SIX WEEKS SUMMER TRAINING

ON

Data Science with Machine Learning

PROJECT NAME: BREAST CANCER DETECTION

Submitted By

Rishav Rapta

Registration no: 11700946

Programme name: Computer Science and Engineering

Under the guidance of

Teach Tech Services

School of Computer Science and Engineering

Lovely Professional University, Phagwara

(June-July, 2019)

Index:

Declaration

I hereby declare that I have completed my six weeks summer training at Teach Tech Services .From 10th June 2019 to 25th July 2019 under the guidance of Mr. Sai Saurab sir (Faculty – IIT Kanpur).I have declare that I have worked with full dedication during these six weeks of training and my learning outcomes fulfill the requirements of training for the award of degree of Machine Learning with Python.

Name: Rishav Rapta

Reg no: 11700946

TEACH-TECH SERVICES

iSMRITI
Industry Experts & Faculties from
IIT-KANPUR

Certificate

OF COMPLETION

THIS IS TO CERTIFY THAT

QUALITY ASSURED
ISO 9001:2015 CERTIFIED

Rishav Rapta

has successfully completed the

Artificial Intelligence with Machine Learning and Deep Learning

conducted between **10 June, 2019 to 25 July, 2019** at **Lovely Professional University.**

Verification ID : TTS18830

https://teachtechin.com/certificate/verify.php

Authorised Signature
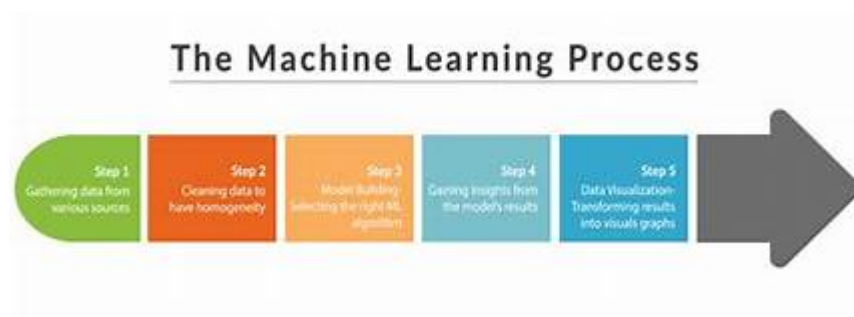
1. **Introduction**

    Python:
    *   As we all know python is a fastest growing programming language.
    *   Python is a multi – purpose language, High – level language, Huge Community, Cross – platform, Large Ecosystem(Libraries and Tools)
    *   The reason for using python is we can solve complex problems in less time with fewer lines of code.
    *   Python is a case – sensitive language.
    *   We can use python in following items :
        1. Data Analysis
        2. Artificial Intelligence /Machine Learning
        3. Automation.
        4. Web Applications.
        5. Mobile Applications.
        6. Desktop Applications.
        7. Testing.
   8. Hacking.

2. **Machine Learning:**

    In machine learning, computers apply statistical learning techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

The Machine Learning Process

Step 1 — Gathering data from various sources

Step 2 — Cleaning data to have homogenity

Step 3 — Model building Selecting the right ML algorithm

Step 4 — Gaining insights from the model's results

Step 5 — Data Visualization-Transforming results into visuals graphs

- This report mainly gives you all about the technology that I have learnt in my summer training which held in the past six weeks. It mainly gives you all an idea what the technology mainly and how it is useful in the day to day life. And you all will know in more detail about its Applications, implementations and also about its scope in the real world software development. And the technology that I have learnt is "Data Science with Python". You will know this in detail in upcoming pages.

Machine Learning Flow:
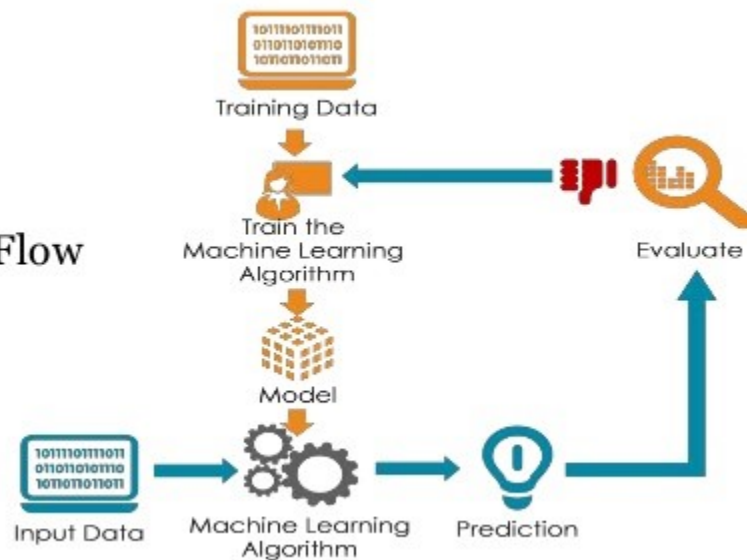
3: SOFTWARE REQUIREMENT ANALYSIS

- We require some sort of deep knowledge about python concepts and methodologies for the better implementations. We need JUPYTER Software for the project to test and implement for the result.

- We need Jupyter note book for writing the code and compiling it.

DIFFERENT LIBRARIES:

Pandas:

- Pandas is an open source, high performance, easy to use data structures and data analysis tools for the Python programming language.

- Pandas series is a one

-

- -dimensional labelled array capable of holding data of any type (int, str, float, etc..)

- The axis labels are collectively called index.

- Generally we use "pd" as aobject for declaring pandas.

- If we want to use the pandas, first we need to import it by "import" keyword.

Data Visualization using Matplotlib:

- Matplotlib is a plotting library for Python.

- Conventionally, the package is imported into the Python script by adding the following statement –

-        - frommatplotlib import pyplot as plt.

- Here, pyplot () is most important function in matplotlib library, which is used to plot 2D data.
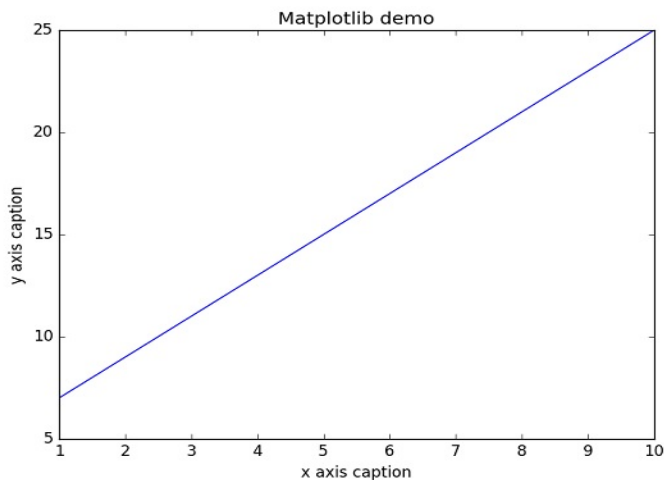
Numpy:

- Numpy is not another programming language but a python extension module.
- It provides fast and efficient operation on arrays of homogeneous data.
- Numpy extends python into high-level language for manipulating numerical data, similar to MATLAB.
- Numpy is a package in Python used to Scientific Computing.
- Numpy package is used to perform different operations.
- The ndarray (Numpy Array) is a multidimensional array used to store values of same data type.

SciPy:

- Scipy, pronounced as Sigh Pi, is a scientific python open source, distributed under the BSD licensed library to perform Mathematical, Scientific and Engineering Computations.
- The SciPy library depends on Numpy, which provides convenient and fast N-dimensional array manipulation.
- Scipy is another Python library for researches, developers and data scientists. Do not get confused with the SciPy.
- Stack and library, It provides statistics, optimizations, integration jand linear algebra packages for computation.
- It is based on Numpy concept to deal with complex mathematical problems.
- It provides numerical routines for optimization and integration.
- It inherits varieties of sub-modules to choose from.
- If you have just started your data science career, SciPy can be very helpful to guide science career, SciPy can be very helpful to guide you throughout the whole numerical computations thing.

Applications of Tensor flow:

- It's a powerful machine learning framework

- It is used for deep learning.

- It helps classify and cluster data like that with sometimes superhuman accuracy.

Keras:

- Keras Applications are deep learning models that are made available alongside pre-trained weights. These models can be used for prediction, feature extraction, and fine-tuning.
- Weights are downloaded automatically when instantiating a model.
- Keras has broad adoption in the industry and the research community

MACHINE LEARNING TYPES:

SUPERVISED LEARNING

- **Supervised learning** is the machine 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. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples

UNSUPERVISED LEARNING

- **Unsupervised learning** is a type of self-organized Hebbian learning that helps find previously unknown patterns in data set without pre-existing labels. It is also known as self-organization and allows modeling probability densities of given inputs. It is one of the main three categories of machine learning, along with supervised and reinforcement learning. Semi-supervised learning has also been described, and is a hybridization of supervised and unsupervised techniques.

2: PROFILE OF THE PROBLEM

- Breast Cancer prediction from given raw data.
- Here we have display various sections by using different modules of the data science with python.
- Here we have to show and check whether we have tried to apply major concepts like Data cleaning, Data manipulation along with different techniques.
- We have to prepare a simple and effective regression technique for the result.

3: SOFTWARE REQUIREMENT ANALYSIS:

- We require some sort of deep knowledge about python concepts and methodologies for the better implementations.
- We need Python idle Software for the project to test and implement for the result.

4:TESTING & IMPLEMENTATION:

File   Edit   Shell   Debug   Options   Window   Help

```
Python 3.7.3 (v3.7.3:ef4ec6ed12, Mar 25 2019, 22:22:05) [MSC v.1916 64 bit (AMD64)] on w:
Type "help", "copyright", "credits" or "license()" for more information.
>>>
 RESTART: C:\Users\hp india\Desktop\machine learning\classifier on breast data.py
   clump  cell_size  cell_shape  ...  bland_chromnnorm_nuclei  mitoses classes
0      5          1           1  ...                        1        1       2
1      5          4           4  ...                        2        1       2
2      3          1           1  ...                        1        1       2
3      6          8           8  ...                        7        1       2
4      4          1           1  ...                        1        1       2

[5 rows x 10 columns]
```

*classifier k nearest lgorithm.py - C:\Users\hp india\Desktop\machine learning\classifier k nearest lgorithm.

File   Edit   Format   Run   Options   Window   Help

```python
import pandas as pd
from math import sqrt
import numpy as np
from collections import Counter
distance=[]
def knn(data,predict,k):
    for i in data:
        for s in data[i]:
            di=np.sqrt(np.sum((np.array(s)-np.array(predict))**2))
            distance.append([di,i])
    s= [j[1] for j in sorted(distance) [:k]]
    x=Counter(s).most_common(1)[0][0]
    return x


df=pd.read_csv("breast-cancer-wisconsin.data")
df.replace('?',-9999,inplace=True)
df=df.drop(['id'],1)

test_size=0.2
df=df.astype(float).values.tolist()
random.shuffle(df)
train_x=df[:-int(test_size*len(df))]
test_x=df[-int(test_size*len(df)):]
tr_set={2:[],4:[]}
test_set={2:[],4:[]}
c=0
t=0
for i in train_x:
    tr_set[ i[-1]].append(i[:-1])

for x in test_x:
    test_set[x[-1]].append(x[:-1])
for group in test_set:
    for data in test_set[group]:
        s=knn(tr_set,data,5)
        #print(group)
        if group==s:
            c+=1
        t+=1
print("accuracy",c/t)


103 139
accuracy 0.7410071942446043
>>>
```

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
import numpy as np
s=pd.read_csv('breast-cancer-wisconsin.data')
s.replace('?',-9999,inplace=True)
k=s.drop(['id'],1)
full_data=k.astype(float).values.tolist()
print(full_data)

s=s.drop(['id'],1)
inputs=np.array(s.drop(['classes'],1))
outputs=np.array(s['classes'])
x_train,x_test,y_train,y_test=train_test_split(inputs,outputs,test_size=0.2)
model=KNeighborsClassifier()
model=model.fit(x_train,y_train)
y=model.predict(x_test)
accuracy=model.score(x_test,y_test)
print(accuracy)
al=model.score(x_test,y)
print(al)
```

Python 3.7.3 Shell

File   Edit   Shell   Debug   Options   Window   Help

```
Python 3.7.3 (v3.7.3:ef4ec6ed12, Mar 25 2019, 22
4)] on win32
Type "help", "copyright", "credits" or "license(
>>>
 RESTART: C:\Users\hp india\Desktop\machine lear
y
```

Squeezed text (462 lines).

```
0.9714285714285714
1.0
>>>
```

TECHNOLOGY LEARNT:

Python with Machine Learning

- Machine Learning is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.
- To become a perfect machine Learning course, we must be perfect in some of the programming languages like Python, Perl, C/C++ and SQLare the required in roles of Machine Learning.
- Companies need to use data to run and grow their everyday business.
- Machine Learning is the study of to learning of the machine
- It involves developing methods of recording, storing, and analyzing data to effectively extract useful information.
- The goal of Machine Learning is to gain insights and knowledge from any type of data – that is both structured and unstructured.
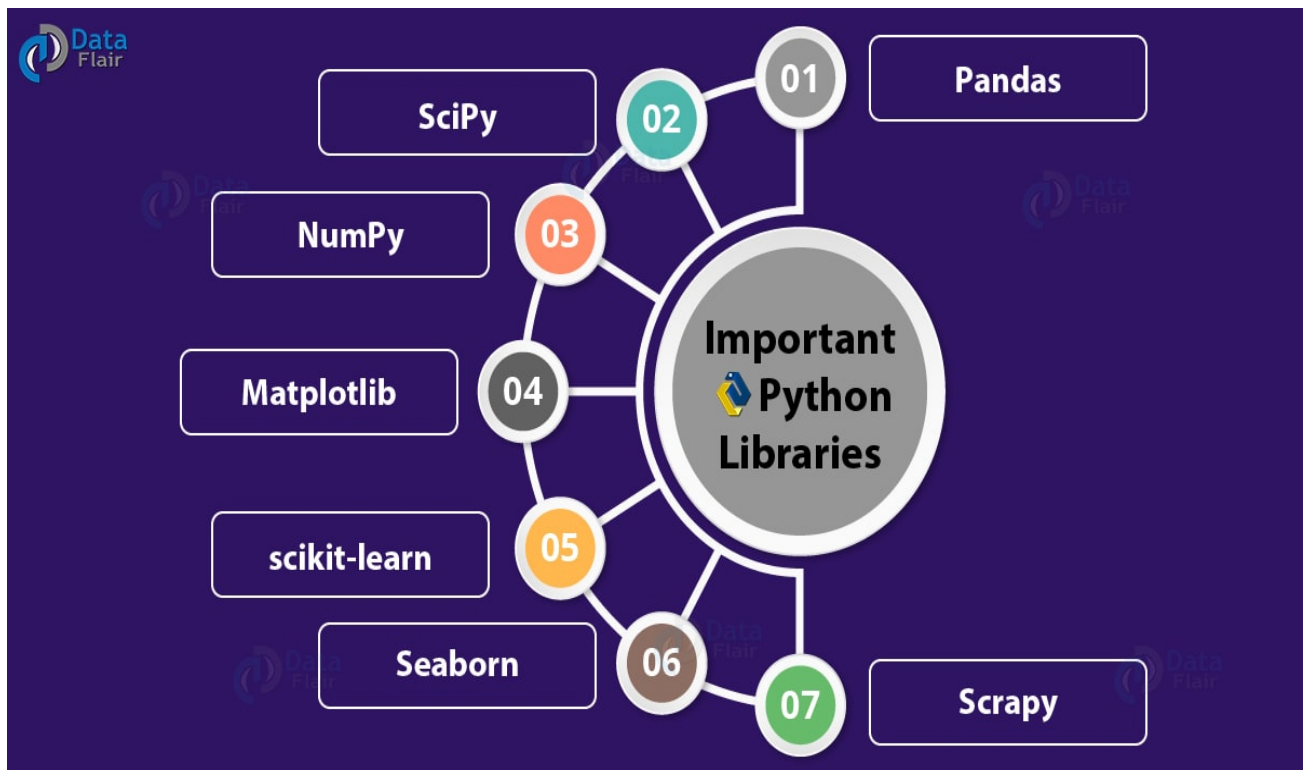
APPLICATIONS OF THE TECHNOLOGY:

Machine learning is the application of Artificial Intelligence which makes the computers to predict the outcomes automatically without the intervention of human beings.

Some of the various application used by the Machine Learning is:

- **Financial Services**

- **Virtual Personal Assistants**

- **Marketing and Sales**

- **Predictions while Traveling**

- **Healthcare**

LEARNING OUTCOMES FROM THE TECHNOLOGY:

1. Improve Adaptive Learning

2. Better Parent Engagement.

3. Better Organization.

4. Student Recruitment.

CONCLUSION:

- To implement cutting edge machine learning algorithms to solve real world problems.

- We have carefully selected the projects which will cover important aspect of Machine learning such as Supervised Learning, Unsupervised learning and Neural network with deep learning.

- You will start with real world data available publicly to create these Machine Learnings Projects

BIBLIOGRAPHY:

https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_understanding_data_with_visualization.htm

https://www.listendata.com/p/statistics-tutorials.html

https://en.wikipedia.org/wiki/Library(computing)