Predictive model for Diabetes using Machine Learning

Project report submitted in fulfillment of the requirement for the degree of Bachelor of Technology

In

Computer Science and Engineering/Information Technology

HimanshuNadda(161366)

Piyush Thakur(161373)

Under the supervision of

Dr. Ekta Gandotra

to

Department of Computer Science &Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan- 173234, Himachal Pradesh

Candidate's Declaration

I hereby declare that the work presented in this report entitled "Diabetes Prediction System Using Machine Learning" in fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/ Information Technology submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat is an authentic record of the work carried out over a period from July 2019 to July 2020 under the supervision of Dr. Ekta
Gandotra(Assistant Professor,Computer science and Engineering).The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Himanshu Nadda(161366) Himanshu

Piyush Thakur (161373) Piyush

This is to certify that the above statement made by the candidates is true to the best of my knowledge.

Ekta Gandotra

Dr. Ekta Gandotra
Assistant Professor
Department of Computer Science & Engineering
Dated:

ACKNOWLEDGEMENT

We would like to express our special thanks of gratitude to our project guide Dr. Ekta Gandotrawho helped us in conceptualizing the project and actual building of procedures used to complete the project. We would also like to thank our Head of department for providing us this golden opportunity to work on a project like this, which helped us in doing a lot of research and we came to know about so many things.

Thanking you,

Himanshu Nadda (161366)
Piyush Thakur(161373)

Table of Contents

List of Abbreviations

- ACEI — Angiotensin-converting-enzyme inhibitor
- ANN — Artificial Neural Networks
- BUN — Blood urea nitrogen
- CAD — Coronary Artery Disease
- CKD — Coronary Kidney Disease
- COPD — Chronic obstructive pulmonary disease
- DL — Deep Learning
- DM — Diabetes Mellitus
- FN — False negatives
- FP — False positives
- Hb — Hemoglobin
- HTN — Hypertension
- IHD — Ischemic heart disease (IHD)
- LR — Logistic Regression
- ML — Machine Learning
- NaN — Not-A-Number
- PCO — Polycystic Ovary Syndrome
- RFC — Random Forest Classifier
- SES — Social Economic Status
- SVM — Support Vector Machine
- T3 — Triiodothyronine
- TN — True negatives
- TP — True positives
- WHO — World Health Organization

List of Figures

List of Tables

Abstract

Diabetes has become a common disease to the mankind from young to the old Personsnowadays. There are various reasons due to which the population of diabetic patients is increasing day by day such as obesity, bad diet, auto immune reaction, change in lifestyle, eating habits, environmental pollution etc . Hence, early prediction of diabetes is very essential to save the human life from diabetes. Data analytics is one of the branches of computer science ,which is a process of examining the large datasets and find some useful hidden patterns and draw conclusion based upon those patterns .This analytical process is carried out using machine learning algorithms in health care system.To carry out medical diagnoses,machine learning algorithms are used for analysing large medical data to build the machine learning models. This project presents a diabetes prediction system to diagnosis of diabetes.Early detection of diabetes is possible with the help of this model.

Chapter-1

Introduction

1.1Introduction

Diabetes is brought about by the expansion level of the sugar (glucose) in the blood. The diabetes can be of two sorts, for example, type 1 diabetes and type 2 diabetes. Type 1 diabetes is an immune system malady. The cells are demolished by body which are fundamental to create insulin to assimilate sugar .This sort can be caused paying little heed to heftiness. The weight is the expansion of weight list (BMI) than the typical degree of BMI of an individual . Type 1 diabetes can be found in kids and grown-ups at times. The grown-ups who are corpulent are predominantly influenced by this sort. For the most part moderately aged individuals are influenced by type 2 diabetes. Diabetes is a major reason to different ailments, for example, coronary illness, stroke ,kidney sickness, eye issues, dental infection, nerve harm, foot issues. Side effects which can cause diabetes are over the top discharge of pee (polyuria), thirst, consistent appetite, weight reduction, vision changes, and weakness, can happen suddenly.[1]

1.2 Problem Statement

The serious issue which is executing a great many individuals all through the world is diabetes. In any case, with the progressions in innovations human life is succeeding. Hence for the better eventual fate of human life and medicinal services why not utilize these advances. Different AI and deep learning calculations are utilized for some sort of forecast offices. Often these calculations are utilized by business giants for benefit in deals. Given the subject how might we utilize innovations for the human advancement. Different calculations utilized and learnare to be tested for expectation of something whose specialization just lives in the hands of specialists. So as to learn different complexities of different highlights of bio mechanics of human body and foresee the entangled issues of individuals. the machine must be prepared with the attitude of a specialist with different highlights and outer components gave from a valid dataset.

1.3 Objectives

The principle target of this expectation framework for diabetic patients is to discover a helpful model to serve humankind and can be comprehended be the accompanying focuses.

a) Implementing the essentials of AI

b) To discover connection between Diabetic Patient and his different components that influences the malady

c) Compare performances of all algorithms and in the end use the most efficient model.

1.4 Methodology

In clinical field grouping of information into various classes is finished by utilizing diverse order systems as indicated by some compels relatively an individual classifier. Diabetes influences the capacity of the body in creating the hormone insulin.which brings about the raise the degrees of glucose in the blood and turn makes the digestion of sugar anomalous. An individual for the most part experiences high glucose. Strengthen thirst, Intensify hunger and Frequent pee are a portion of the side effects caused because of high glucose. Numerous difficulties happen if diabetes stays untreated. Diabetic ketoacidosis and nonketotic hyperosmolar trance like state are a portion of the significant complexities. Numerous scientists are directing investigations for the conclusion of diabetes utilizing different order calculations of AI approaches like Naive Bayes, Decision Tree, Random timberland and so on. In Machine Learning we can prepare the PC to gain from different datasets By applying different calculations and experiencing a lot of cost work the idea of AI calculations can be utilized in prescient model regularly applying a connection between these factors.[2]

Chapter-2

Literature survey

2.1 Introduction

Diabetes is a disease of various problems. The World Health Organization (WHO) survey has found that 1.2 million people died due to this chronic disease. Moreover 2.2 million people suffered the same fate due to cardiovascular diseases.The risk of stroke and other heart diseases increases due to increase in level of glucose.. Type 2 diabetes patientsare prone to biological factors of hypertension, chest pain, obesity etc. making them more vulnerable to this disease.Blood pressure, cholesterol, triglycerides, obesity, sedentary lifestyle, abnormal sugar levels and smoking etc are factors that are mutually exclusive for both of these chronic diseases. Abundant amount of research is done in past 10 years and this research has created an opportunity to develp an important tool for future references.

2.2 Research aboutDiabetes

Diabetes is an incessant infection wherein levels of sugar and glucose are very unsteady. A few illnesses are the consequence of this shakiness. Once in a while these medical problems can cause abrupt passing moreover. Diabetes is an ailment which results in light of turmoil for digestion .It can be arranged in three types.

There are numerous individuals who are experiencing this sickness and number of these kind of individuals are expanding step by step. It has been found in ongoing overview that one out of 11 grown-ups are experiencing this sickness. As indicated by an ongoing study it has been discovered that one of every 11 grown-ups are

experiencing this ailment. It's a serious hazardous measurement for a malady to spread that way. [3]

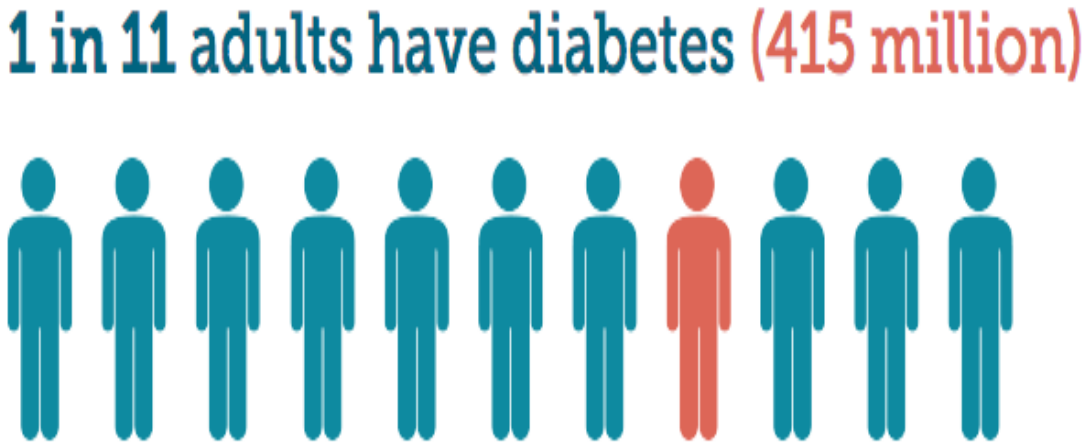


Figure1-Diabetic Rate among adults[12]

The inability of body to produce very less amount of insulin or nothing can lead to many complications. There is an extraordinary hazard on pancreas of the individual experiencing type 1 of the ailment. A recent study shows that type1 diabetes generally happens in age group of 1-20.

3]Type 2:

The inability of body to deny or resist any kind of insulin produced by the body which results in non-availability of insulin to the body. Type 2 diabetic patients are more prone to heart related ailments.According to recent survey of World Health Organisation (WHO) has found that maximum of patients suffer from type2 diabetes.

Type 3:

It is a rare type of diabetes which have a serious damage on the brain of a person,which is commonly known as Gestational Diabetes.

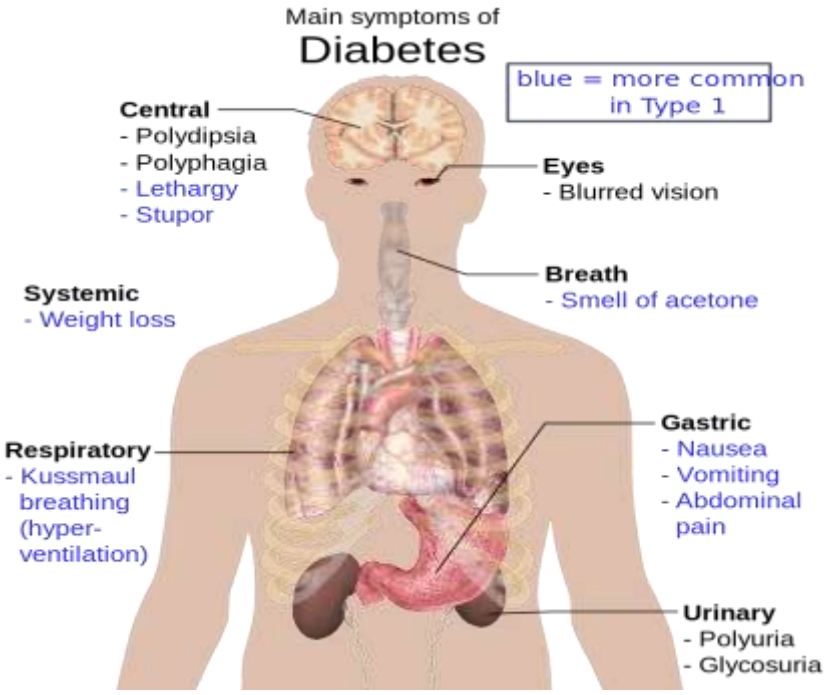


Figure2 Main symptoms of diabetes

Treatment related to low blood sugar in most cases is same for type1 and type2.Most cases are considered to be mild not medical emergencies.Feeling of unease, sweating, trembling etc. are serious effects .There are more dangerous serious effects such as aggressiveness, permanent brain damage and death in severe cases.

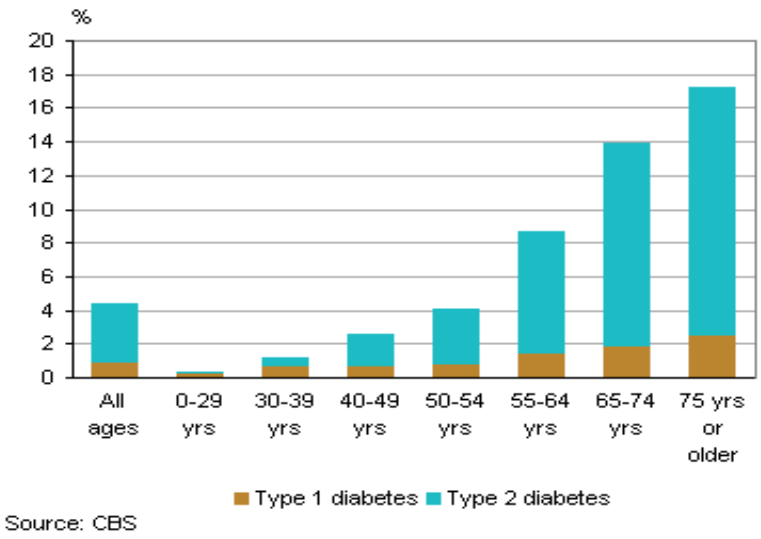


Figure3- Statistic of Type 1,type2 and type 3diabetes among people

2.3Conclusion

Diabetes is a hurtful malady which have interrelated reactions on the human body. There can be numerous highlights that are basic in both the sickness. Through those highlights here it very well may be built up that there can be an important expectation from the applicable information.

Chapter -3

System Development

We have implemented various methods or approaches to use our data systematically and in synchronized way for the purpose of the development of our model.Moreover the test plan is according to our model and can be helpful if we wants to make further improvements and developments to our model.

3.1Data

This dataset depicts clinical records for Pima Indians and whether every patient will have a beginning of diabetes inside five years.

The dataset have data related to 768 women with 8 characterstics:

i) Number of times pregnant(NTP)

ii) Plasma glucose concentration

iii) Diastolic blood pressure (mm Hg)

iv) Skin fold thickness (mm)

v) 2-Hour serum insulin (mu U/ml)

vi) Body mass index (weight in kg/(height in m)^2)

vii) Diabetes pedigree function

viii) Age (years)

Attribute no.	Attribute
1	No. of time pregnant(NTP)
2	Plasma glucose concentration(PGC)
3	Distolic blood pressure(mmHg)(DBP)
4	Triceps skin-fold thickness(mm)(TSFT)
5	2-h serum insulin(mu U/mL)(H2SI)
6	Body mass Index(kg/m2)(BMI)
7	Diabetes Pedigree Function(DPF)
8	Age
9	Outcome

Brief description of all eight features is given in table1.

Table-1: Features of patientsin dataset[17]

3.2Data Preprocessing

We may wind up drawing an off base surmising about the information, if the missing qualities are not dealt with appropriately. Since all the segments or columns probably won't be helpful for the model or the informational index that is accessible isn't in the structure wherein it tends to be utilized for the preparation of the machine in every one of these cases information pre-handling is a significant factor that decides the sound beginning of the model. Information pre-handling is a procedure which is utilized to turn crude information to valuable organization. Information Pre-handling is one of the significant highlights required for the preparation of the model. Data pre-processing incorporates checking for invalid values on the off chance that these invalid values are supplanted by mean of entire section. In data pre-processing straight out information can be changed into numerical information .label_encoder is object which help us in moving Categorical information into Numerical information.

Relationship shows the quality and course of the straight relationship between two quantitative factors. It takes esteems between - 1 and +1. A positive incentive for r shows a positive affiliation and a negative an incentive for r demonstrates a negative affiliation.

The last step in datapre-processing is thesplitting of data into training and testing data.In our ML model we have used cross_validation object from sklearn library train_test_split.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesP
Pregnancies	1.000000	0.129459	0.141282	-0.081672	-0.073535	0.017683	
Glucose	0.129459	1.000000	0.152590	0.057328	0.331357	0.221071	
BloodPressure	0.141282	0.152590	1.000000	0.207371	0.088933	0.281805	
SkinThickness	-0.081672	0.057328	0.207371	1.000000	0.436783	0.392573	
Insulin	-0.073535	0.331357	0.088933	0.436783	1.000000	0.197859	
BMI	0.017683	0.221071	0.281805	0.392573	0.197859	1.000000	
igreeFunction	-0.033523	0.137337	0.041265	0.183928	0.185071	0.140647	
Age	0.544341	0.263514	0.239528	-0.113970	-0.042163	0.036242	
Outcome	0.221898	0.466581	0.065068	0.074752	0.130548	0.292695	

Table-2 correlation table

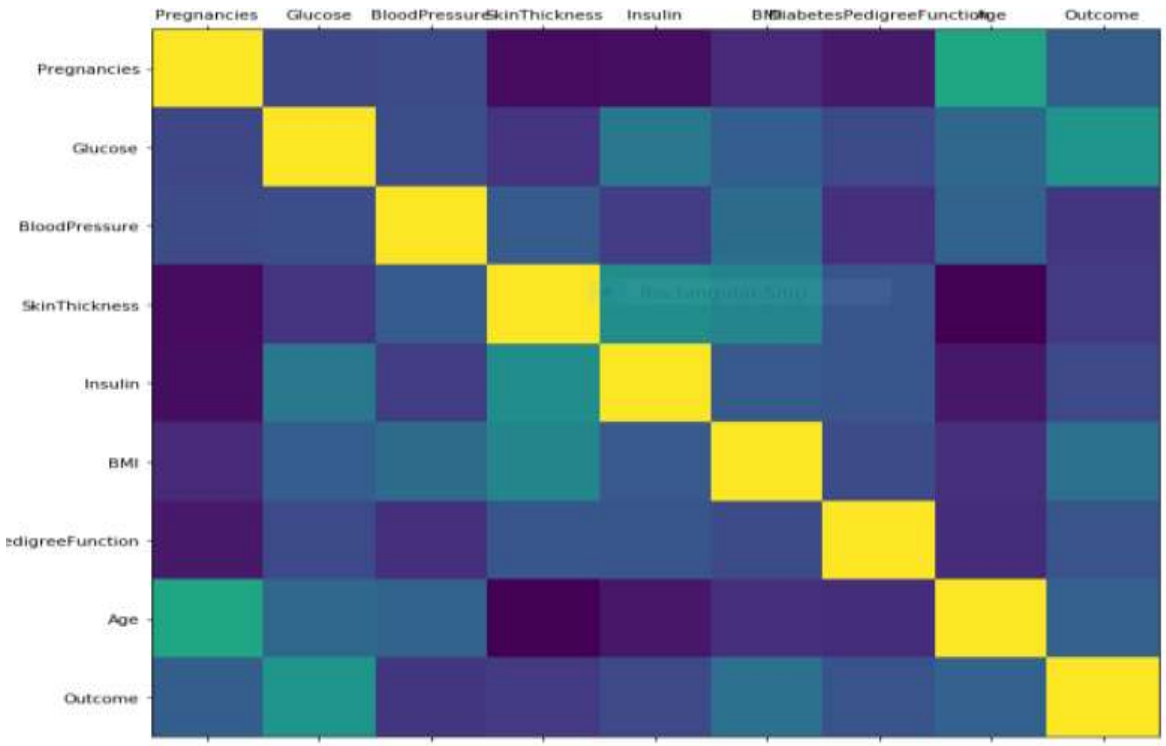


Figure4 -Correlation matrix

3.3 Conclusion

We have pre-processed our data and made it useful to the further implementation.Various missing values are replaced,many columns are deleted and converted into numerical values in order to have positive impact on model.

Chapter -4

Algorithms

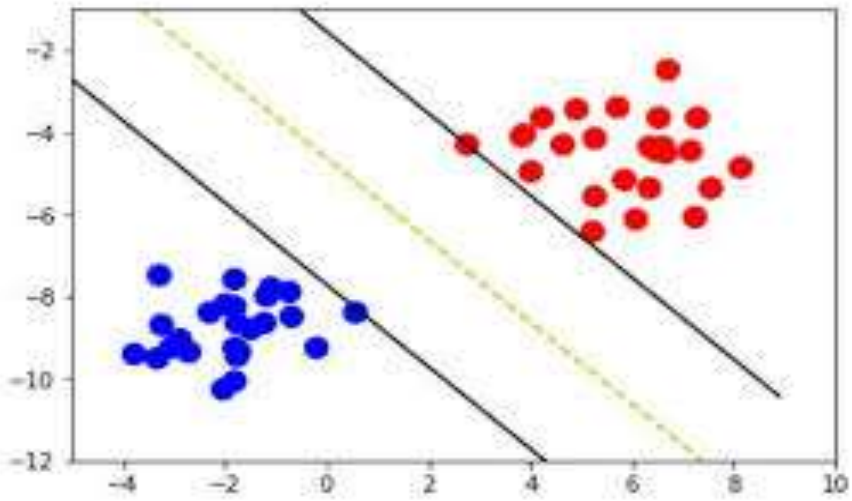
Method

Various ML and DL algorithms have been used to predict Diabetes in our dataset in this section.We will utilize logistic regression ,Random forest,Decision Tree,ArtificialNeural Networks etc. algorithms to predict and analyse results and compare these algorithms on the basis of performance.

4.1 Support Vector Machine (SVM)

It is a supervised machine learning algorithm which can be used in regression as well as for classification purposes. But it is mostly used for classification as we are using in our project. In this algorithm we plot each data item in a n dimensional space in which n is number of features. By finding the hyper-plane these points are differentiated between two different classes.One side of hyper-plane has a place with one classification and opposite side has a place with other class. This algorithm is very effective where number of dimensions are high. This algorithm identifies those extreme points also known as support vectors in order to find the hyper-plane .Every one has a place with it is possible that one classification or has a place with other classification.

The SVM graph usually looks like as following figure.



This algorithm not only just draw a simple line between two different categories, but also have a region of certain width about the line .We will fit aSVM Classifier to our pre-processed data. While the mathematical details of the likelihood model are interesting.

4.2 Logistic Regression

It is a supervised machine learning algorithm. It is chiefly utilized on the off chance that when target variable is clear cut, It utilizes a strategic capacity, additionally called the sigmoid capacity which takes an incentive somewhere in the range of 0 and 1On the premise of which we can anticipate class of the objective variable. This is sigmoid capacity

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

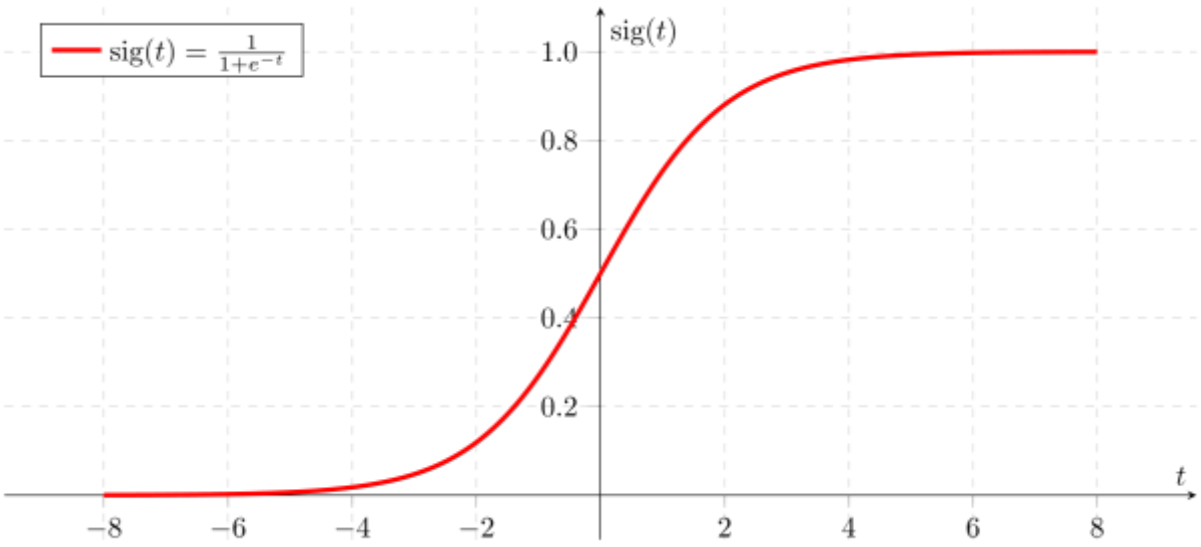


Figure6sigmoid function[11]

4.3 Random Forest Classifier

Random forest classifier falls in the category of supervised machine learning algorithm.It creates manyDTs and in the end it merges them together to give the final result. Many conditional statements are associated with multiple decision trees which helps in great accuracy.We can avoid over fitting ,due to such great number of random DTs.

The working of this algorithm can be better understood with the help of following steps-

Step1- Randomly select samples from the given date set .

Step2- This algorithm will construct DT for every sample and also get prediction from each DT.

Step3- Voting will be done on each and every predicted result.

Step4- In the last most voted prediction will be given as result.

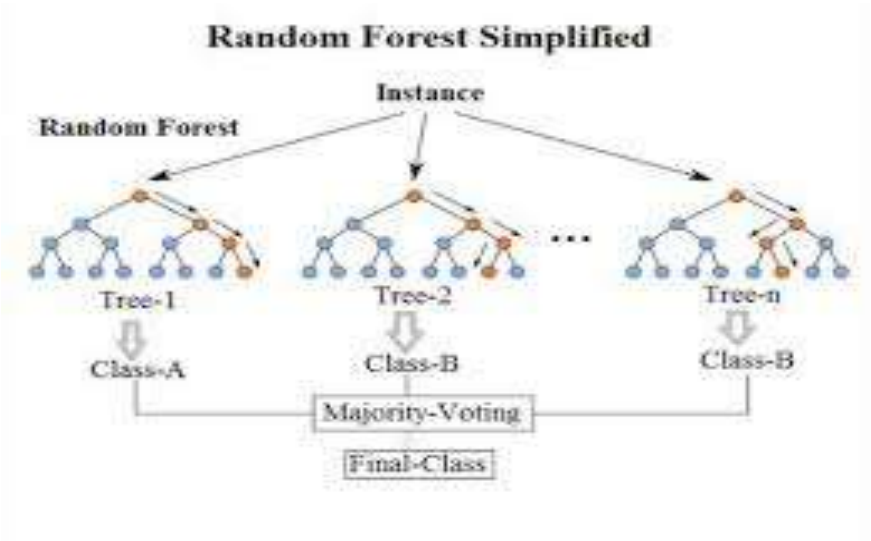


Figure7-classification in random forest[10]

4.4 Naive Bayes

Naive Bayes is probabilistic algorithmwhich is based on Bayes' Theorem. This algorithm do incrediblein grouping issues dependent supposition each element

equivalent or autonomous. Mathematical representation of Naïve Bayes algorithm is as follows-

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Probability of event A,when given event B is already true..

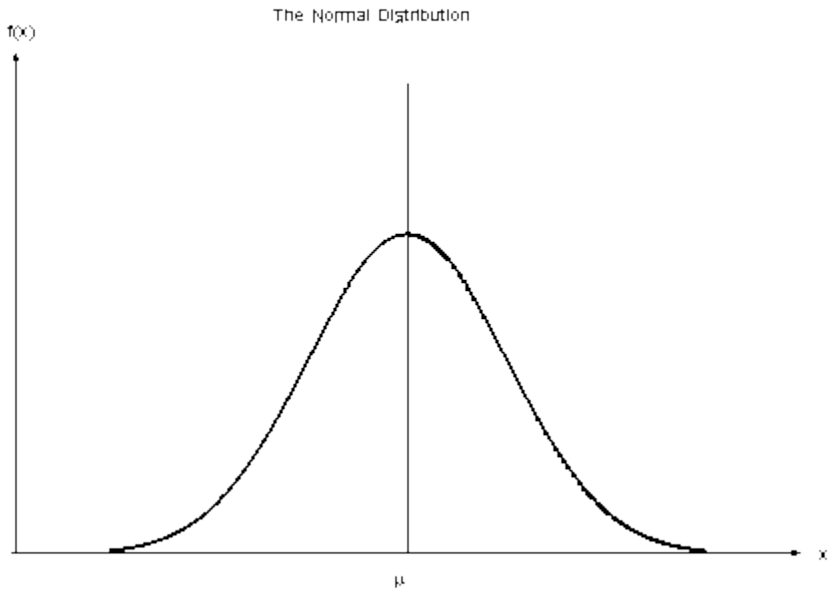


Figure8 Gaussion naïve bayes normal distribution[11]

Each component is given a similar weight(importance). For instance, knowing just temperature and mugginess alone can't anticipate the result precisely. All properties are similarly pertinent and thought to be contributing similarly to the result.

4.5 Decision Tree :

DT isthe preferred tool for classification and prediction. Decision tree looks like tree, in which the middle nodes contain the test on attributes and the leaf node contains class label.

Decision trees arrange occasions by arranging them down the tree from the establishment to some leaf hub, that gives the grouping of the case. An example is classed by starting at the establishment hub of the tree, testing the trait such by this hub, at that point descending the appendage likened to the value of the property as appeared inside the in beneath of figure. This strategy is then lasting for the subtree frozen in place at the new hub

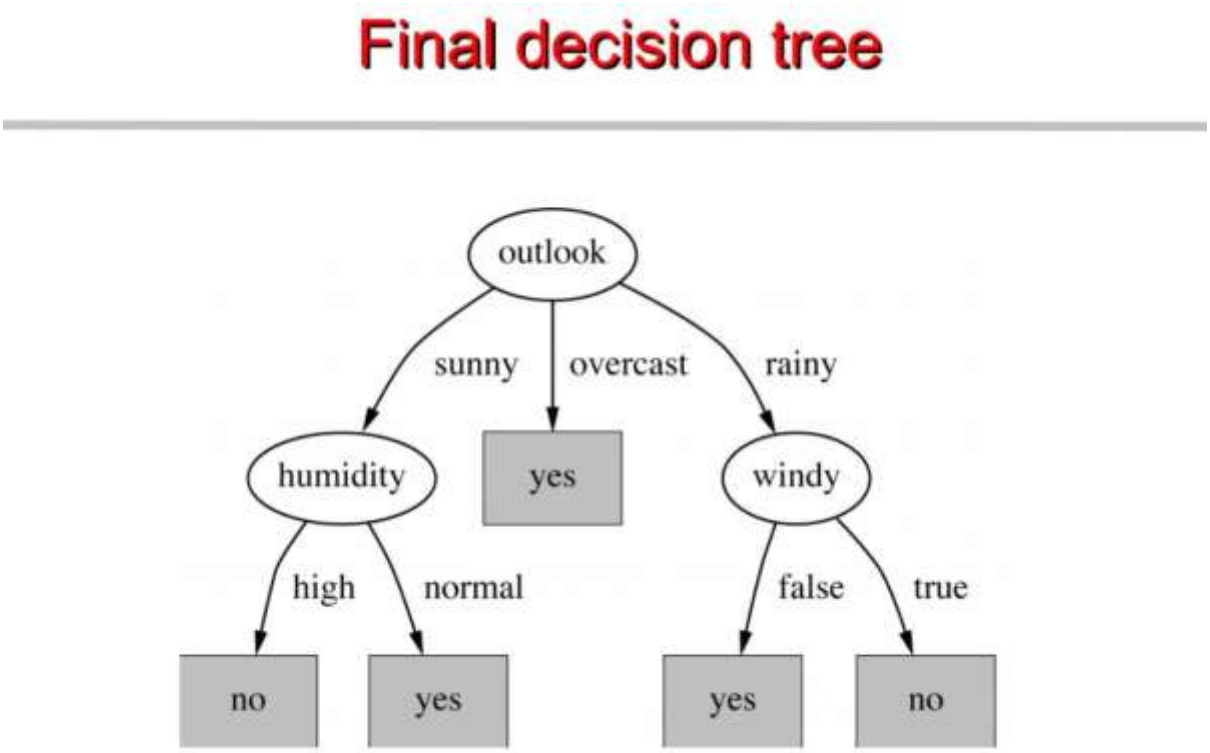


Figure9 Decision tree[14]

4.6 Artificial Neural Network (ANN)

This algorithm is simply based on human brain.It works exactly same as how humans take decisions on different conditions. Decision making is done on various internal noses of ANN.Structural representation of ANN is shown in below figure.

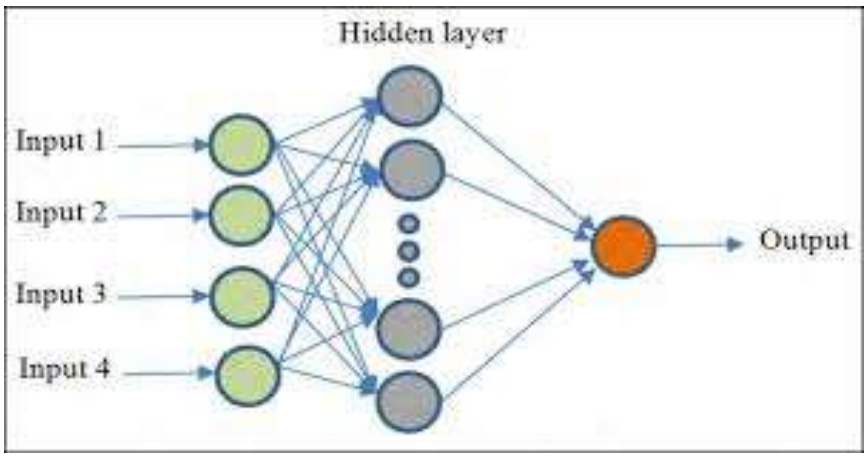


Figure10 -Structure of a Basic ANN[16]

Each info esteem has some specific weightage related with the expectation of yield parameter, Then these parameters are passed from these hubs to shrouded hubs. In these concealed layer summation and actuation work attempts to foresee the last yield. ANN basically takes a shot at input technique , where for good forecast organize is compensated and for terrible expectation it is rebuffed.

4.7 Feature Importance

So here we have 8 criteria that can be used to estimate and predict the diabietes. These criteria are the pregnancy (that is condition of women when give birth to child), quantity of glucose , the number or value of blood pressure , the thickness of human skin , Body mass Ratio also called BMI, the function of diabetes pedigree and the year to which human body has gone through with disease & the age , these all collectively going to predict the diabetes of homosapien. The collective dataset of feature are going to give crucial information about the estmation of diabetes in a human being.

All the features of our dataset play an important role in the prediction of diabetes. Calculating importance of each feature will help us to find that how much each feature is relevant in finding the output of our model.

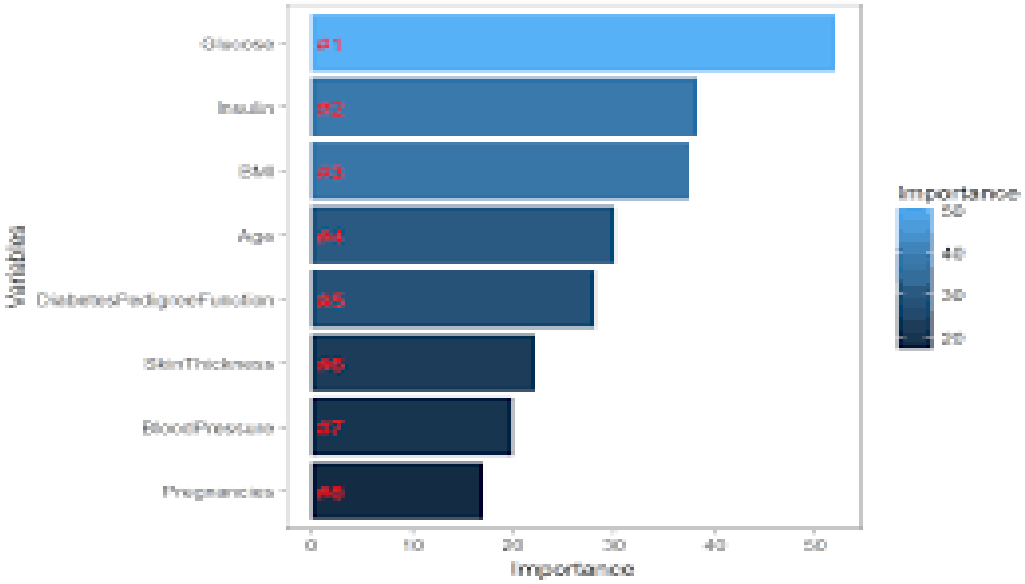


Figure11 Showing importance of all features [17]

4.8 Conclusion

We have gone through many more different kind of model and weanalyze that the top most efficient algorithim among them till now is the random forest one , that is producing best output and result from the all tried model.

Chapter -5

Result and Evaluation

We can explore the right outcomes through the maginificentstuding and going through variety of literature survey and books on the project , and here we are implementing various model for this approach deending on the accuracy and various parameter of them. The different kind of approaches and classifier of AI , deep CNN and the machine language learning are implemented and then a graph and matrix of confusion is found out with the help of them and used to find out the accuracy , specificity , uniqueness and measure or strength of performance and at last we will find out which one is closer to the real data output . at last we plot a graph for acquiring which model is closest to the original data by implementing LSM(least square method) and other distance formula.[6]

5.1 Evaluation Parameters Used

5.1.1 Confusion Matrix

Confusion matrix is just a matrix that is consists of 4 elemnts or variable namely TP, FP, FN & TN. These values are estimated as by the classifier or any model of AI and it predicts them and then check whether shows correct result or not . depending upon that the value for each of the 4 variable is inserted . Always remember this that the sum of the row of matrix is 1 and value for each of the four variable is a probability value lie between 0& 1. So this is the way the confusion matrix is formed.

TP: This indicates that this value is true or right and estimates also true.

FP: This indicates that this value is false or wrong and estimated as true.

FN: This indicates that this value is true or right and estimated as fasle.

TN: This indicates that this value is fasle or wrong and estimated also false.

		Predictions	
		Class Positive	Class Negative
Actual	Class Positive	TP	FN
	Class Negative	FP	TN

Table-3: A table about confusion matrix

5.1.2 Sensitivity

Sensitivity refers to positives which are actually positive and is estimated true with the help of modelfrom all true or positive ones. In maths or statsical expression of sensitivity can be seen below.

$$Sensitivity = \frac{TP}{TP + FN}$$

5.1.3 Specificity

Specificity refers to which are actually positive and predicted to be wrong or falsefrom themodel of AIfrom all negative or wrong. The formulatic way of representing in the Mathematical expression of specificity can be seen below.

$$Specificity = \frac{TN}{TN + FP}$$

5.1.4 Accuracy

Accuracy tells us about the correctly predicted results. This one give us reprsent or call about data that upto which extent it predicted it is accurateby a model. This could be find out by mathematical notation given below-

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

5.2Result Analysis

We used four Machine Learning Algorithms which are Random Forest, Decision Tree, logistic Regression and Naïve Bayes algorithms on the pre-processed data. The Confusion Matrix obtained is shown in table4

Classifier	Confusion Matrix
Decision Tree	55 25 44 107
Support Vector Machine	46 34 27 124
Artificial Neural Network	40 40 29 122
Logistic Regression - Cross Validation	53 27 42 109
Logistic Regression	48 32 23 128
Random Forest Classifier	48 32 28 123
Naive Bayes	51 27 32 119

Table 4-Confusion Matrix Obtained

Further from above confusion matrix we can deduce values of sensitivity ,specificationetc.In medical field we will always choose ML model with good sensitivity and specification.

Classifier	Sensitivity	Specificity	Accuracy
Logistic Regression	0.60	0.84	0.76
Random Forest Classifier	0.60	0.81	0.74
Naive Bayes	0.66	0.78	0.74
Decision Tree	0.68	0.70	0.70
Support Vector Machine	0.67	0.82	0.78
Artificial Neural Network	0.60	0.80	0.80

Table 5 Comparison of machine learning algorithms on the basis of sensitivity, specificity and accuracy

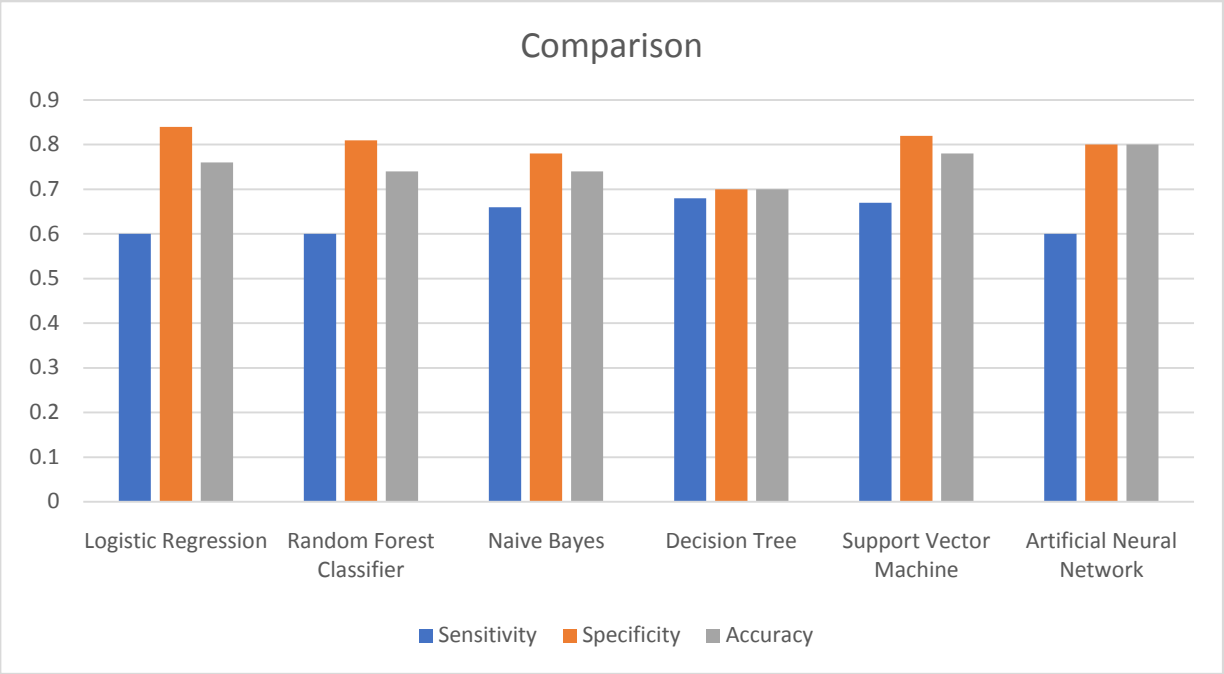


Figure 12 :Comparison of machine learning algorithms on the basis of sensitivity

Figure 12 shows the visualization of results.It is clear from the figure random forest algorithm have highest accuracy i.e 86% and Naïve Bayes algorithm have lowest accuracy i.e 80%

5.3 Conclusion

We have used six Machine Learning Algorithms which are Random Forest, Decision Tree, logistic Regression and Naïve Bayes, Support Vector Machine, Artificial Neural Network algorithms on the pre-processed data .It is clear from the result and all evaluation parameters used that Artificial Neural Network has the highest accuracy i.e 0.80. and Decision Tree algorithm has Least accuracy .

Chapter-6

Conclusion and future Work

6.1 Conclusion

Here we applied different Machine Learning and Deep Learning techniques to construct a diabetes classifier .We have accomplished ideal exactness through Artificial Neural Network classifier i.e 0.80 . One of the significant true clinical issues is the discovery of diabetes at its beginning period. In this investigation, synchronized endeavors are made in structuring a framework which brings about the forecast of ailment like diabetes. During this work, four AI order calculations are considered and assessed on different measures. Investigations are performed on Pima Indians Diabetes database.

6.2 Future Work

For future work, it is necessary to bring in hospital's real and latest patients' data for continuous training and optimization of our proposed model. The quantity of the dataset should be large enough for training and predicting. Some advanced algorithms and models should be applied in the study of diabetes. Grading forecasting standards are also necessary for potential diabetes patients. Developing a series of rules and standards is a valid method to prevent people from developing diabetes. Based on that, a more effective model for predicting diabetes and grading potential patients is presented. This will help to lower the growth rate of diabetes and eventually decrease the risk of developing diabetes.

References

[1]Journal of Computational Intelligence in Bioinformatics ISSN 0973-385X Volume 10, Number 1 (2017) pp. 1-8 © Research Foundation http://www.rfgindia.com

[2]A Method for Classification Using Machine Learning Technique for Diabetes Aishwarya. School of Computing Science and Engineering, VIT University, Vellore – 632014, Tamil Nadu, India.

[3]*2)* Causes of vision loss worldwide, 1990-2010: a systematic analysis. Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H et al. Lancet Global Health 2013;1:e339-e349

[4] A. Nather, C.S. Bee, C.Y. Huak, J.L.L. Chew, C.B. Lin, S. Neo, E.Y. Sim, Epidemiology of diabetic foot problems and predictive factors for limb loss, J. Diab. Complic. 22 (2) (2008) 77–82.

[5] Shiliang Sun, A survey of multi-view machine learning, Neural Comput. Applic. 23 7–8) (2013) 2031–2038.

[6]R.Alizadehsani, M.H. Zangooei, M.J. Hosseini, J. Habibi, A. Khosravi, M. Roshanzamir, F. Khozeimeh, N. Sarrafzadegan, S. Nahavandi, Coronary artery disease detection using computational intelligence methods, Knowledge-Based Systems, 109 (2016) 187-197.

[7] P. Sattigeri, J.J. Thiagarajan, M. Shah, K.N. Ramamurthy, A. Spanias, A scalable feature learning and tag prediction framework for natural environmentsounds, Signals Syst. and Computers 48th Asilomar Conference on Signals,Systems and Computers, 2014, 1779–1783.

[8] M.W. Libbrecht, W.S. Noble, Machine learning applications in genetics andgenomics, Nat. Rev. Genet. 16 (6) (2015) 321–332.

[9] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct.Biotechnol. J. 13 (2015) 8–17.

[10] E.M. Hashem, M.S. Mabrouk, A study of support vector machine algorithm for liver disease diagnosis, Am. J. Intell. Sys. 4 (1) (2014) 9–14.

[11] W. Mumtaz, S. Saad Azhar Ali, M. Azhar, M. Yasin, A. Saeed Malik, A machine learning framework involving EEG-based functional connectivity to diagnosemajor depressive disorder (MDD), Med. Biol. Eng. Comput. (2017) 1–14.

[12] D.K. Chaturvedi, Soft computing techniques and their applications, in:Mathematical Models, Methods and Applications, 31–40. Springer Singapore,

2015.
[13] A. Tettamanzi, M. Tomassini, Soft computing: integrating evolutionary, neural, and fuzzy systems, Springer Science & Business Media (2013).
[14] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intell. Syst. Appl. 13 (4) (1998) 18–28.
[15] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.
[16] S.A. Dudani, The distance-weighted k-nearest-neighbor rule, IEEE Trans. Syst. Man Cybernet. SMC-6 (4) (1976) 325–327.
[17]*Zhi-Hua Zhou and Yuan Jiang. NeC4.5: Neural Ensemble Based C4.5. IEEE Trans. Knowl. Data Eng, 16. 2004.*
.

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 7/15/2020

Type of Document (Tick): **PhDThesis M.Tech Dissertation/ Report B.Tech Project Report Paper**

Name: Himanshu Nadda, Piyush Thakur **Department:** Computer Science **Enrolment No** 161366, 161373 **Contact No.** 9418197210, 8350833681 **E-mail.** himanshunadda94@gmail.com, piyushthakur547@gmail.com **Name of the Supervisor:** Dr. Ekta Gandotra

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): PREDICTIVE MODEL FOR DIABETES USING MACHINE LEARNING

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 36
- Total No. of Preliminary pages = 10
- Total No. of pages accommodate bibliography/references = 2

Himanshu
Piyush

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index**at.....(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

Ekta Gandotra

(Signature of Guide/Supervisor)Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	**Excluded**	**Similarity Index (%)**	**Abstract & Chapters Details**	
	• AllPreliminary Pages • Bibliography/ Images/Quotes • 14 Words String	14%	Word Counts	
Report Generated on			Character Counts	
		Submission ID	Page counts	
			File Size	

Checked by

Name & Signature **Librarian**

Please send your complete Thesis/Report in (PDF) &DOC (Word File) through yourSupervisor/Guide at plagcheck.juit@gmail.com