NATIONAL INSTITUTE OF TECHNOLOGY UTTARAKHAND

Summer training Presentation

Submitted to:
Computer science and engineering, NIT UTTARAKHAND

Submitted by : Pankaj Kumar
Roll no: BT18CSE010

ICT Summer Training Certificate

This is to certify that

Dr. / Mr. / Ms. **Pankaj Kumar** has successfully

completed a 6 weeks course on

Artificial Intelligence

from ICT - Indian Institute of Technology, Kanpur

Prof. B. V. Phani

Prof. Amey Karkare

E & ICT Coordinators

Date of Issue: 29-07-2020

System Identification No.:
14757-399186-6026aecb81620920

OVERVIEW WHAT I LEARNED DURING SUMMER TRAINING

- Python
- Machine learning
- Different libraries of python
- Different algorithm for machine learning

Different strategy to separate data

- Train test split technique
- K-fold cross validation technique
- Leave one out cross validation

Different performance evaluation metrices

For logistic regression
- Classification accuracy
- Logarithmic loss
- Area under ROC curve
- Confusion matrix

For Linear regression
- Mean absolute error
- Mean squared error

Summer training project

I made two project in summer training. One with help of class teacher and second one was assignment for marks evaluation and I done that perfectly.

1. **Brain tumor**: Model is based on brain tumor dataset and you can find out it easily on Kaggle site. This is done with help of class teacher and got an accuracy of 96 % (approx.)

2. **Heart attack** : This is my real work. This is trained on heart attack dataset which is also available on Kaggle. This was given as a assignment by professor for marks evaluation purpose. I worked successfully and got accuracy as 83% approx. using BernouliiNB algorithm.
   So lets discuss it in detail now….

Libraries used in project

- NumPy
  This is a python libraries to work smoothly with n dimensional array. In this libraries we can play with array easily and this is easy to use.

- Pandas
  This is one of the important libraries of python for machine learning. This libraries is mainly used for data reading and cleaning. It read every thing like a dataframe. Dataframe is a kind of data structure and pandas read everything as a dataframe which allow to change information of any row or column easily.

- Matplotlib
  This is a plotting libraries and it run on top of NumPy. We can use it for data visualization.

- Seaborn
  Seaborn is also a plotting libraries of python but it is used to draw statistical graphics like confusion matrix, normal distribution

- Sklearn
  It is a free python libraries which has all the algorithm of regression, classification, etc.

- Joblib
  I used it only to save model for further prediction.
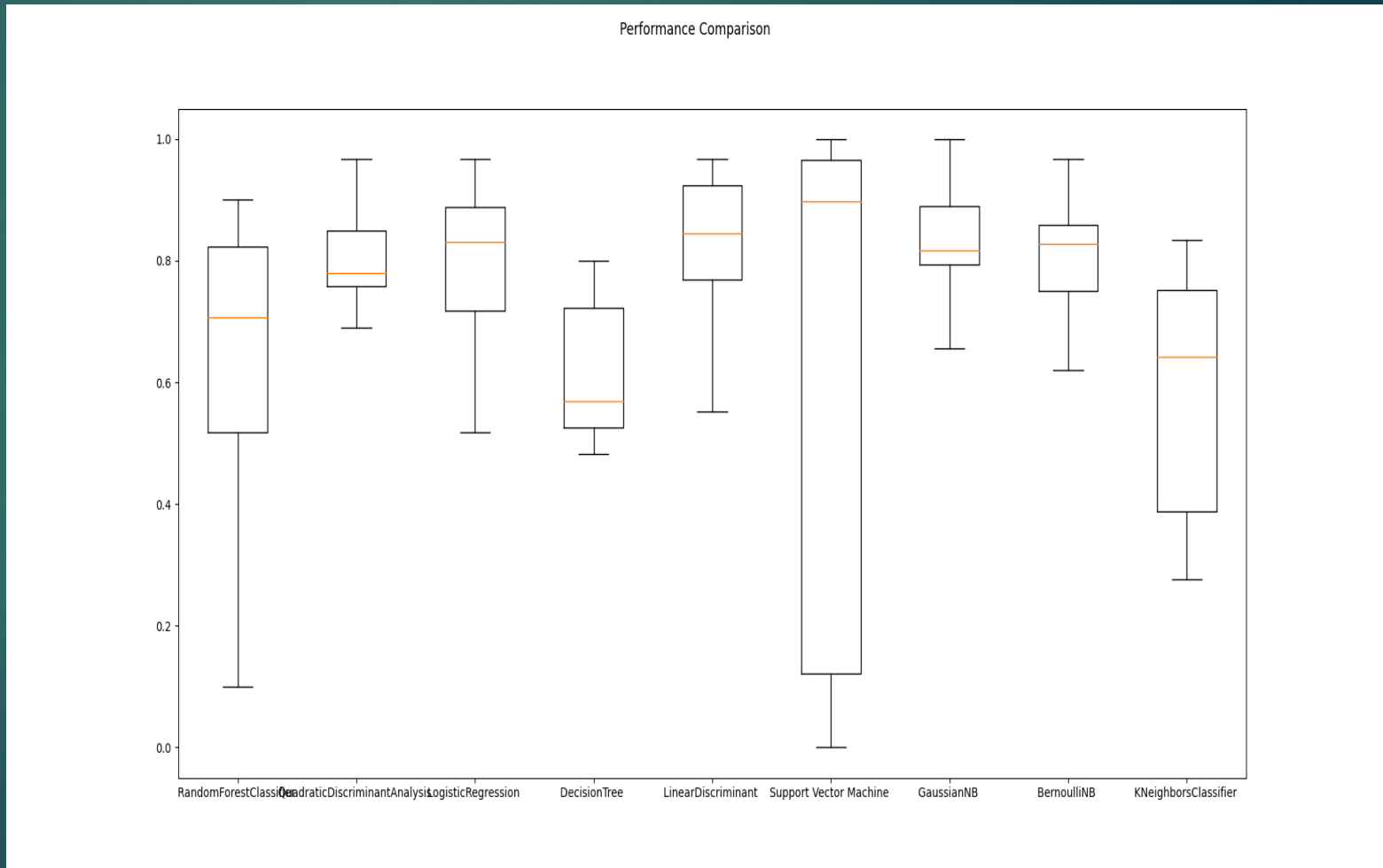
Techniques used in project

- Data observation

- Data cleaning

- Data formatting

- Data standardization

- Spot cheeking techniques

- Concept of pipeline
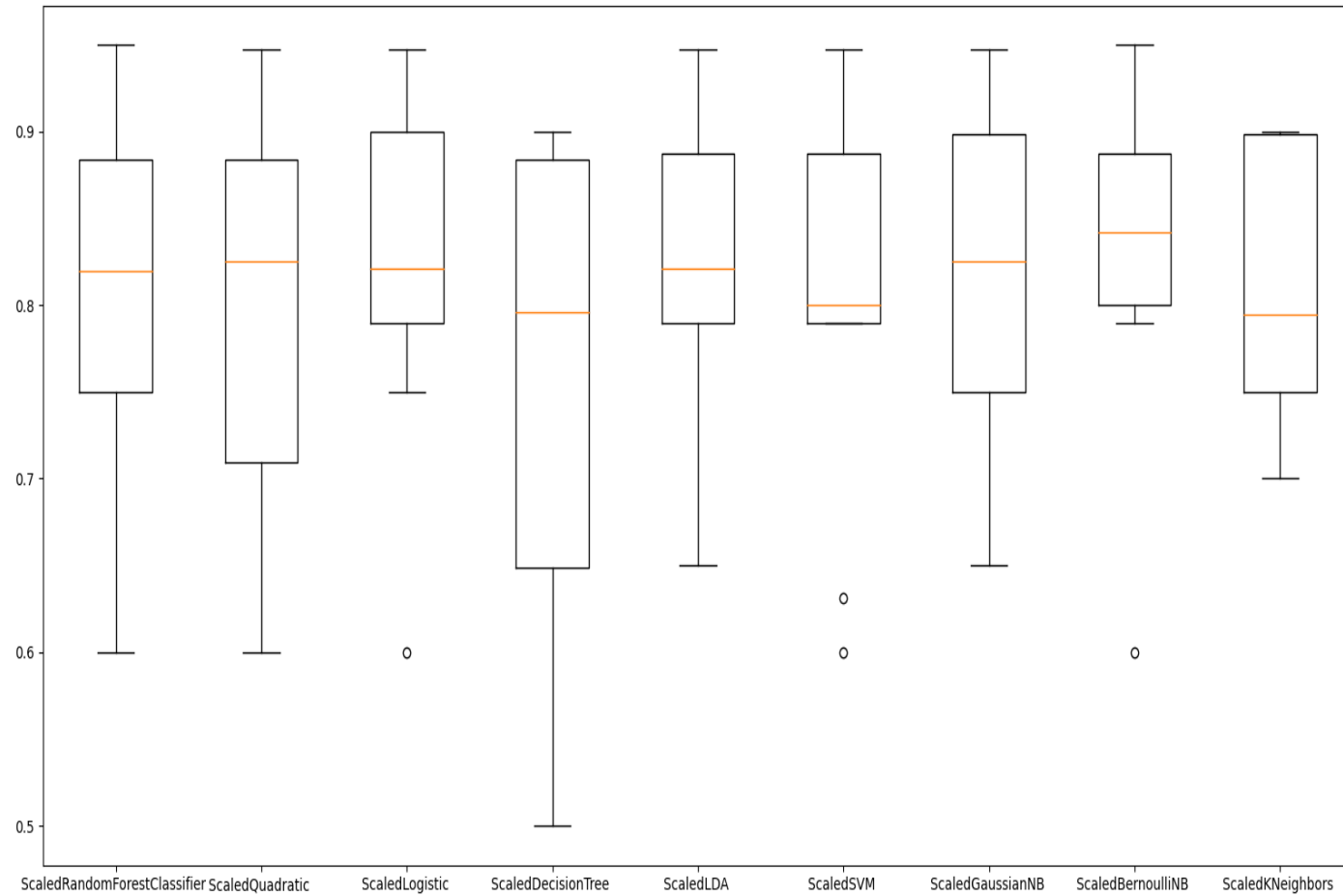
Challenges That I faced

- Prepare data for training
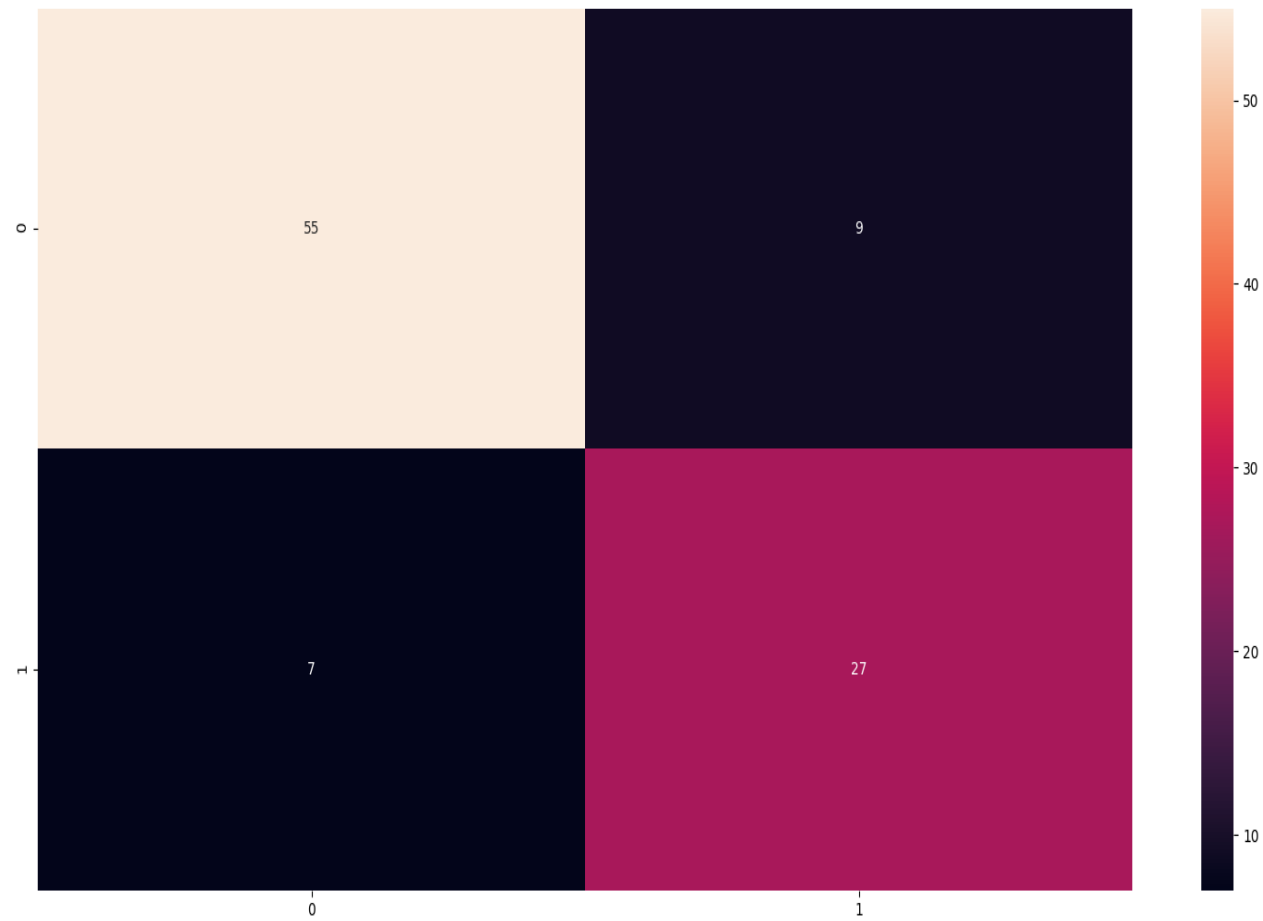- How to increase accuracy and decrease training time.

Some picture

Accuracy of algorithm without standardization.

Performance Comparison

Increase in accuracy after standerization.

Final confusion matrix

THANK YOU