

Relative Analysis of GUI based Healthcare Prediction for Diabetes and Heart Stroke Diseases Using Machine Learning Approach

¹Narendra Yejarla, ²R.Senthil Kumar

¹Student, Department of CSE, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai, India, narendray0406@gmail.com

²Assistant Professor, Department of CSE, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Chennai, India, rsenthilmecse@gmail.com

Article Info

Volume 83

Page Number: 5178-5185

Publication Issue:

May-June 2020

Abstract

Healthcare area might be an exceptionally conspicuous research field with fast innovative progression and expanding information step by step. So as to affect enormous volume of healthcare data we'd like Machine learning which is a developing methodology in healthcare domain. Numerous patients look for medicines round the globe With different strategy. Examining the patterns in treatment of patients for analysis of a particular ailment will help in settling on educated and productive choices to improve the general nature of healthcare. AI might be a promising methodology which helps in early finding of ailment and might help the professionals in choosing for conclusion. Diabetes Mellitus and Heart stroke are growing to be extremely fatal diseases everywhere in the planet. Medical professionals need a reliable prediction system to diagnose Diabetes and Heart Stroke. Distinctive AI procedures are helpful for looking at the assorted points of view and synopsisizing it into significant data. The Openness and accessibility of monster measures of information will be prepared to give us valuable information if certain AI systems are applied there on. Diabetes and Heart stroke data will contributes to identifying the renal disorder, nerve damage and blindness. This paper targets constructing a specialized model utilizing Anaconda guide device to foresee diabetes and Heart stroke maladies by utilizing Supervised AI approach calculation. The Exploration would like to suggest the least complex calculation bolstered effective execution result for the expectation of diabetes and heart stroke infections. Test aftereffects of each calculation utilized on the dataset was assessed. It's seen that Supervised AI approach best in expectation of the infection having greatest exactness.

Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 16 May 2020

Keywords: Healthcare, Machine Learning, Disease Prediction, Anaconda Navigator, Supervised Machine Learning.

1. Introduction

The significant test that the Healthcare business faces now-a-days is predominance of office. Diagnosing the

ailment accurately and giving successful treatment to patients will characterize the standard of administration.

Poor finding causes sad results that aren't acknowledged. Healthcare domain gives huge amounts of degree to inquire about as it is massively developed. There is a need of updating the predominant medicinal service innovation by grasping digitalization of clinical data, both as far as patient gave information likewise as clinical outcomes created from cutting edge hardware. A Standard result of this data transformation that we are confronted with the overwhelming errand of deciphering and understanding the huge information accumulated. Since there's gigantic measure of information consequently Machine learning approach includes salvage. The paper centres around the forecast of the basic ailments heart stroke and diabetes.

Regardless of whether cardiovascular and diabetes ailments are found on the grounds that the significant wellspring of passing in world in old years. These are reported on the grounds that generally avoidable and reasonable maladies. The Whole and exact administration of an ailment lay on the all round planned judgment of that ailment. A Right and efficient instrument for perceiving high-hazard patients and preparing information for convenient investigation of heart disease and diabetes looks a critical need. Distinctive individual body can show various indications of heart condition which can shift as needs be. However, they regularly incorporate back agony, jaw torment, neck torment, stomach issue and smallness of breath, chest torment, arms and shoulders torments. These are a spread different of different heart illnesses which joins coronary failure and arteria coronaria ailment.

Diabetes, being the fourth significant clarification of mortality around the world, might be a significant open sick wellbeing that presents a major raising predominance.

Overall there are 415 million individuals living with diabetes and around 193 million individuals have undiscovered diabetes.

Diabetes represents very 90% of patients with diabetes and results in microvascular and full scale vascular intricacies that cause significant mental and physical misery to the two patients and parental figures. The Scope of complexities emerging from diabetes, because of the harming idea of glucose atoms on the smaller scale and large scale vascular framework incorporates: clutter, coronary heart condition, visual impairment, nephropathy, fringe neural malady, removals and male erectile brokenness.

Several diabetes and heart stroke related complications are concentrated through various old style AI and information handling strategies Identification of the dangers factors identified with these confusions are of extraordinary incentive to the clinical administration of individuals with diabetes and heart stroke On account of the elevated level of incapacity and gradual expenses of the illnesses, it's

important to inquire about the most factors required inside the beginning of complexities.

2. Methods

1. Variable Identification Process / data validation process:

A Study was performed including all patients with diabetes and heart stroke. Major challenge is the way to extract the knowledge from these data because the quantity is extremely large so data processing and machine learning techniques are often used. Also, the expected outcome and scope of this project is that if disease are often predicted than early treatment are often given to the patients which may reduce the danger of life and save lifetime of patients and price to urge treatment of diseases are often reduced up to some extent by early recognition. For this problem, a machine learning approach will train for prediction of disease. The Quick appropriation of electronic wellbeing records has made an abundance of most recent information about patients, which might be a goldmine for improving the comprehension of human health. The Above method is employed to predict diseases using patient treatment history and health data. The Dataset is now supplied to machine learning model supported this data set the model is trained. within the initiative of amassing data, information from beforehand patients datasets from like(www.kaggle.com/www.data.gov.in).

These Datasets are converged to make a standard dataset, on which examination will be finished. Exploratory data joins are Heart Disease Data Set is accessible at UCI which is Machine Learning Repository. Prime Indians Diabetes Dataset is accessible on KAGGLE.

Validation techniques in machine learning are wont to get the error rate of the Machine Learning (ML) model, which may be considered as on the brink of truth error of the dataset. On the off chance that the data volume is sufficiently huge to be illustrative of the populace, you'll not need the approval systems. Be that as it may, in certifiable situations, to figure with tests of information which won't be a genuine agent of the number of inhabitants in given dataset. To Finding the missing worth, copy worth and blueprint of information type whether it's buoy variable or number. The Example of information wont to give an impartial assessment of a model fit on the preparation dataset while tuning model hyper parameters. The Assessment turns out to be progressively one-sided as ability on the approval dataset is consolidated into the model arrangement. The Approval set is utilized to guage a given model, however this is regularly for visit assessment. It as AI engineers utilizes this information to calibrate the model hyper parameters. Information assortment, information investigation and hence the way toward tending to information substance, quality, and structure can mean a tedious plan for the day. During the strategy for information

ID, it assists with knowing your information and its properties, this information will help you pick which calculation to use to make your model. for example, statistic data are often analysed by regression algorithms; classification algorithms are often wont to analyse discrete data.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

Figure 1: Sample dataset for diabetes

	id	gender	age	hypertension	heart_disease	ever_married
0	30669	Male	3.0	0	0	No
1	30468	Male	58.0	1	0	Yes
2	16523	Female	8.0	0	0	No
3	56543	Female	70.0	0	0	Yes
4	46136	Male	14.0	0	0	No

Figure 2: Sample dataset for heart stroke

2. Data Validation/ Cleaning/Preparing Process:

Importing the library packages with loading given dataset. To analysing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A Validation dataset may be a sample of knowledge held back from training your model that's wont to give an estimate of model skill while tuning model's and procedures that you simply can use to form the simplest use of validation and test datasets when evaluating your models. Data cleaning preparing by rename the given dataset and drop the segment and so forth to explore the univariate, bi-variate and multi-variate process. The Means and procedures for information cleaning will shift from dataset to dataset. The Main objective of information cleaning is to recognize and remove mistakes and irregularities to broaden the value of information in examination and choosing.

3. Exploration data analysis of visualization and training a model by given attributes

Training dataset

- The Main line imports iris informational collection which is as of now predefined in sklearn module and informational index is

basically a table which contains data about different assortments.

- For instance, to import any calculation and train Test split class from sklearn and numpy module to be utilized during this program.
- To exemplify load_data() strategy in information dataset variable. Further gap the dataset into preparing information and test information utilizing train_test_split strategy. The X prefix in factor indicates the component esteems and y prefix signifies targets.
- This technique isolates dataset into preparing and test information haphazardly in proportion of 67:33/70:30. At that point we exemplify any calculation.
- In resulting line, we fit our preparation information into this calculation all together that PC can get prepared utilizing this information. Presently the preparation part is finished. Testing dataset:
- Now, the size of latest features during a numpy array called 'n' and it want to predict the species of this features and to try to using the predict strategy which accepts this cluster as information and lets out anticipated objective incentive as yield.
- So, the foreseen target esteem comes twisted be 0. Finally to seek out the test score which is that the ratio of no. of predictions found correct and total predictions made and finding accuracy score method which basically compares the particular values of the test set with the anticipated values.

To validate Age of patients:

```
print("Minimum value of Age of patient is:", df.Age.min())
print("Maximum value of Age of patient is:", df.Age.max())
```

Minimum value of Age of patient is: 21
Maximum value of Age of patient is: 81

Figure 3: Validation of diabetes patient ages

To validate Age of patients:

```
print("Minimum value of Age of patient is:", df.age.min())
print("Maximum value of Age of patient is:", df.age.max())
```

Minimum value of Age of patient is: 10.0
Maximum value of Age of patient is: 82.0

Figure 4: Validation of heart stroke patient ages

Exploration data analysis of visualization:

Sometimes data doesn't add up until it can check out during a visual form, like with charts and plots. Having the ability to quickly visualize of knowledge samples. A Crucial skill both in applied statistics and

in applied machine learning, it'll discover the various sorts of plots that you simply will got to know when visualizing data in Python and the way to use them to raised understand your own data.

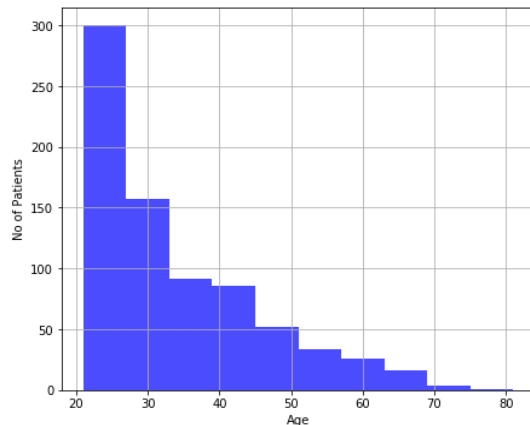


Figure 5: Age distribution of diabetes patient

Indeed, even before prescient models are set up on preparing information, exceptions may bring about deluding portrayals and progressively deceptive translations of gathered information. Exceptions can slant the synopsis dissemination of trait esteems in spellbinding measurements like mean and fluctuation and in plots like histograms and scatterplots, compacting the body of the information. At long last, exceptions can speak to tests of information examples that are pertinent to the issue like irregularities inside the instance of extortion identification and PC security.

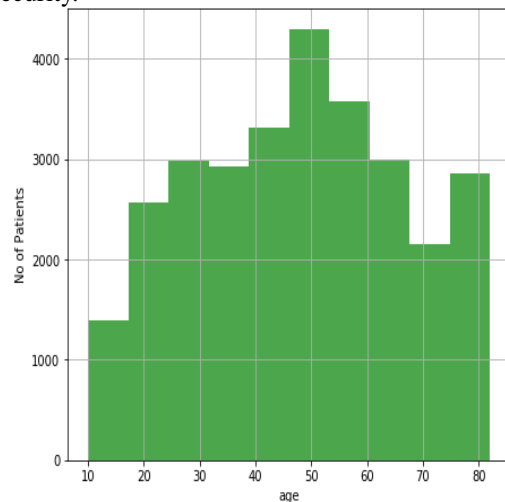


Figure 6: Age distribution of heart stroke patient

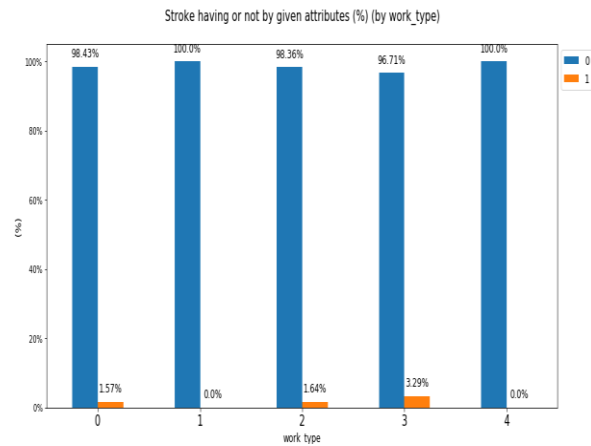


Figure 7: Percentage level of stroke patient by given attributes

4. Performance measurements of logistic regression and decision tree

Logistic Regression

It is a statistical for investigating an information set during which there are at least one autonomous factors that decide a result.

The Outcome is estimated with a dichotomous variable (where there are just two potential results). The Objective of strategic relapse is to search out the least difficult fitting model to clarify the association between the dichotomous trait of intrigue (subordinate variable = reaction or result variable) and a gathering of free (indicator or informative) factors. Strategic relapse might be a Machine Learning characterization calculation that is wont to anticipate the likelihood of an unmitigated variable . In Calculated relapse, The Variable might be a paired variable that contains information coded.

```
from sklearn.linear_model import LogisticRegression
logR= LogisticRegression()

logR.fit(X_train,y_train)

predictR = logR.predict(X_test)

print("")
print('Classification report of Logistic Regression Results:')
print("")

print(classification_report(y_test,predictR))
x = (accuracy_score(y_test,predictR)*100)

print('Accuracy result of Logistic Regression is:', x)
print("")
cm1=confusion_matrix(y_test,predictR)
print('Confusion Matrix result of Logistic Regression is:\n',cm1)
print("")
sensitivity1 = cm1[0,0]/(cm1[0,0]+cm1[0,1])
print('Sensitivity : ', sensitivity1 )
print("")
specificity1 = cm1[1,1]/(cm1[1,0]+cm1[1,1])
print('Specificity : ', specificity1)
print("")
```


Classification report of Logistic Regression Results:

	precision	recall	f1-score	support
0	0.77	0.89	0.82	150
1	0.71	0.49	0.58	81
accuracy			0.75	231
macro avg	0.74	0.69	0.70	231
weighted avg	0.75	0.75	0.74	231

Accuracy result of Logistic Regression is: 75.32467532467533

Confusion Matrix result of Logistic Regression is:

```
[[134 16]
 [ 41 40]]
```

Sensitivity : 0.8933333333333333

Specificity : 0.49382716049382713

Figure 8: Logistic regression for diabetes

Classification report of Logistic Regression Results:

	precision	recall	f1-score	support
0	0.98	1.00	0.99	8558
1	0.00	0.00	0.00	164
accuracy			0.98	8722
macro avg	0.49	0.50	0.50	8722
weighted avg	0.96	0.98	0.97	8722

Accuracy result of Logistic Regression is: 98.10823205686769

Confusion Matrix result of Logistic Regression is:

```
[[8557 1]
 [ 164 0]]
```

Sensitivity : 0.9998831502687544

Specificity : 0.0

Figure 9: Logistic regression for Heart stroke

It is one among the principal incredible and mainstream calculation. Choice tree calculation falls under the class of regulated learning calculations. It works for both nonstop additionally as clear cut yield factors. Choice tree constructs grouping or relapse models inside the kind of a tree structure. It separates an information set into littler and littler subsets while at a comparable time a related choice tree is gradually evolved. A Decision hub has at least two branches and aleaf hub speaks to a grouping or choice. The highest choice hub during a tree which relates to the most straightforward indicator called root hub. Choice trees can deal with both straight out and numerical information. Choice tree manufactures order or relapse models inside the kind of a tree structure. It uses an on the off chance that standard set which is totally unrelated and comprehensive for grouping. The standards are found out successively utilizing the preparation information each in turn. At whatever point a standard is found out, the tuples secured by the standards are evacuated.

```
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier()

dtree.fit(X_train, y_train)

predictDT = dtree.predict(X_test)

print("")
print('Classification report of Decision Tree Classifier Results:')
print("")

print(classification_report(y_test, predictDT))
x = (accuracy_score(y_test, predictDT)*100)

print('Accuracy result of Decision Tree Classifier is', x)
print("")
cm2=confusion_matrix(y_test, predictDT)
print('Confusion Matrix result of Decision Tree Classifier is:\n', confusion_matrix(y_test, predictDT))
print("")

sensitivity1 = cm2[0,0]/(cm2[0,0]+cm2[0,1])
print('Sensitivity : ', sensitivity1)
print("")
specificity1 = cm2[1,1]/(cm2[1,0]+cm2[1,1])
print('Specificity : ', specificity1)
```

Classification report of Decision Tree Classifier Results:

	precision	recall	f1-score	support
0	0.78	0.77	0.78	150
1	0.59	0.59	0.59	81
accuracy			0.71	231
macro avg	0.68	0.68	0.68	231
weighted avg	0.71	0.71	0.71	231

Accuracy result of Decision Tree Classifier is 70.995670995671

Confusion Matrix result of Decision Tree Classifier is:

```
[[116 34]
 [ 33 48]]
```

Sensitivity : 0.7733333333333333

Specificity : 0.5925925925925926

Figure 10: Decision tree for Diabetes

Classification report of Decision Tree Classifier Results:

	precision	recall	f1-score	support
0	0.98	0.97	0.98	8558
1	0.04	0.06	0.05	164
accuracy			0.96	8722
macro avg	0.51	0.52	0.51	8722
weighted avg	0.96	0.96	0.96	8722

Accuracy result of Decision Tree Classifier is 95.78078422380189

Confusion Matrix result of Decision Tree Classifier is:

```
[[8344 214]
 [ 154 10]]
```

Sensitivity : 0.9749941575134378

Specificity : 0.06097560975609756

Figure 11: Decision tree for Heart Stroke

5. Performance measurements of Support vector classifier and Random forest

Random forests or random decision forests are an outfit learning strategy for characterization, relapse and different undertakings, that work by developing a wreck of choice trees at preparing time and yielding the classification that is the method of the classes (grouping) or mean forecast (relapse) of the individual trees. Irregular choice woods right for choice trees' propensity for over fitting to their preparation set. Irregular backwoods might be a kind of directed AI calculation upheld gathering learning. Troupe learning might be a kind of realizing where you join contrasting sorts of calculations or same calculation on various occasions to make an all the more impressive forecast model. The Irregular timberland

calculation joins numerous calculation of a proportional sort for example various choice trees, prompting a backwoods of trees, henceforth the name "Irregular Forest". The Irregular timberland calculation are regularly utilized for both relapse and order assignments. Coming up next are the fundamental advances associated with playing out the irregular calculation.

- Pick N random records from the dataset.
- Build a choice tree supported these N records.
- Pick the measure of trees you might want
- in your calculation and rehash stages 1 and several . If there should be an occurrence of a relapse issue, for a substitution record, each tree inside the woodland predicts a value for Y.
- A definitive worth are regularly determined by taking the commonplace of the considerable umber of qualities anticipated by all the trees in woods. Or then again, just in the event of an order issue, each tree inside the timberland predicts the classification to which the new record has a place. At last, the new record is appointed to the class that successes the mass vote.

Support-Vector-Machines: A classifier that classifies the information set by setting an ideal hyper plane between information. I Picked this classifier since it is incomprehensibly versatile inside the amount of various kernelling limits which will be applied and this model can yield a high consistency rate. Bolster Vector Machines are maybe one among the premier mainstream and discussed AI calculations. They were incredibly famous round the time they were created inside the 1990s and still be the go-to strategy for a high-performing calculation with mall tuning.

- How to disentangle the various names wont to ask support vector machines.
- The representation employed by SVM when the model is really stored on disk.
- How a learned SVM model representation are often wont to make predictions for brand spanking new data.
- How to find out an SVM model from training data.
- How to best prepare your data for the SVM algorithm.

6. Performance measurements of KNN and Naive Bayes

K-Nearest Neighbour might be a managed AI calculation which stores all occurrences compare to preparing information focuses in n-dimensional space. At the point when an obscure discrete information is gotten, it investigations the nearest k number of occasions spared (closest neighbours) and returns the first basic class on the grounds that the expectation and for genuine esteemed information it restores the mean of k closest neighbours. Inside the separation weighted closest neighbour calculation, it loads the commitment of everything about k neighbours

predictable with their separation utilizing the resulting inquiry giving more noteworthy load to the close neighbours.

Typically KNN is solid to loud information since it's averaging the k-closest neighbours. The k-closest neighbours calculation might be an order calculation, and it's administered: it takes a lot of marked focuses and uses them to discover the best approach to name different focuses. To name a substitution point, it's at the named focuses nearest thereto new point (those are its closest neighbours) and has those neighbours vote, so whichever mark the premier of the neighbours have is that the name for the new point (the "k" is that the quantity of neighbours it checks).Makes predictions about the validation set using the whole training set. KNN makes a prediction a few new instance by rummaging through the whole set to seek out the k "closest" instances. "Closeness" is chosen utilizing a vicinity estimation (Euclidean) over all highlights.

Naive Bayes algorithm

The Naive Bayes algorithm is an intuitive method that uses the possibilities of every attribute belonging to every class to form a prediction. it's the supervised learning approach you'd come up with if you wanted to model a predictive modelling problem probabilistically. Naive bayes disentangles the estimation of probabilities by accepting that the likelihood of each credit having a place with a given class esteem is free of every other trait. This is often a robust assumption but leads to a quick and effective method. The Probability of a category value given a worth of an attribute is named the contingent probability. By increasing the restrictive probabilities together for each quality for a given class esteem, we've a likelihood of an information occasion having a place thereto class to form a prediction we will calculate probabilities of the instance belonging to every class and choose the category value with the very best probability.

7. GUI based prediction results of health care disease or not (Both Heart Stroke / Diabetes)Tkinter may be a python library for developing UI (Graphical User Interfaces). We use the tkinter library for creating an application of UI (User Interface), to make windows and every one other graphical interface and Tkinter will accompany Python as a typical package, it are often used for security purpose of every users or accountants.

3. Results and Discussion

Software involvement steps:

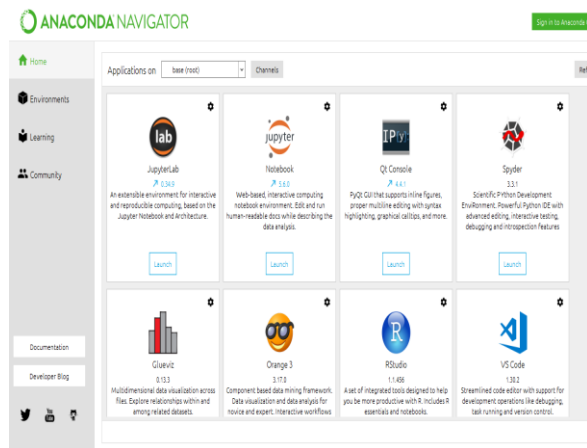


Figure 12: Open the anaconda navigator

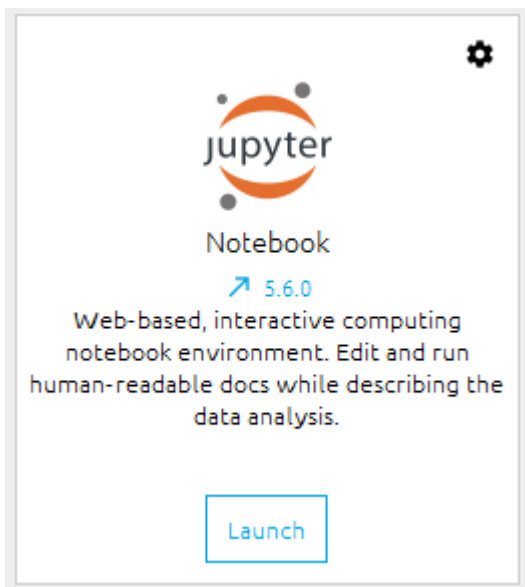


Figure 13: Launch the Jupiter notebook platform



Figure 14: Open the correspondent result Folder

4. Diabetes

Parameters	LR	DT	RF	SVC	KNN	NB
Precision	0.77	0.78	0.73	0.65	0.73	0.76
Recall	0.89	0.77	0.86	1	0.83	0.80
F1-Score	0.82	0.77	0.79	0.79	0.78	0.78
Sensitivity	0.89	0.76	0.86	1	0.83	0.8
Specificity	0.49	0.60	0.41	0	0.41	0.54
Accuracy (%)	75.32	70.99	70.56	64.93	68.83	70.99

Figure 15: Performance measurements of ML algorithm for Diabetes

Parameters	LR	DT	RF	SVC	KNN	NB
TP	40	49	34	0	34	44
TN	134	115	129	150	125	120
FP	16	35	21	0	25	30
FN	41	32	47	81	47	37
TPR	0.49	0.60	0.41	0	0.41	0.54
TNR	0.89	0.76	0.86	1	0.83	0.8
FPR	0.10	0.23	0.14	0	0.16	0.2
FNR	0.50	0.39	0.58	1	0.58	0.45
PPV	0.71	0.58	0.61	-	0.57	0.59
NPV	0.76	0.78	0.73	0.64	0.72	0.76

Figure 16: Performance measurements confusion matrix

5. Heart Stroke

Parameters	LR	DT	RF	SVC	KNN	NB
Precision	0.98	0.98	0.98	0.98	0.98	0.98
Recall	1	0.98	1	1	1	0.98
F1-Score	0.99	0.98	0.99	0.99	0.99	0.98
Sensitivity	1	0.97	0.99	1	0.99	0.97
Specificity	0	0.06	0	0	0	0.13
Accuracy (%)	98.11	96.05	98.08	98.11	98.09	96.34

Figure 17: Performance measurements of ML algorithm for Heart

Parameters	LR	DT	RF	SVC	KNN	NB
TP	0	11	0	0	0	22
TN	8558	8367	8555	8555	8556	8381
FP	0	191	3	0	2	177
FN	164	153	164	164	164	142
TPR	0	0.06	0	0	0	0.13
TNR	1	0.97	0.99	1	0.99	0.97
FPR	0	0.02	0	0	0	0.02
FNR	1	0.93	1	1	1	0.86
PPV	-	0.05	0	-	0	0.11
NPV	0.98	0.98	0.98	0.98	0.98	0.98

Figure 18: Performance measurements confusion matrix

Output screenshots



Figure 19: GUI for diabetes Prediction

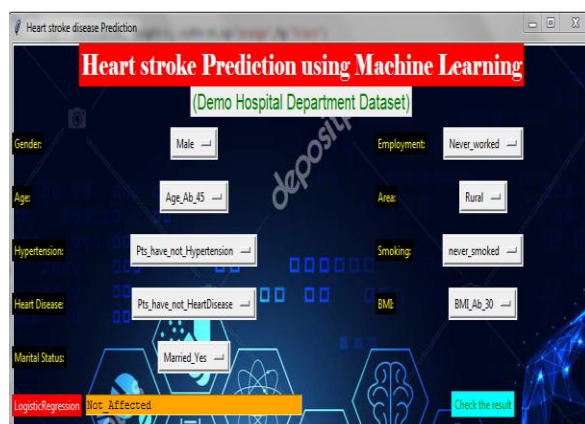


Figure 20: GUI for Heart Stroke Prediction

6. Conclusion

The investigative procedure began from information cleaning and preparing, missing worth, exploratory examination and in the long run model exploratory examination and in the end model structure and assessment. The Simplest accuracy on public test set is higher accuracy score is logistic regression for diabetes – 75.32% and for heart stroke –98.11%. This brings a number of the subsequent insights about diagnose the health care disease. Early diagnosis of stroke is most vital for the patient to scale back its impact. To presented a prediction model with the help of AI to enhance over human accuracy and supply with the scope of early detection.

7. Future Work

- Clinics need to computerize the distinguishing the patient infection from qualification process (ongoing) bolstered the given dataset detail.
- To mechanize this procedure by show the forecast end in web application or work area application.
- To optimize the work to implement in AI condition.

References

- [1] Prediction of Heart Disease Using Machine Learning Algorithms, Santhana Krishnan J.,Dept. of Comput. Applic., Anna Univ., Tiruchirapalli, India, Geetha S, Dept. of Comput. Applic., Anna Univ., Tiruchirapalli, India,2019.
- [2] Non-invasive prediction of diabetes through breath analysis: a literature review, S.Lekha ,School of Electronics, VIT Vellore, Chennai Campus, Chennai, India, 2017.
- [3] ANALYSIS AND PREDICTION OF DIABETES USING MACHINE LEARNING. S.Saru ,UG student,Sri Krishna College of Technology, Coimbatore,2017.
- [4] Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data, NIyati Gupta, A. Rawal, and V. Narasimhan, 2013.
- [5] Heart Disease Prediction Using Effective Machine Learning Techniques, Avinash Golande, Pavan Kumar T,V. Krishnaiah, G. Narasimha, N. Subhash Chandra, 2019.
- [6] Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, C. Beulah Christalin Latha S, CarolinJeeva,2019.
- [7] A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms, Amin Ul Haq ,Jian Ping Li ,Muhammad Hammad Memon , Shah Nazir and Ruinan Sun ,2018.
- [8] Edward W. GREGG, Naveed SATTAR, and Mohammed K. ALI, "The changing face of diabetes complications," The Lancet Diabetes& Endocrinology, 2016.
- [9] L. M., et.PEÑA-LONGBARDO, "Is quality of life different between iabetic and non-diabetic people? The importance of cardiovascular risks," PloS one, 2017.
- [10] Diagnosis of Diabetes Using Classification Mining Techniques, A. Ayer, J. S and R. Sumbala, 2015.
- [11] Efficient Prediction and Classification of Diabetic Patients from big data using R,K. Sharmila and S. Manickam, Sep 2015.\ Analysis of Diabetic Data Set Using Hive and R,I S. Sadhana and S. Savitha, July 2014.