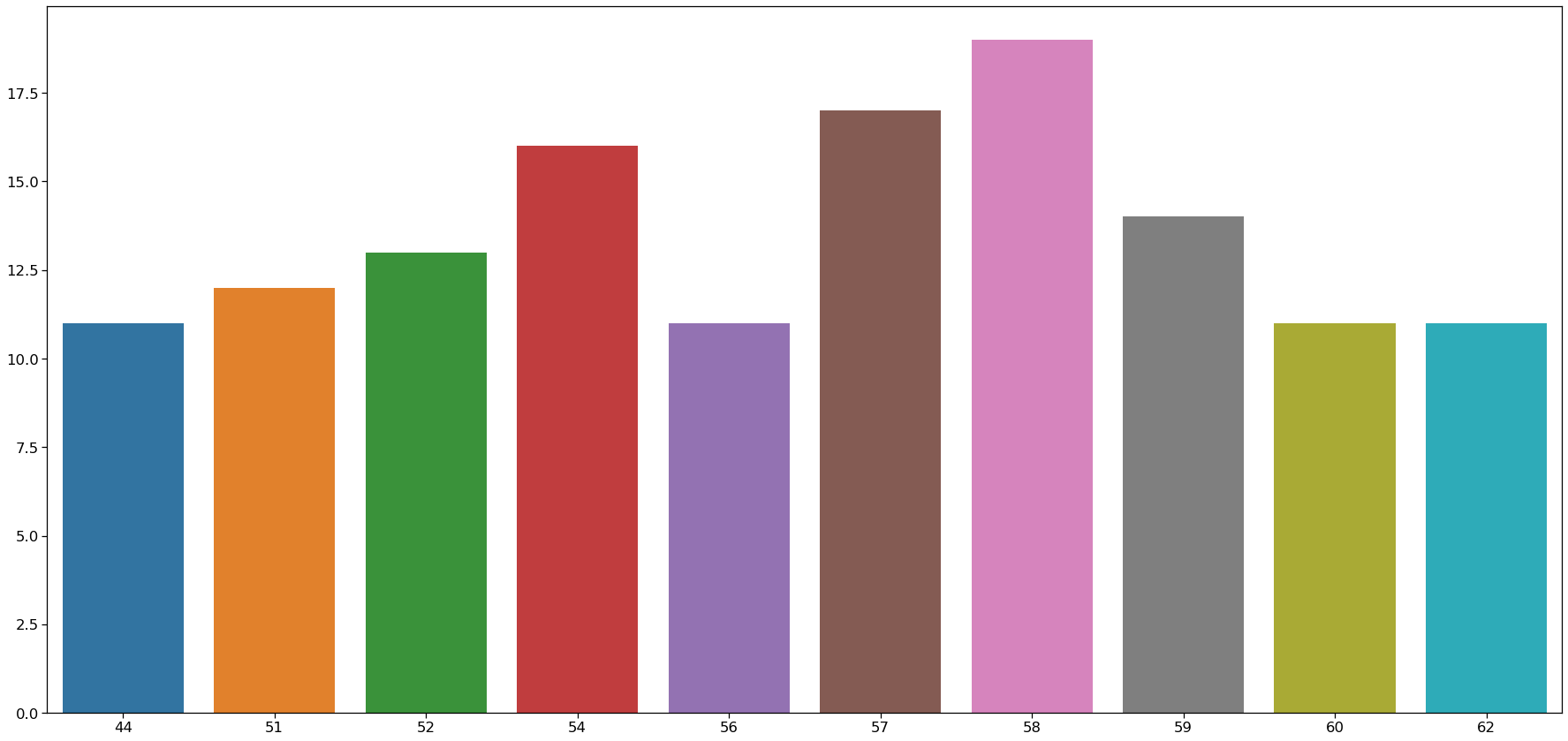
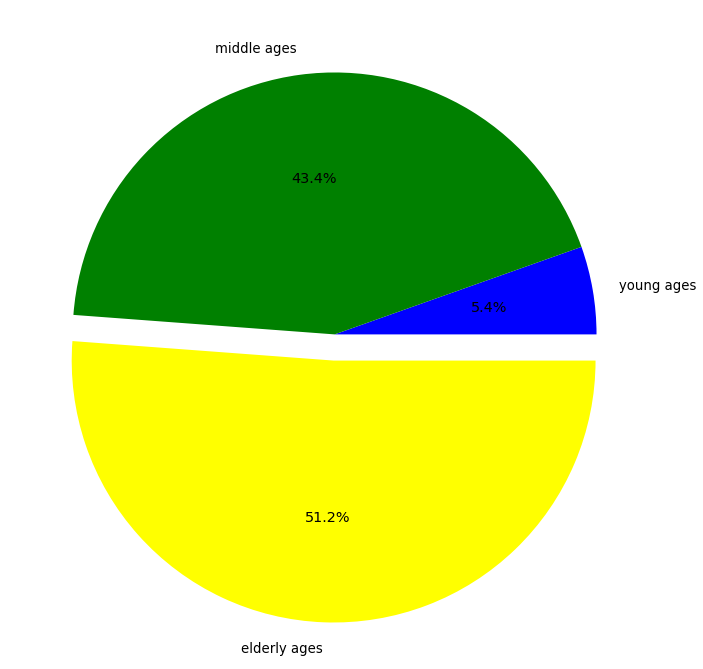
 

Fig.2: Age Analysis

Here we can see that elder people are the most affected by heart disease and young ones are the least affected.



Fig.3:Target analysis

Patient without heart problems - labeled as 0 Patient with heart problems - labeled as 1

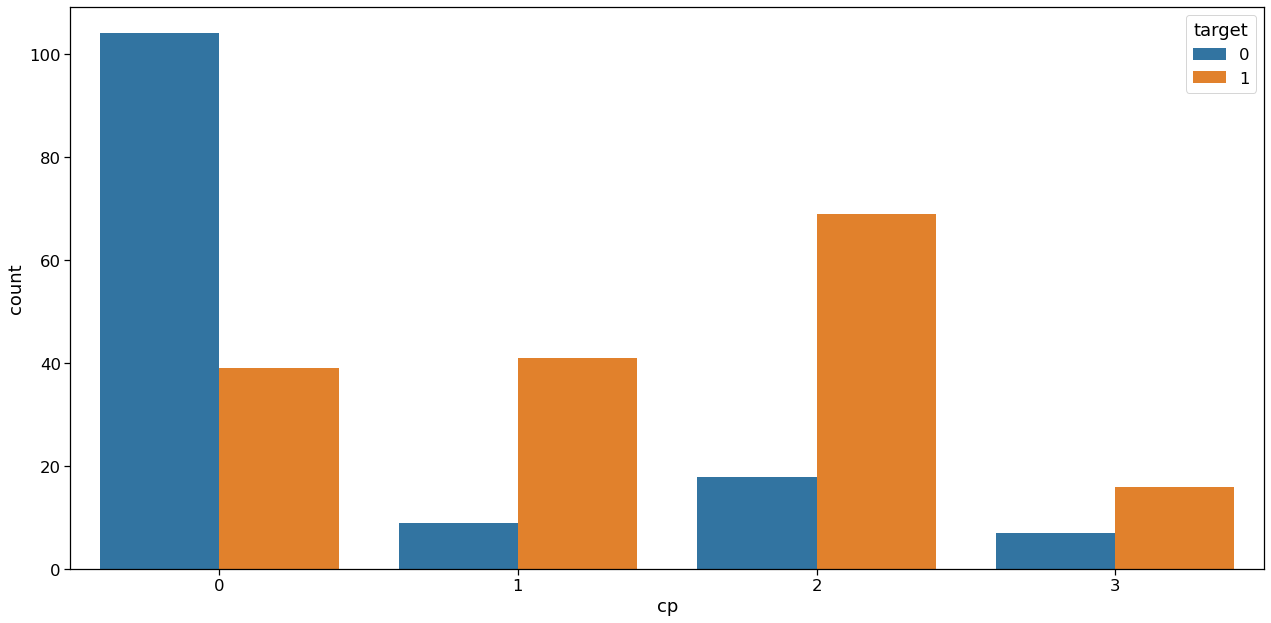


Fig.4: Analyzing cp vs target column

Inference: From the above graph we can make some inferences,

People having the least chest pain are not likely to have heart disease.

People having severe chest pain are likely to have heart disease.

Elderly people are more likely to have chest pain.

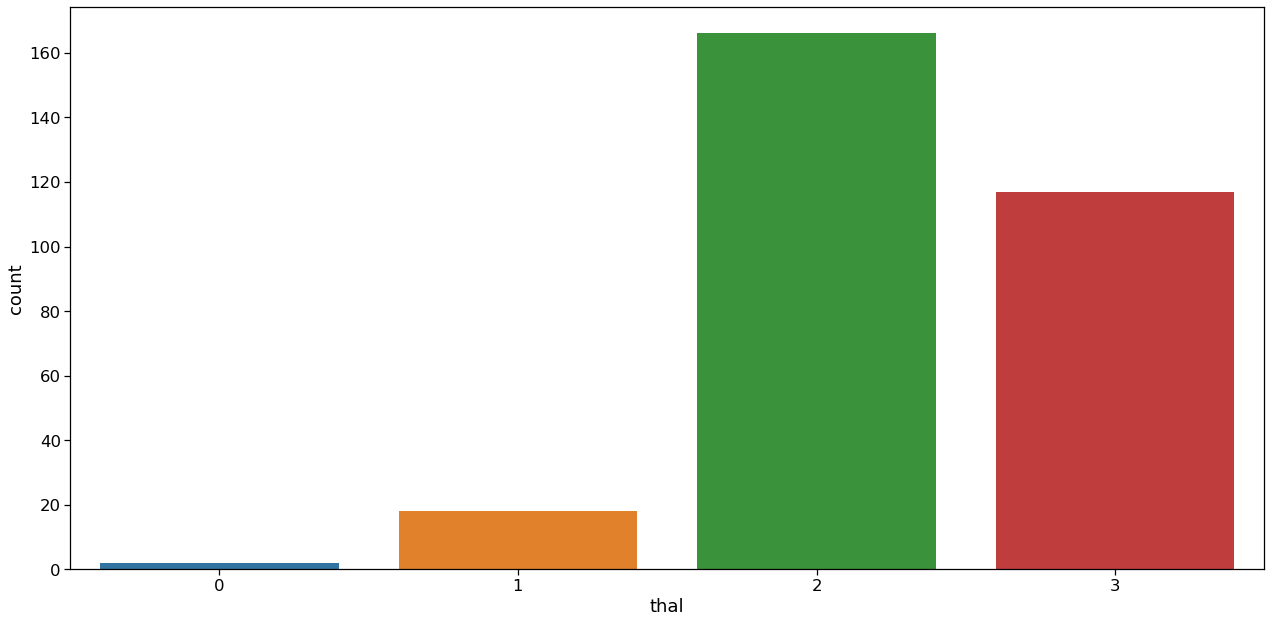


Fig.3: Thal Analysis

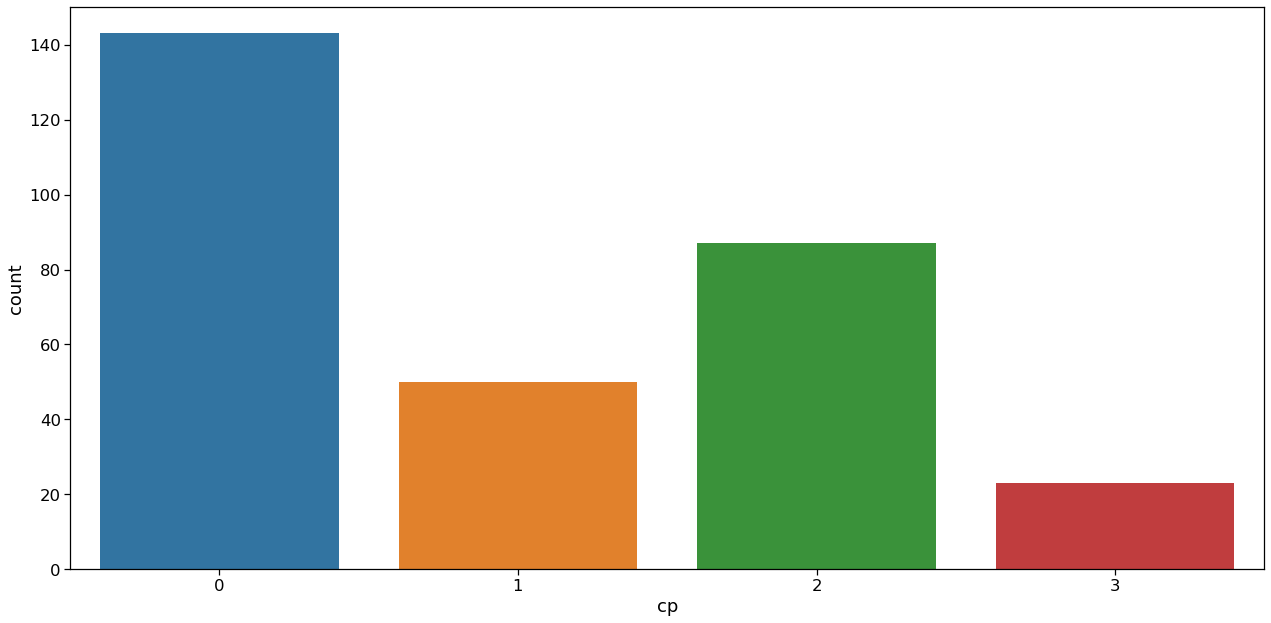


Fig.4: Chest Pain Type(“cp”) Analysis

The CP feature has values from 0 to 3. We notice, that chest pain of '0', are much less likely to have heart problems. As seen, there are 4 types of chest pain

1)status at least 2) condition slightly distressed

3)condition medium problem 4) condition too bad.

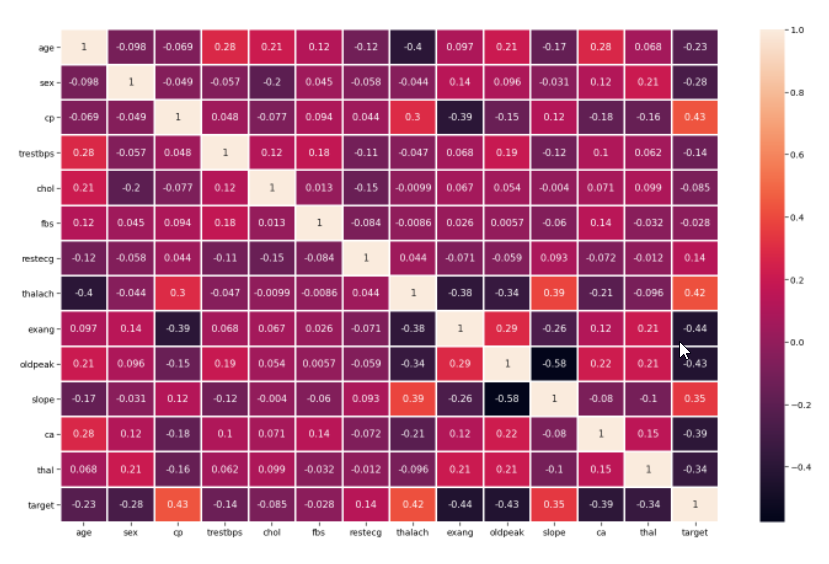


Fig.5: Visualization to find Correlation between data.

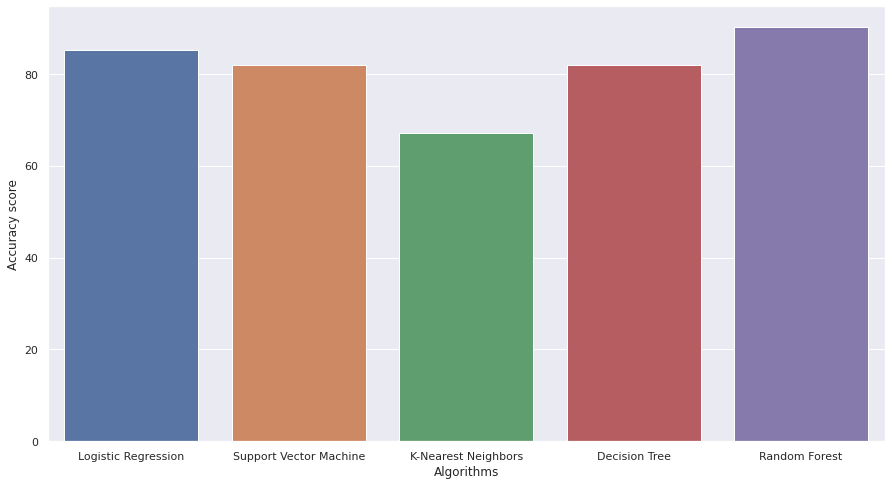


Fig.6: Comparison between different ML algorithms

Chapter 6

**ACKNOWLEDGEMENT**

Internship is an integral part of engineering curriculum providing engineers with first hand and practical aspects of their studies. It gives us the knowledge about the work and circumstances existing in the company. The internship opportunity I had with **Technical Coding Research Innovation** was a great chance for learning and professional development. Therefore, I consider myself very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to meet so many wonderful people and professionals who led me though this internship period.

It gives me great pleasure in completing training at **Technical Coding Research Innovation** and submitting the training report for the same. I am thankful to the company for the whole experience.

We are indebted to our Internship Supervisor **Prof. Dikshendra Sarpate** from Zeal college of engineering & Research. We feel it’s a pleasure to be indebted to our guide for her valuable support, advice and encouragement and we thank her for her superb and constant guidance towards this project.

We sincerely thank **Prof. Dikshendra Sarpate** Head of the Department of Computer Engineering and for all the facilities provided to us in the pursuit of this internship.

Chapter 7

**FUTURE SCOPE**

Future of Machine Learning is as vast as the limits of human mind. We can always keep learning, and

teaching the computers how to learn. And at the same time, wondering how some of the most complex

machine learning algorithms have been running in the back of our own mind so effortlessly all the time.

There is a bright future for machine learning. Companies like Google, Quora, and Facebook hire people with machine learning. There is intense research in machine learning at the top universities in the world. The global machine learning as a service market is rising expeditiously mainly due to the Internet revolution. The process of connecting the world virtually has generated vast amount of data which is boosting the adoption of machine learning solutions. Considering all these applications and dramatic improvements that ML has brought us, it doesn't take a genius to realize that in coming future we will definitely see more advanced applications of ML, applications that will stretch the capabilities of machine learning to an unimaginable level.

Chapter 8

**CONCLUSION**

This internship at TCR INNOVATIONS, has aided me in gaining practical experience. Online internship of this kind helped me to get used to an online project atmosphere and thus prepared me to face the world of engineering in much better and confident way. I’m sure that after completing this internship I will be able to visualize the work situation in a better way.

Thus, I confidently conclude that this internship was beneficial in its technical and educational aspect, the experience is bound to help me in the years to come.

Chapter 9

**REFERENCES**

[1] Mozilla Developer Network (MDN) - https://developer.mozilla.org/en-US/docs/Web

[2] W3Schools - https://www.w3schools.com/

[3] Stack Overflow - https://stackoverflow.com/

[4] GitHub - https://github.com/

[5] Codecademy - https://www.codecademy.com/

[6] CodePen - https://codepen.io/

[7] Smashing Magazine - https://www.smashingmagazine.com/

[8] A List Apart - https://alistapart.com/

[9] Google Developers - https://developers.google.com/

[10] Web Platform Docs - https://webplatform.github.io/docs/