**HEALTH INSURANCE PREDICTION**

**PROJECT REPORT**

**Submitted by**

**SEALIONS**

**T. PURUSHOTHAM**

**P. LOKESH**

**A. LOKESH**

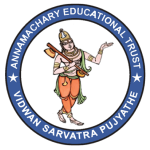
**N. PRATHAP**

**S. MUNIBHASKER**

***In partial fulfilment for the award of the Certificate***

**of**

**SUMMER INTERNSHIP PROGRAM**



**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village, Renigunta Mandal, Tirupati, Andhra Pradesh 517520**

**July 2019.**

### BONAFIDE CERTIFICATE

This is to certify that the project entitled ”**HEALTH INSURANCE**” submitted by **T.PURUSHOTHAM**,**P.LOKESH, A.LOKESH, N.PRATHAP, S.MUNIBHASKER** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### Signature of Supervisor                                       Signature of Head of the Department

 Akshay Kumar Kothuri, M.Tech       Mrs.B.Rupa Devi M.Tech,Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS,Tirupati.

**CONTENTS**

1. **Health insurance**
   1. **Introduction**
   2. **Objective of research**
   3. **Problem statement**
   4. **Industrial profile**
2. **Review of literature**
3. **Data collection**
4. **Methodology**
   1. **Choosing a best model**
   2. **Model visualization**
   3. **Prediction**
5. **Findings & Suggestions**
6. **Conclusion**

**HEALTH INSURANCE**

1.1>INTRODUCTION

**Python:**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk and Unix shell and other scripting languages.

Libraries used in python:

* Pandas
* Numpy
* Matplotlib
* Scikitlearn
* Seaborn

Softwares used:

* Anaconda Navigator
* Python 3.7
* IBM cloud

SOFTWARES:

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, macOS, and Linux.

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages and use multiple environments to separate these different versions.

The command line program conda is both a package manager and an environment manager, to help data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

**Installation:**

We can get anaconda navigator from the link given below

<https://www.anaconda.com/distribution/#windows>

**Navigator** is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages and update them, all inside Navigator.

The following applications are available by default in Navigator:

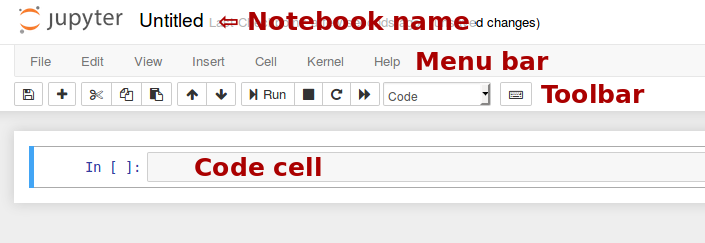
* JupyterLab
* Jupyter Notebook
* QTConsole
* Spyder
* VSCode
* Glueviz
* Orange 3 App
* Rodeo
* RStudio

**JUPYTER NOTEBOOK** extends the console-based approach to interactive computing in a qualitatively new direction, providing a web-based application suitable for capturing the whole computation process: developing, documenting, and executing code, as well as communicating the results. The Jupyter notebook combines two components:

* **A web application**: a browser-based tool for interactive authoring of documents which combine explanatory text, mathematics, computations and their rich media output.
* **Notebook documents**: a representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images, and rich media representations of objects.

STRUCTURE OF NOTEBOOK:

When you create a new notebook document, you will be presented with the **notebook name**, a **menu bar**, a **toolbar** and an empty **code cell**.



**Notebook name**: The name displayed at the top of the page, next to the Jupyter logo, reflects the name of the .ipynb file. Clicking on the notebook name brings up a dialog which allows you to rename it. Thus, renaming a notebook from “Untitled0” to “My first notebook” in the browser, renames the “Untitled0.ipynb” file to “My first notebook.ipynb”.

**Menu bar**: The menu bar presents different options that may be used to manipulate the way the notebook functions.

**Toolbar**: The tool bar gives a quick way of performing the most-used operations within the notebook, by clicking on an icon.

**Code cell**: the default type of cell; read on for an explanation of cells.

**Keyboard shortcuts:**

All actions in the notebook can be performed with the mouse, but keyboard shortcuts are also available for the most common ones. The essential shortcuts to remember are the following:

* **Shift+Enter: run cell**

Execute the current cell, show any output, and jump to the next cell below. If Shift-Enteris invoked on the last cell, it makes a new cell below. This is equivalent to clicking the **Cell**, **Run** menu item, or the Play button in the toolbar.

* **Esc: Command mode**

In command mode, you can navigate around the notebook using keyboard shortcuts.

* **Enter: Edit mode**

In edit mode, you can edit text in cells.

**Machine learning**

**Machine Learning** is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: **The ability to learn**. Machine learning is actively being used today, perhaps in many more places than one would expect.

1.2>Objective of research

Health insurance is an agreement whereby insurance company agrees to undertake a guarantee of compensation for medical expenses in case the insured falls ill or meets with an accident which leads to hospitalization of the insured. Generally, insurance companies have tie-ups with the leading hospitals so as to provide cashless treatment to the insured. In case the insurance company has no tie-ups with the hospital, they reimburse the cost of expenses incurred by the insured. The government also promotes health insurance by providing a deduction from income tax.

1.3>Problem statement

Buying a health insurance policy for yourself and your family is important because medical care is expensive, especially in the private sector. Hospitalization can burn a hole in your pocket and derail your finances. It will become even tough, if the person who brings in the money, is now in a hospital bed. All this can be avoided by just paying a small annual premium which would lessen your stress in case of medical emergencies. A good health insurance policy would usually cover expenses made towards doctor consultation fees, costs towards medical tests, ambulance charges, hospitalization costs and even post-hospitalization recovery costs to a certain extent.

1.4>Industry profile

A health Insurance policy equips you to get the best healthcare treatment without worrying about the huge costs payable at the time of discharge. Therefore, knowing about the claim process is an essential piece of information that the insured individual should be aware of at all times.

The two main types of health insurance claim which an individual can choose from when making a claim are:

* Cashless Claim Process

When the insured individual provides their health insurance details to the respective hospital, he/she begins to receive treatment. Upon discharge, the hospital will forward the medical bills to the designated health insurance company. The company will then audit the expenses and settle the outstanding payment due to the hospital. This process is hassle-free for the insured as the payments are between the hospital and insurance company.

* Reimbursement Claim Process

In the reimbursement claim process, the insured individual who has been admitted to a certain hospital pays for the entire treatment until discharge. Once the insurer has paid for the treatment and hospitalization costs incurred, he/she has to make a reimbursement claim to the particular insurance company. The insured individual will have to provide original bills of the hospital to the health insurance and claim reimbursement. The insurance company will audit the claim and will then decide to approve or reject it. On approval of the insurance company, the claim will be made to the policyholder. The insurance company will notify the insured individual in case the claim has been rejected.

**2>Review of Literature:**

The literature review for this project was conducted during the spring and summer of 2019.The primary focus was on literature addressing the development of health insurance inputs. However, literature that generally discusses emissions modelling inputs, including data sources, data collection methods and processing methods was included as long as it is relevant to the development of health insurance inputs**.**

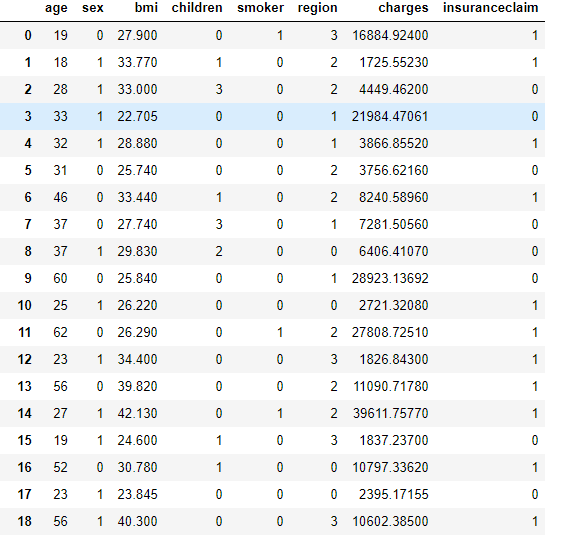
we want to make an analysis on health insurance, so when we are searching for problem statement to work on health insurance, we find problem statement regarding to this project on Kaggle website.

We fetched the required data for our project from the Kaggle website. We have understood the dataset and then we concluded that dataset is appropriate for the project we have taken. After all these finally, we have started working on this project.

**3>DATA COLLECTION:**

we are searching for the datasets on different websites, but finally we get the required dataset from the Kaggle.com website

We can access this dataset for the project by using the link given below

<https://www.kaggle.com/easonlai/sample-insurance-claim-prediction-dataset/downloads/sample-insurance-claim-prediction-dataset.zip/2>****

**Independent attributes:**

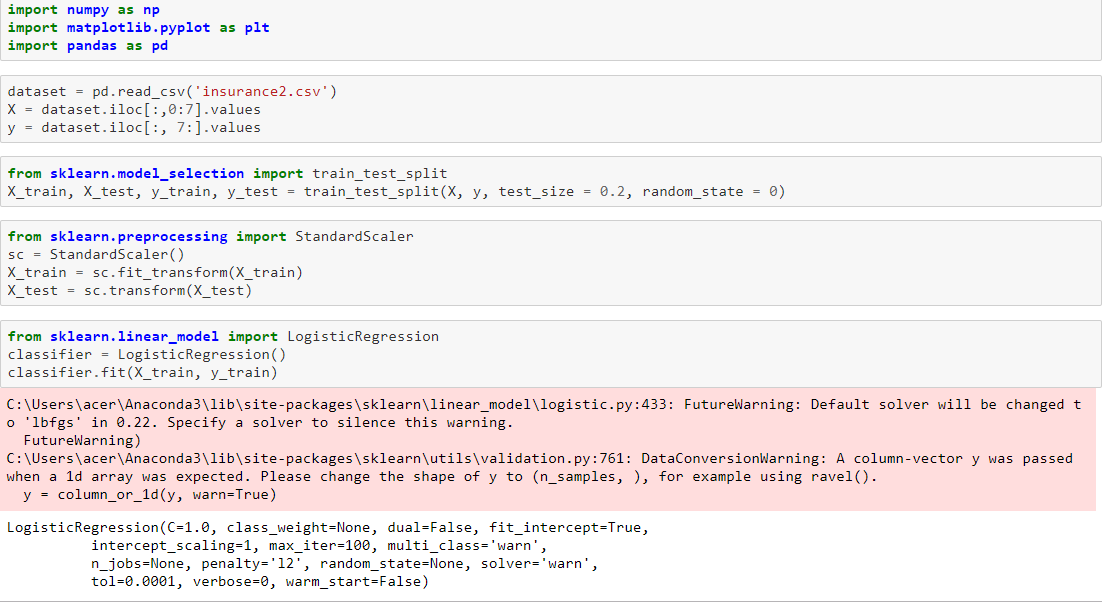
* Age
* Sex
* BMI
* Smoker
* Region
* Children
* Charges

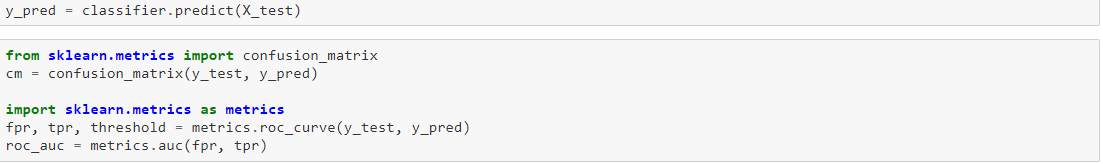
**Dependent attributes**

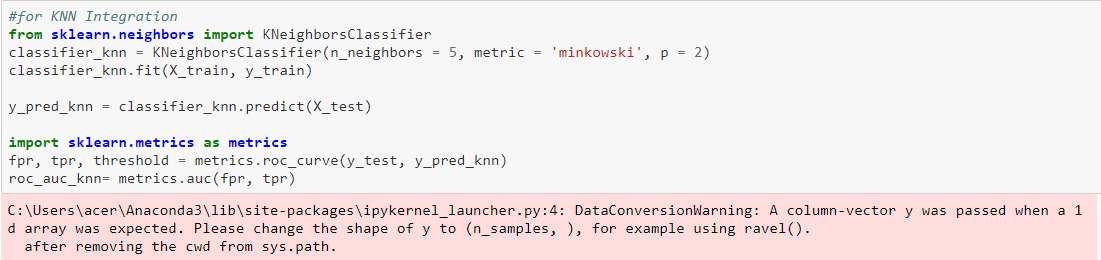
* Insurance claim

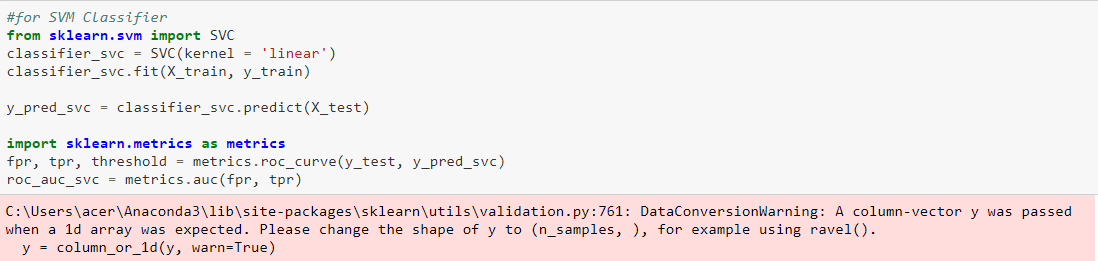
**4>METHODOLOGY:**

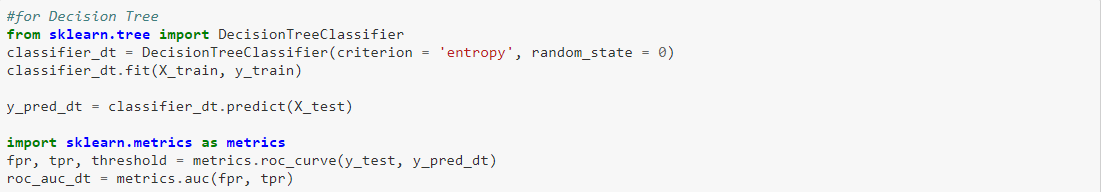
4.1>Choosing a best model:

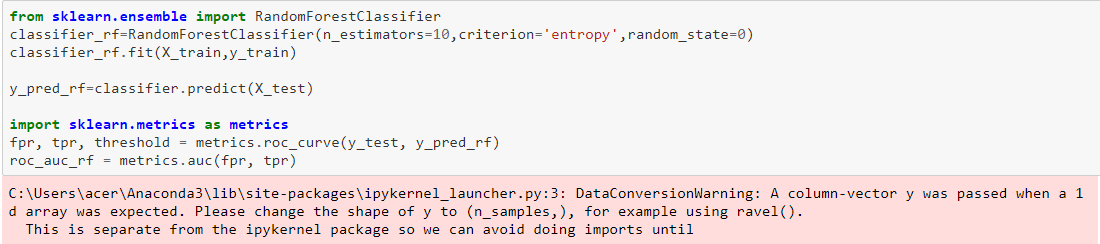


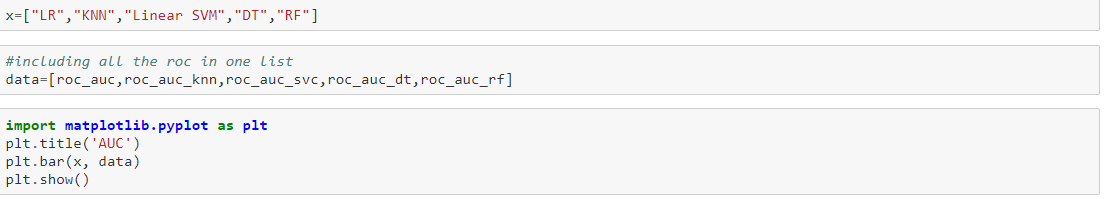


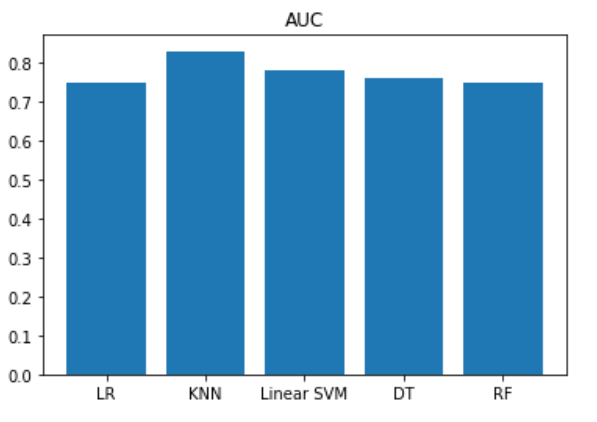






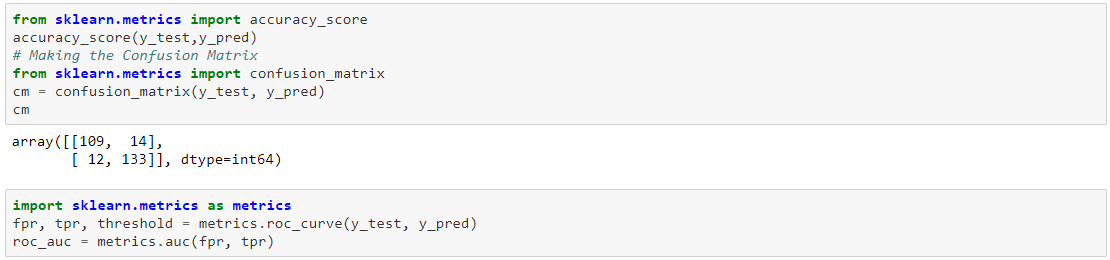
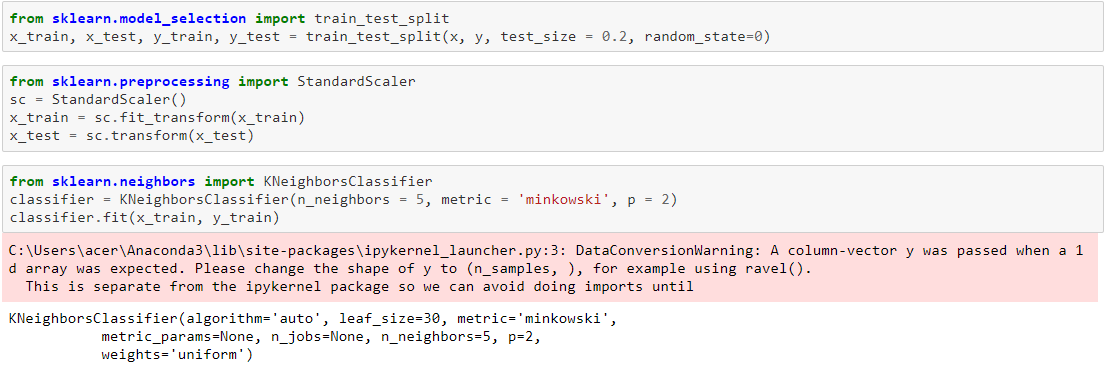
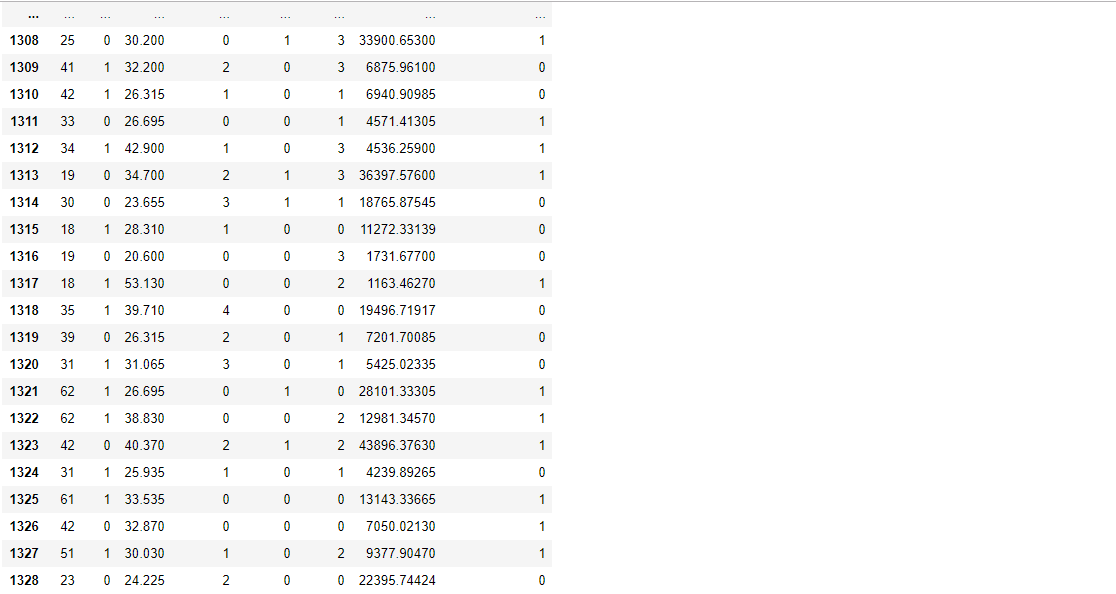
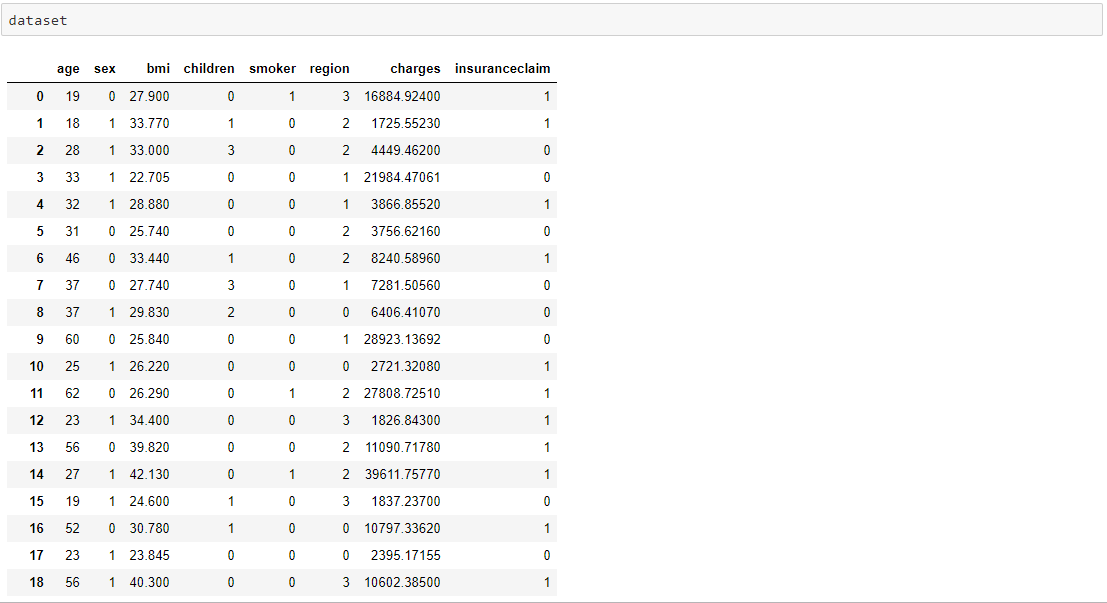


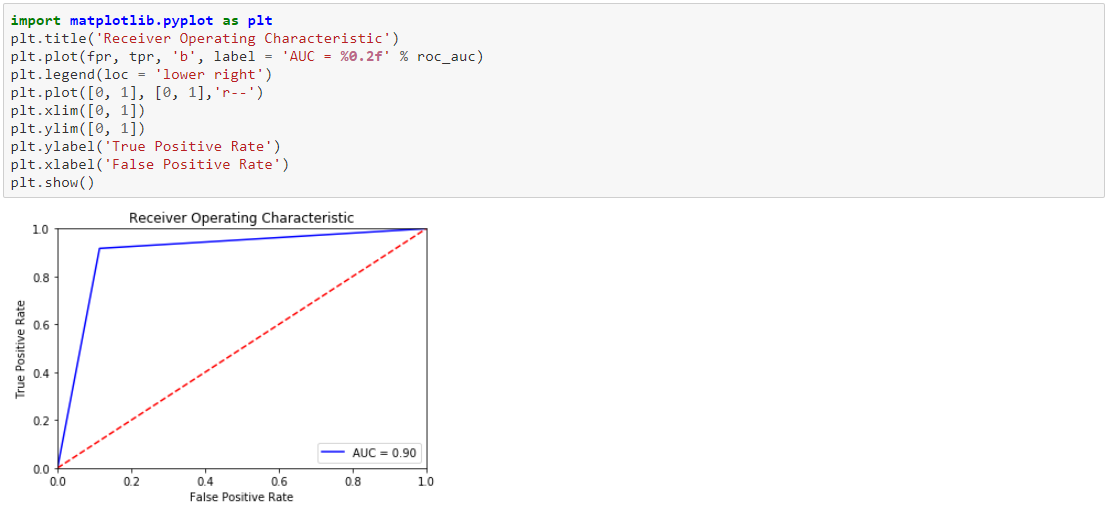




By observing the bar graph, we can conclude that KNN is the best fit model with an accuracy of 0.90 and all other models are having an accuracy approximately equal to 0.80

4.2>Model visualization:

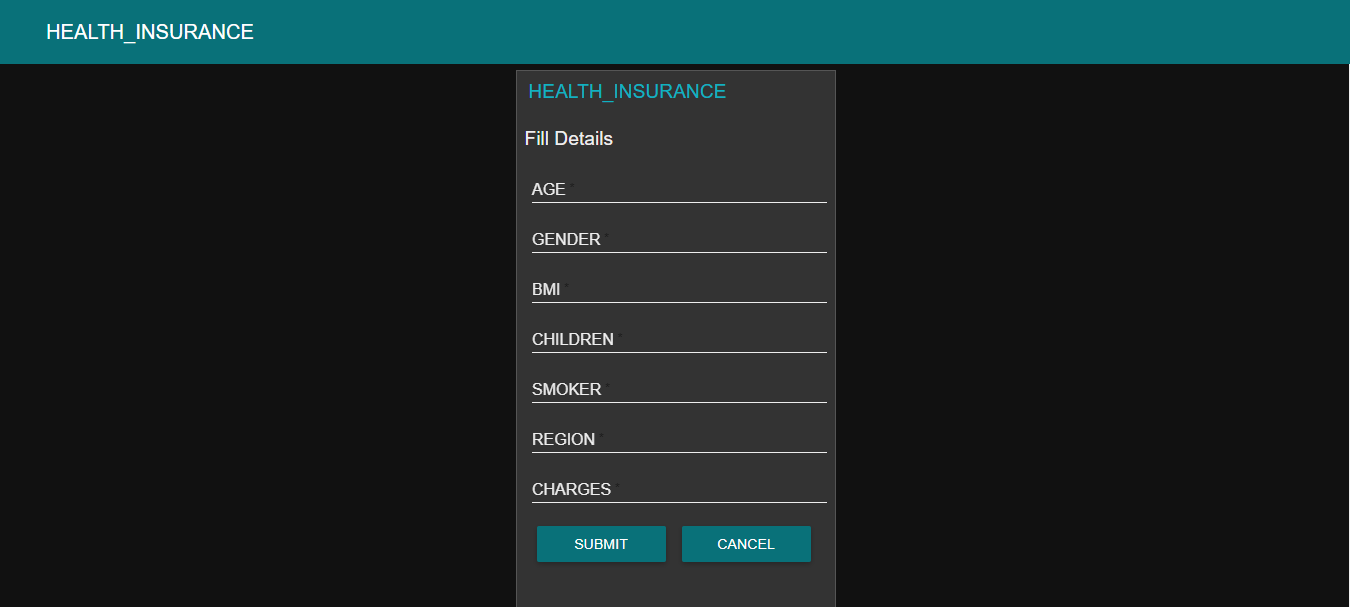




**4.3>Prediction:**

By using the following link, you can predict this model

https:/tinyurl.com/sealoins143



**5>Findings and suggestions**

Findings:

1. From the various predicted values of this project, we find that maximum number of respondents are coming under age group of 26-45 years and minimum number of respondents are above age group of 60 years.
2. The major finding of the study is that the maximum numbers of respondent are graduates while the minimum numbers of respondents are illiterates
3. It is concluded that most of the respondents are aware about the health insurance plans.

Suggestions:

* The present scenario demands almost all the customers to have a general insurance cover in order to protect from future uncertainty.
* Every company should conduct an effective research for making more and more awareness about its policy among potential customers by means of advertisements and efficient insurance agents, which in turn will help in increasing its customers.
* One can apply various models in machine learning for this dataset for acquiring better accuracy. For time being we applied only one model for the above dataset.

The remaining models like:

* + - * 1. Logistic regression
        2. Support vector machine(SVM)
        3. Decision tree
        4. Random Forest

**6>CONCLUSION:**

* We here by conclude that by using this model, we can predict whether a person will get insurance claim or not. By this, user need not contact insurance agents and need not to submit any documents. By giving minimal information he/she can know whether he gets claim or not
* The outcome of the study has proved that the performance of the model is outstanding in predicting the insurance claim