Shri Vaishnav Vidhyapeeth Vishwavidhyalaya

"DATA SCIENCE USING PYTHON "

Training Project report
on

"Disease Prediction by Symptoms using GUI"

Submitted By-:

Sakshi Shukla[1710DMTCSE01307]

Anunay Sahu[1904MCA0006481]

Radhika Sharma[1904MCA0006493]

Vipul Jain[1804MCA0004590]

COMPUTER SCIENCE & ENGINEERING IN
BACHELOR OF TECHNOLOGY

2020-2021

A Report of 3 Weeks Industrial Training

at

. **“WebTek Labs Pvt. Ltd.”**

Industrial Training report submitted in partial fulfillment Of the degree of

BACHELOR OF TECHNOLOGY
COMPUTER SCIENCE & ENGINEERING

Submitted By-:

Sakshi Shukla[1710DMTCSE01307]

Anunay Sahu[1904MCA0006481]

Radhika Sharma[1904MCA0006493]

Vipul Jain[1804MCA0004590]

Department of Information Technology

Shri Vaishnav Institute of Information Technology

Shri Vaishnav Vidyapeeth Vishwavidyalaya

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

DECLARATION

We hereby declare that we have undertaken industrial training at “**WEBTEK LABS PVT. LTD.**” during a period from 2**7th July to 21st August** in partial fulfilment of requirements for the award of degree of B.Tech (COMPUTER SCIENCE & ENGINEERING) at SHRI VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY INDORE. The work which is being presented in the training report submitted to Department of COMPUTER SCIENCE & ENGINEERING at VAISHNAV INSTITUTE OF INFORMATION TECHNOLOGY INDORE is an authentic record of training work.

Students Name-:

Sakshi Shukla[1710DMTCSE01307]

Anunay Sahu[1904MCA0006481]

Radhika Sharma[1904MCA0006493]

Vipul Jain[1804MCA0004590]

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

ACKNOWLEDGEMENT

After the completion of Industrial Training, words are not enough to express my feelings about all those who helped us to reach my goal, feeling above this is my indebtedness to The Almighty for providing me this moment in life.

It’s a great pleasure and moment of immense satisfaction for me to express my profound gratitude to MS. **Mousita Dhar** Webtek Labs Pvt. Ltd. whose constant encouragement enable us to learn and work enthusiastically. Their perpetual motivation, patience and excellent expertise and guidance in discussion during the training period, have benefited us to an extent, which is beyond expression. Working under their guidance has been a fruitful and unforgettable experience. She has helped us to accomplish the challenging task in a very short period of time.

we also express my sincere thanks and gratitude to **Dr. Anand Rajawat** Prof. & Head, Computer Science Department, SVIIT, SVVV, Indore, for providing all the necessary facilities and true encouraging environment to bring out the best of my endeavors.

Finally, we express the constant support of our friends, family and professors for inspiring us throughout and encouraging us.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that Sakshi Shukla, Anunay Sahu,Radhika Sharma and Vipul Jain has satisfactorily completed the Industrial Training “Data Science using Python ”. towards the partial fulfilment of the degree in Bachelor of Technology (Computer Science) awarded by Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore during the academic year 2020-2021.

Training In-Charge
Ms. Mousita Dhar

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

ABSTRACT

This Report Introduces the process of creating Machine Learning Model Prediction using python . Python is a general purpose programming language , hence can be used for Data Science Analysis.Python is designed with a feature to facilitate data analysis and visualisation. Python provides various Libraries to work efficiently one such and mostly used library is Sklearn ,Numpy,Pandas etc.Disease Prediction by Symptoms using GUI is one such project with has been deployed using the language python

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

TABLE OF CONTENT

Introduction about the company

WebTek Labs Pvt. Ltd. is recognized as a leading IT solution providing organization with a dynamic and fast growing team of diversely talented individuals. Incorporated in 2001, in our aim to provide the best talent, we initially started with Recruitment & Staffing services. We paralleled this by providing knowledge and skill development certification training programs. WebTek Certified Tester (WCT) Program that aims to provide IT companies trained software Testers has reached soaring heights of recognition over the years. Few years later after its inception,

Having partnered and worked with some of the leading names across Education, IT, ITES, Banking, Insurance, Aviation, Retail, Healthcare, Hospitality, Media, Manufacturing and FMCG sectors, WebTek Labs has explored business opportunities in software solutions with the Government, Corporate and Institutes.

With over a decade of experience we create and deliver high-impact solutions, enabling our clients to achieve their business goals and enhance their competitiveness. In our pursuit of excellence, WebTek's Research & Development team consistently innovates to provide up-to-date solutions keeping in pace with changing times. Our mission is for businesses to leverage the internet and mobility to work smarter and grow faster.

Overview of company

VISION

Is to lead in the creation and delivery of innovative solutions and services that enable our clients to win in the changing world of work.

TEAM

Has expertise ranging from design to development, training to placements. We combine this knowledge with proactive thinking and strategic planning to approach new challenges with your overall business objectives in mind. WebTek Lab's management team brings together a wealth of experience in both technological and organizational development that is critical in helping our customers achieve their goals.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

Introduction to python

What is Python?

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a functional way.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

FEATURES OF PYTHON:

Easy 01
Expressive 02
Free and Open Source 03
High-Level 04
Portable 05
Object Oriented 06
Extensible 07
Embeddable 08
Interpreted 09
Large Standard Library 10
GUI Programming 11
Dynamically Typed 12
Features of Python
Data Flair

1. *Easy:*

When we say the word ‘easy’, we mean it in different contexts.

- Easy to code: As we have seen in earlier lessons, Python is very easy to code. Compared to other popular languages like Java and C++, it is easier to code in Python. Anyone can learn python syntax in just a few hours. Though sure, mastering Python requires learning about all its advanced concepts and packages and modules. That takes time. Thus, it is programmer-friendly

- Easy to read: Being a high-level language, Python code is quite like English.

2.Free andOpen-Source

Firstly, Python is freely available. Secondly, it is open-source. This means that its source code is available to the public. You can download it, change it, use it, and distribute it. This is called FLOSS (Free/Liber and Open Source Software). As the Python community, we’re all headed toward one goal- an ever-bettering Python.

2. High- Level

It is a high-level language. This means that as programmers, we don’t need to remember the system architecture. Nor do we need to manage the memory. This makes it more programmer-friendly and is one of the key python features.

3. Portable

Let’s assume we have written a Python code for our Windows machine. In other words, we can take one code and run it on any machine, there is no need to write different code for different machines. This makes Python a portable language.

4. Interpreted

In Python, there is no need to compile it. Internally, its source code is converted into an immediate form called byte code,

5. *Object-Oriented*

A programming language that can model the real world is said to be object-oriented. It focuses on objects, and combines data and functions. Contrarily, a procedure-oriented language revolves around functions, which are code that can be reused. Python supports both procedure-oriented and object-oriented programming which is one of the key python features. It also supports multiple inheritance, unlike Java. A class is a blueprint for such an object. It isan abstract data type, and holds no values.

6. Large StandardLibrary

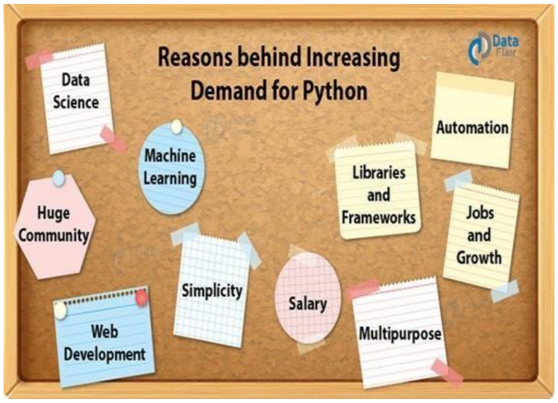
Python downloads with a large library that we can use so we don't have to write your own code for every single thing. There are libraries for regular expressions, documentation- generation, unit-testing, web browsers, threading, databases, CGI, email, image manipulation, and a lot of other functionality.

7. GUI Programming

You can use Toolkits to create basic GUIs.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

DEMAND OF PYTHON



1. *Data Science*

Python holds a special place in the hearts of Data Scientists compared to any available language, such as R or C++. Data Science is all about dealing with data at huge amounts (Big Data). Hence with simple usage and a large set of libraries and frameworks, Python has become the most promising option to handle it! e.g. PyBrain, PyMySQL, and NumPy are one of the big reasons. Another step forward is because of Python’s easy integration with other programming languages, making it more scalable and future-oriented.

1. Machine Learning

Python being an interpreted language makes it comprehensive enough for the language to be interpreted by virtual machine against any other machine language which is what the hardware understands. It can even be used in complicated scenarios by making use of variables, objects, complex arithmetic or Boolean expressions and other concepts to make its demand and usability increase exponentially. Even the growth of machine learning has been phenomenal in the last couple of years and is rapidly changing everything around us. Algorithms are becoming sophisticated every day.

2. Web Development

While concerning backend development, Python is chosen by 2 out of 3 developers who initially worked with PHP. Python’s trend has been witnessing a steep upward in the past two years as it is serving as a better alternative. It offers so many good libraries and frameworks, e.g. Flask and Django, which make web development easy. After adopting Python some of the product based platforms have become the biggest names – YouTube, Instagram, Facebook, Google, Netflix, and Spotify. Considering the general perception towards python, in web development Python creates more robust code that can be used to form versatile use cases.

3. Simplicity

Python is readable as well as simple. It’s even easy to set up Python; There’s nothing like class path problems like that in Java and compiler issues present in C++. Just install it and run it!

1. *Libraries and Frameworks*

Not just a framework but it has got “superior framework”. Django is the clear winner when *python’s frameworks* are considered. Frameworks are easy to use, secure and fast. Mostly developers prefer these frameworks, as their use results in shorter development time and ease of setup. The richer the framework, deeper one should learn as it will translate into a lot of long-term valuable results.

2. *Automation*

You are required to write test scripts to automate tasks, that’s where Python comes into existence in automation. You’ll be impressed with the time and the number of lines required to write codes for tools. As python supports with lots of tools and modules, it makes things easier and even highlight the power of Python. With basic python codes, one can reach the advanced level of automation easily. Software testing is one of the tedious tasks in automation and python becomes its performance booster!

3. *Multipurpose*

Swiss Army knife-like nature describes well the overall work of python. It’s not tied to just one discipline, you can do many things. You don’t really need to fetch data from a SQL server or a MongoDB database; Python supports all these sources of data with very clean syntax use. Python API called PySpark can be used to distribute computing. It also provides support for Natural language processing through NLTK.

Python Applications

1. Applications of Python Programming in Desktop GUI

Most binary distributions of Python ship with Tk, a standard GUI library. It lets you draft a user interface for an application. Apart from that, some toolkits are available:

- wxWidgets
- Kivy – for writing multitouch applications
- Qt via pyqt or pyside

2. Science and NumericApplications

This is one of the very common applications of python programming. With its power, it comes as no surprise that python finds its place in the scientific community. For this, we have:

- SciPy – A collection of packages for mathematics, science, and engineering.
- Pandas- A data-analysis and -modelling library
- Also, NumPy lets us deal with complex numerical calculations.

3. Database Access

With Python, you have:

- Custom and ODBC interfaces to MySQL, Oracle, PostgreSQL, MS SQL Server, and others. These are freely available for download.
- Object databases like Durus and ZODB

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

IDLE

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, which are both not free.

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python. There is also a small, bootstrap version of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

Packages

Numpy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays..

NumPy is the fundamental package for scientific computing with Python. It contains:

- Powerful N-dimensional array objects
- Tools for integrating C/C++, and Fortran code
- It has useful linear algebra, Fourier transform, and random number capabilities

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

Scipy

SciPy is a free and open-source Python library used for scientific computing and technical computing.

SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.

.

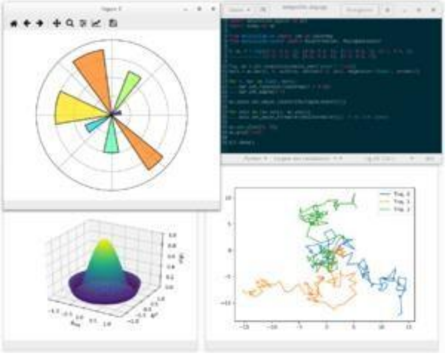
As the name suggests, it is a scientific library that includes some special functions:

- It currently supports special functions, integration, ordinary differential equation (ODE) solvers, gradient optimization, and others
- It has fully-featured versions of the linear algebra modules
- It is built on top of NumPy

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.



Scikit-learn

Scikit-learn (formerly **scikits.learn** and also known as **sklearn**) is a free software machine learning library for the Python programming language It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-learn provides machine learning libraries for python.Some of the features of Scikit-learn includes:

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

Pandas

In computer programming, **pandas** is a software library written for the Pytho programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
Pandas is used for structured data operations and manipulations.

- The most useful data analysis library in Python
- Instrumental in increasing the use of Python in the data science community
- Used extensively for data mugging and preparation

Next, in our learning of Data Science with Python let us learn the exploratory analysis using Pandas.

Seaborn

Seaborn is a library for making statistical graphics in Python. It is built on top of matplotlib and closely integrated with pandas data structures.

Here is some of the functionality that seaborn offers:

- A dataset-oriented API for examining relationships between multiple variables
- Specialized support for using categorical variables to show observations or aggregate statistics
- Automatic estimation and plotting of linear regression models for different kinds dependent variables

Seaborn aims to make visualization a central part of exploring and understanding data. Its dataset-oriented plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

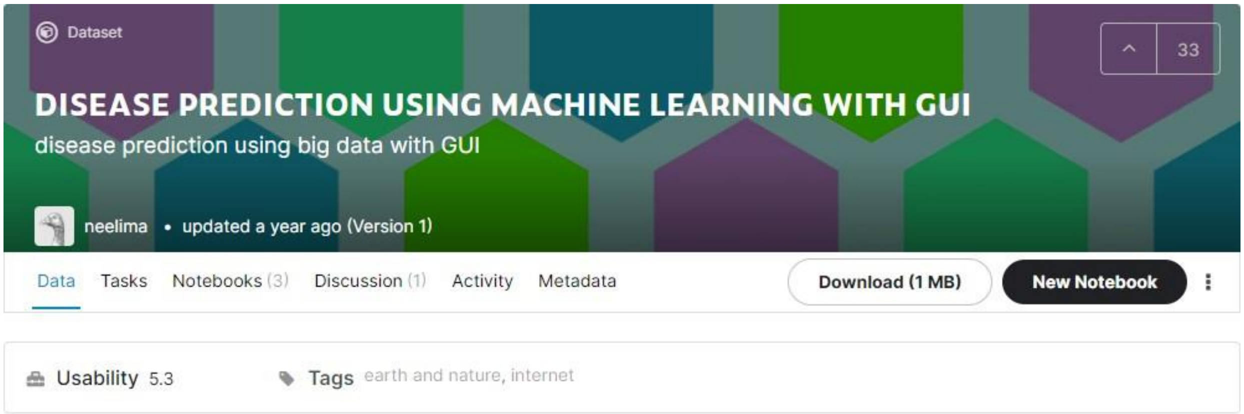
Data Science

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining. Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

Turing award winner JiGray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge. When Harvard Business Review called it "The Sexiest Job of the 21st Century" the term became a buzzword, and is now often applied to business analytics, business intelligence, predictive modeling, or any arbitrary use of data, or used as a glamorized term for statistics. In many cases, earlier approaches and solutions are now simply rebranded as "data science" to be more attractive, which can cause the term to become "dilute beyond usefulness." While many university programs now offer a data science degree, there exists no consensus on a definition or suitable curriculum contents. Because of the current popularity of this term, there are many "advocacy efforts" surrounding the field. To its discredit, however, many data science and big data projects fail to deliver useful results, often as a result of poor management and utilization of resources.

PROJECT-ANALYSIS

1) Dataset and its Description



DISEASE PREDICTION USING MACHINE LEARNING WITH GUI

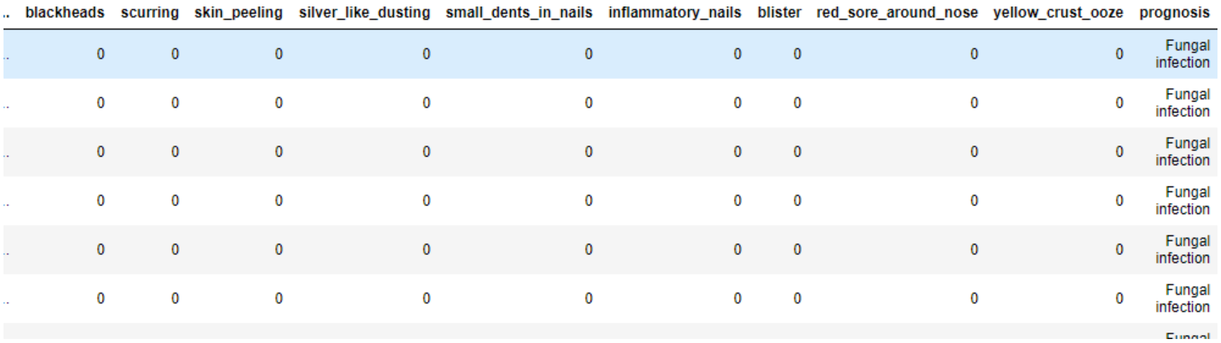
disease prediction using big data with GUI

neelima • updated a year ago (Version 1)

Due to big data progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Moreover, different regions exhibit unique appearances of certain regional diseases, which may results in weakening the prediction of disease outbreaks. In this project, it bid a Machine learning Decision tree map, Random forest algorithm by using structured and unstructured data from hospital. It also uses Machine learning algorithm for partitioning the data. To the highest of gen, none of the current work attentive on together data types in the zone of remedial big data analytics. Compared to several typical calculating algorithms, the scheming accuracy of our proposed algorithm reaches 100% with an regular speed which is quicker than that of the unimodal disease risk prediction algorithm and produces report.

2) *Target coloumn –prognosis and other features(symptoms)*

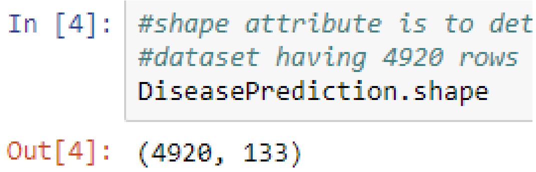
The target of a supervised model is a special kind of attribute. The target column in the training data contains the historical values used to train the model. The target column in the test data contains the historical values to which the predictions are compared. The act of scoring produces a prediction for the target.



..	blackheads	scurring	skin_peeling	silver_like_dusting	small_dents_in_nails	inflammatory_nails	blister	red_sore_around_nose	yellow_crust_ooze	prognosis
..	0	0	0	0	0	0	0	0	0	Fungal infection
..	0	0	0	0	0	0	0	0	0	Fungal infection
..	0	0	0	0	0	0	0	0	0	Fungal infection
..	0	0	0	0	0	0	0	0	0	Fungal infection
..	0	0	0	0	0	0	0	0	0	Fungal infection
..	0	0	0	0	0	0	0	0	0	Fungal infection

3) *Shape of Dataset*

Returns tuple of shape (Rows, columns) of dataframe/series



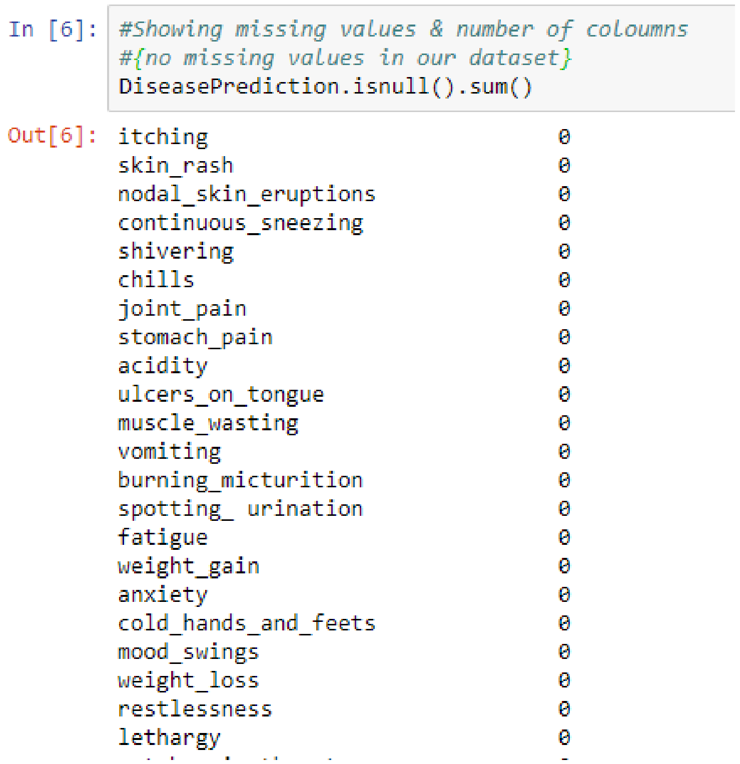
```
In [4]: #shape attribute is to det
        #dataset having 4920 rows
        DiseasePrediction.shape

Out[4]: (4920, 133)
```

4) *Finding Missing Values in Dataset*

Data can have missing values for a number of reasons such as observations that were not recorded and data corruption.

Handling missing data is important as many machine learning algorithms do not support data with missing values.

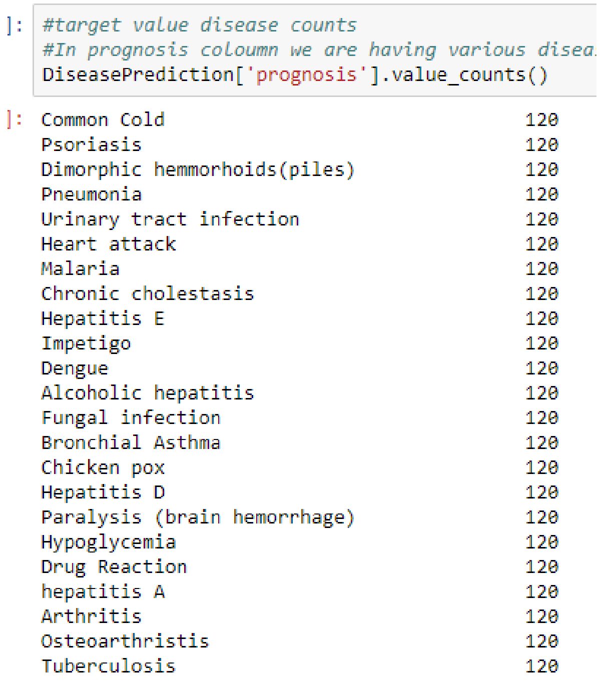


```
In [6]:  #Showing missing values & number of coloumns
          #{no missing values in our dataset}
          DiseasePrediction.isnull().sum()

Out[6]:  itching                           0
         skin_rash                         0
         nodal_skin_eruptions              0
         continuous_sneezing               0
         shivering                         0
         chills                            0
         joint_pain                        0
         stomach_pain                      0
         acidity                           0
         ulcers_on_tongue                  0
         muscle_wasting                    0
         vomiting                          0
         burning_micturition               0
         spotting_ urination               0
         fatigue                           0
         weight_gain                       0
         anxiety                           0
         cold_hands_and_feets              0
         mood_swings                       0
         weight_loss                       0
         restlessness                      0
         lethargy                          0
```

5) *Total Counts of Symptoms of target disease*

Counting the number of rows in a Pandas DataFrame determines how many rows exist in the DataFrame.

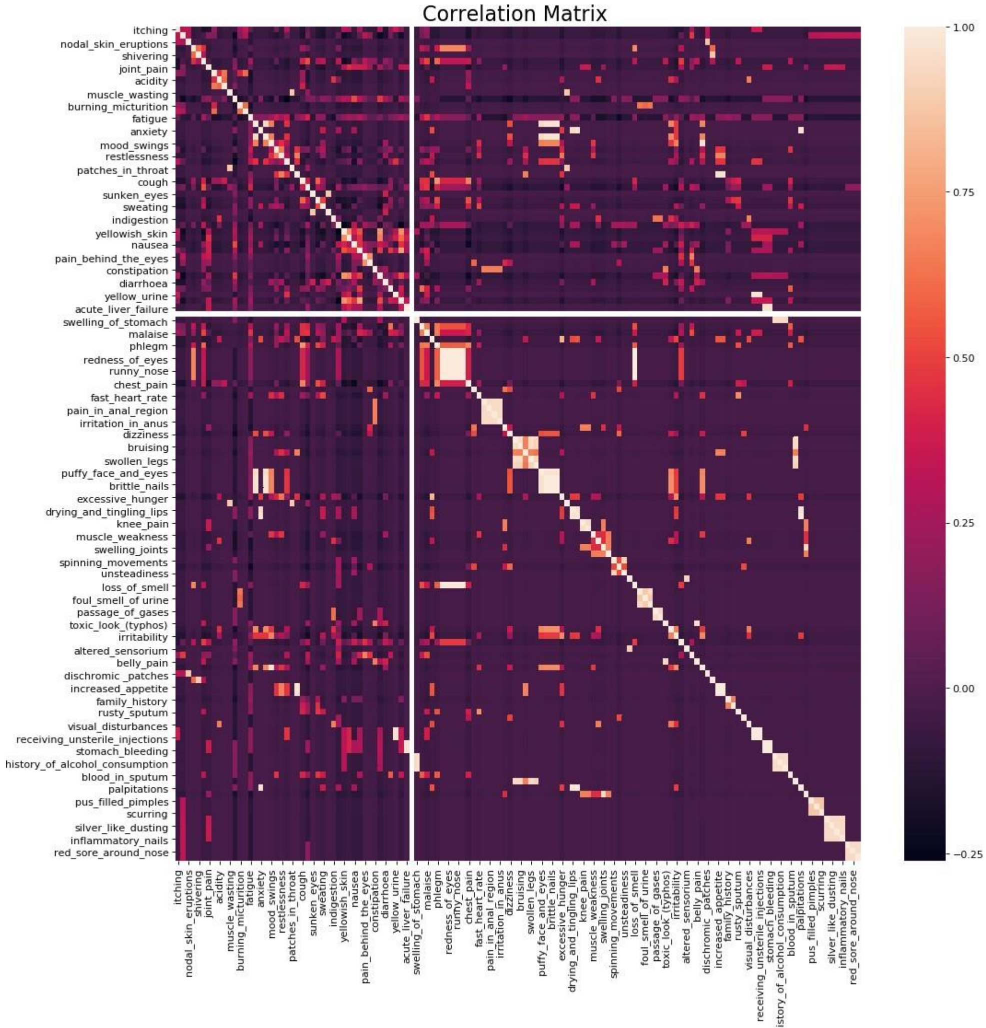


```
]: #target value disease counts
   #In prognosis coloumn we are having various disea
   DiseasePrediction['prognosis'].value_counts()
```

```
]: Common Cold                                120
   Psoriasis                                  120
   Dimorphic hemmorhoids(piles)               120
   Pneumonia                                  120
   Urinary tract infection                    120
   Heart attack                               120
   Malaria                                    120
   Chronic cholestasis                        120
   Hepatitis E                                120
   Impetigo                                   120
   Dengue                                     120
   Alcoholic hepatitis                        120
   Fungal infection                           120
   Bronchial Asthma                           120
   Chicken pox                                120
   Hepatitis D                                120
   Paralysis (brain hemorrhage)               120
   Hypoglycemia                               120
   Drug Reaction                              120
   hepatitis A                                120
   Arthritis                                  120
   Osteoarthristis                            120
   Tuberculosis                               120
```

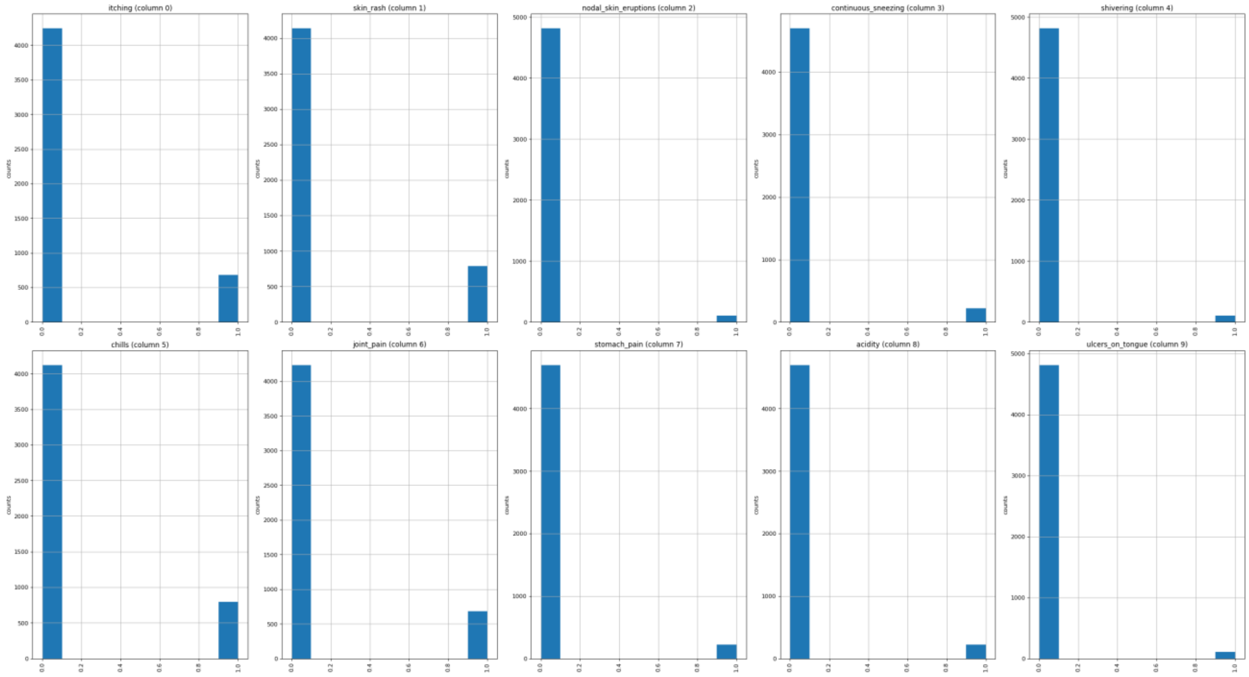
5) *Heat-Map*

To find co-relation among symptoms



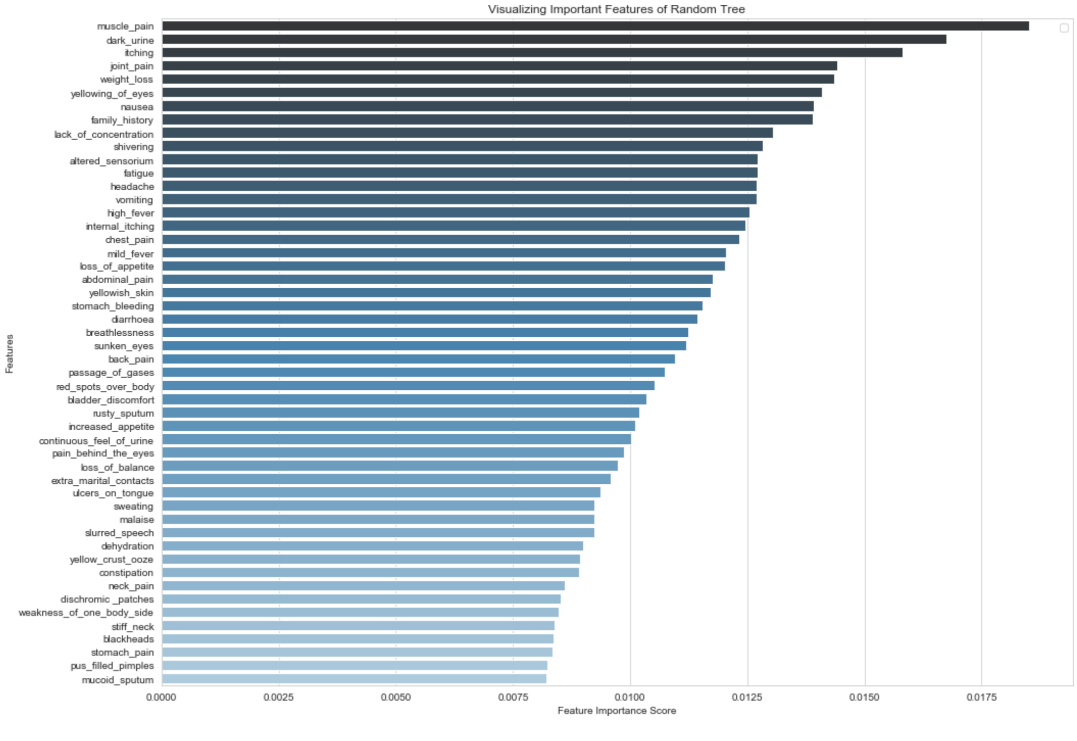
6) *Per coloumn distribution*

It shows the coloumn distribution of symptoms 0 and 1



7) *Visualizing—Random Forest importance features in bar graph*

Random forests are among the most popular machine learning methods thanks to their relatively good accuracy, robustness and ease of use. They also provide two straightforward methods for feature selection: mean decrease impurity and mean decrease accuracy.



Random Forests

Random forests (RF) construct many individual decision trees at training. Predictions from all trees are pooled to make the final prediction; the mode of the classes for classification or the mean prediction for regression. As they use a collection of results to make a final decision, they are referred to as Ensemble techniques.

Feature Importance

Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. *The higher the value the more important the feature.*

Implementation in Scikit-learn

For each decision tree, Scikit-learn calculates a nodes importance using Gini Importance, assuming only two child nodes (binary tree):

$$ni_j = w_jC_j - \ w_{left(j)}C_{left(j)} - \ w_{right(j)}C_{right(j)}$$

- ni sub(j)= the importance of node j
- w sub(j) = weighted number of samples reaching node j
- C sub(j)= the impurity value of node j
- left(j) = child node from left split on node j
- right(j) = child node from right split on node j

sub() is being used as subscript isn’t available in Medium

The importance for each feature on a decision tree is then calculated as:

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k}$$

- fi sub(i)= the importance of feature i
- ni sub(j)= the importance of node j

These can then be normalized to a value between 0 and 1 by dividing by the sum of all feature importance values:

$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j}$$

The final feature importance, at the Random Forest level, is it's average over all the trees. The sum of the feature's importance value on each trees is calculated and divided by the total number of trees:

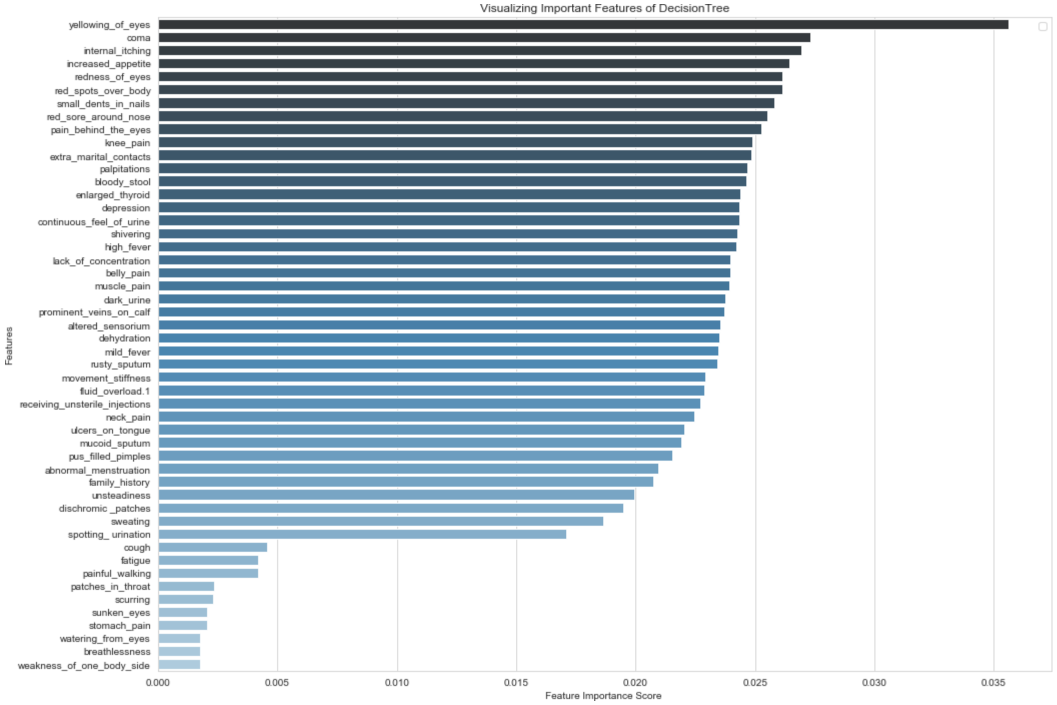
$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

- RFfi sub(i)= the importance of feature i calculated from all trees in the Random Forest model

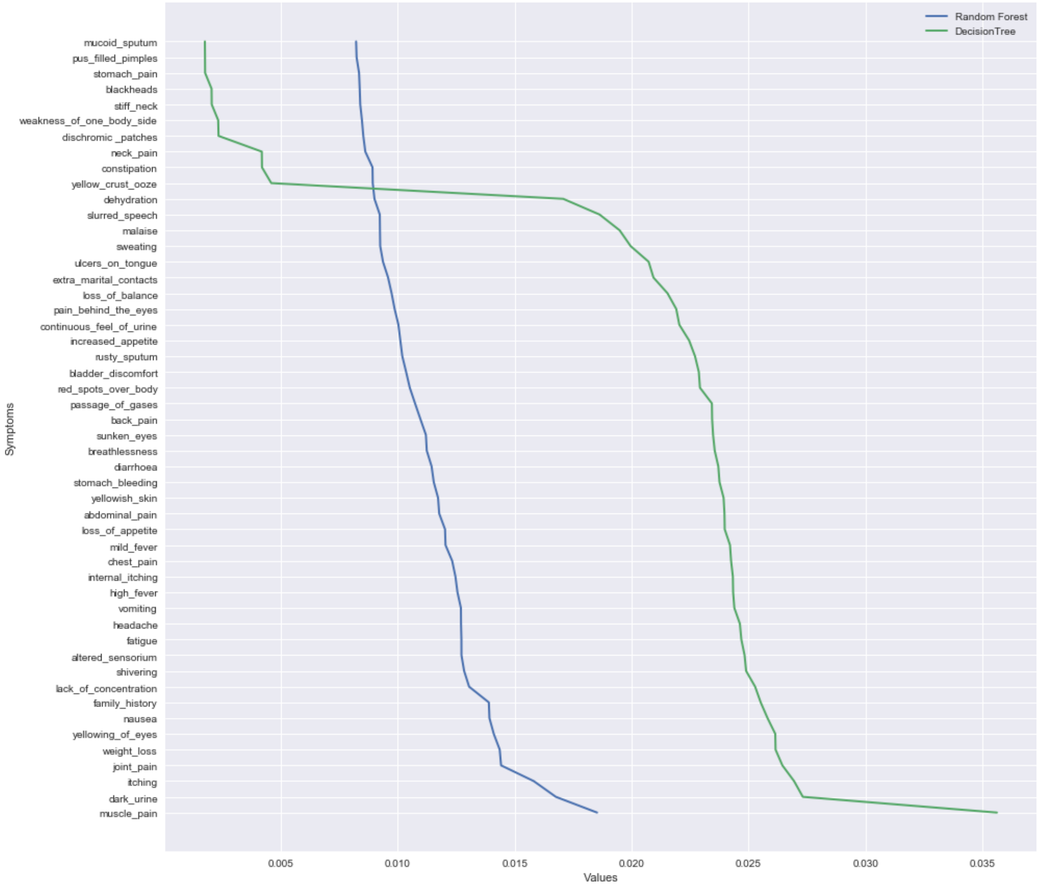
- normfi sub(ij)= the normalized feature importance for i in tree j

- T = total number of trees

8) *Visualizing—Decision Tree importance features in bar graph*

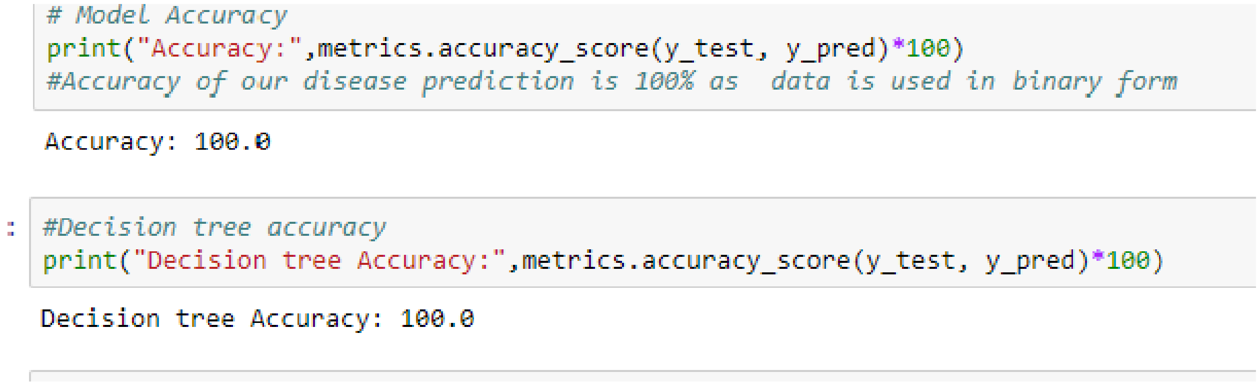


9) Comparison of Random Forest and Decision Tree



Random forest shows better results as compared with Decision tree.

ACCURACY OF THE MODEL



```
# Model Accuracy
print("Accuracy:",metrics.accuracy_score(y_test, y_pred)*100)
#Accuracy of our disease prediction is 100% as  data is used in binary form

Accuracy: 100.0
```

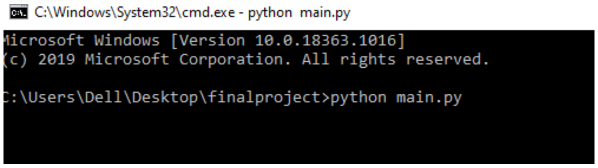
```
#Decision tree accuracy
print("Decision tree Accuracy:",metrics.accuracy_score(y_test, y_pred)*100)

Decision tree Accuracy: 100.0
```

100% as our dataset contains symptoms in binary form and test dataset and same kind of dataset so it predicts the perfect disease according to dataset

SHRI VAISHNAV INSTITUE OF INFORMATION TECHNOLOGY

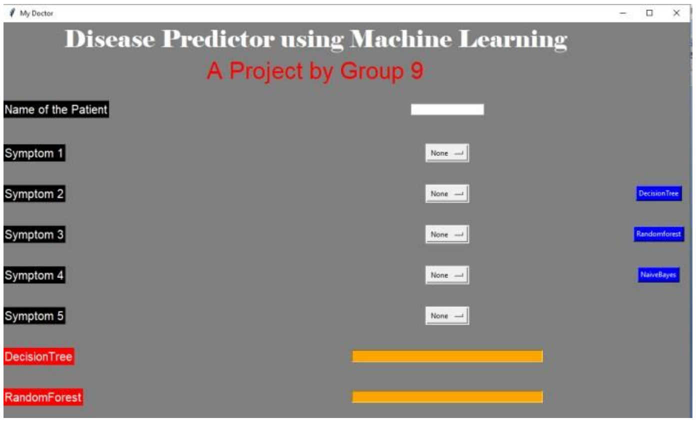
GUI Interface Of Disease Prediction



C:\Windows\System32\cmd.exe - python main.py

```
icrosoft Windows [Version 10.0.18363.1016]
(c) 2019 Microsoft Corporation. All rights reserved.

:\Users\Dell\Desktop\finalproject>python main.py
```



My Doctor

Disease Predictor using Machine Learning

A Project by Group 9

Name of the Patient

Symptom 1 None

Symptom 2 None

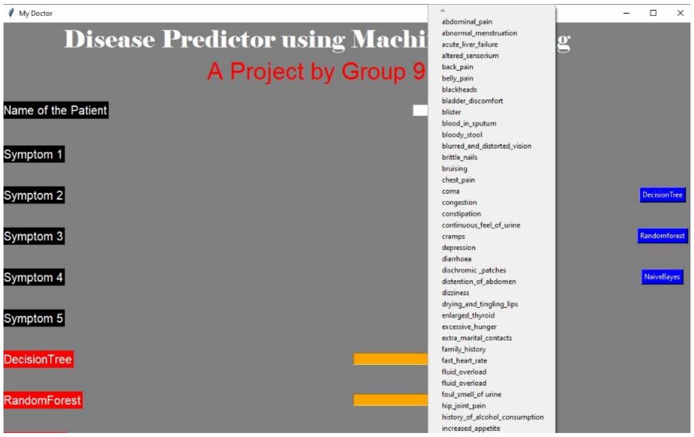
Symptom 3 None

Symptom 4 None

Symptom 5 None

DecisionTree

RandomForest



My Doctor

Disease Predictor using Machi g

A Project by Group 9

Name of the Patient

Symptom 1

Symptom 2

Symptom 3

Symptom 4

Symptom 5

DecisionTree

RandomForest