iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40856

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Disease Prediction Application Using Machine Learning

Arnab Das[1], A. Udith Sai[2], P. Asha[3]
[1,2]*Student,* [3]*Professor, Department of Computer Science, Sathyabama Institute of Science and Technology, Chennai, India*

Abstract: ***The health care systems collect data and reports from the hospitals or patient's database by machine learning and data processing techniques which is employed to predict the disease so as to create reports supported the results which used for various kinds of predictions for disease and which is that the leading explanation for the human's death since past years. Medical reports and data had been extracted from various databases to predict a number of the required diseases which are commonly found in people nowadays breast cancer, heart disease and diabetes disease and make their life more critical to measure. Nowadays technology advancement within the health care industry has been helping people to create their process easier by suggesting hospitals and doctors to travel to for his or her treatment, where to admit and which hospitals are the simplest for the treating the desired disease. we've implemented this sort of system in our application to form people's life simpler by predicting the disease by inputting certain data from their reports which can give the result positive or negative supported the disease prediction they are going to be having a choice to get recommendation of best hospitals with best doctors nearby from the past users or guardians.***

Keywords: Machine learning, Logistic Regression, Random Forest, disease prediction Introduction.

I. LITERATURE REVIEW

[1] Physicians and System Physicians Author: Elliot et al [HCAHPS] might be a symptomatic device used to decide the experience of patients in numerous clinics. This data is given by medical services and framework stores, supported by the US Department of Health and Research. Medicaid and Medicare organizations use scores to survey clinical and medical services clients. Quality clinical consideration is straightforwardly connected with clinic confirmations, and numerous clinics are searching for ways of working on tolerant consideration and accomplish higher scores on HCAHPS. This study gives a short outline of the issues of fulfillment with HCAHPS gear and the classes they contain. The exploration questions are partitioned into six areas, each with various decision questions. For instance, the "doctor" area estimates patient fulfillment and doctor care through three inquiries: regard, hearing, and understanding. Every issue has four choices (Never, Sometimes, Always, Always).

Research has been directed on the connection between end-stage horribleness and mortality and how patients see a medical procedure. Creator: Sheetz and others. Fulfillment scores were utilized in their article, alongside the Michigan Joint Clinical Register, which was utilized to quantify patient fulfilment. One review analysed the connection between one patient's fulfilment with data given by at least one patient.

[3] led a racial and ethnic investigation of patients' view of treatment. Creator: Goldstein and others. Thinking back, they reasoned those non-Hispanic whites will quite often follow emergency clinics that give better quiet consideration at American, Hispanic, Asian/Pacific medical clinics, or unfamiliar patients.

[4] analysed the connection between eagerness to demand a medical clinic and other leaders. Writer: Klink Enberg in this article, I see that there are emergency clinics that attention on a great deal of things, like regard, regard, compliance, and room cleanliness for medical attendants and specialists. This proviso doesn't determine its starting point or beginning. The books in the HCAHPS documents are essentially founded on thoughts and just cover explicit parts of patient fulfilment or registration. Conversely, the strategy portrayed in this article doesn't imply that you just have thoughts. All things considered, we play out a factual examination in view of measurable investigates of all quiet fulfilment, socioeconomics, and patient insights.

Elliot et al [7] investigated the connection among orientation and various types of fulfilment, and in different examinations broke down changes in treatment classes as per patient wellbeing, skin tone, language, and, less significantly, patient schooling and age [8]. Klink Enberg [9] analysed the connection between readiness to demand an emergency clinic and other leaders. The consequences of the review show that the medical clinic centers around its endeavours to work on human connections, like kindness, regard, hearing, and cleanliness of attendants and specialists, which will expand the fulfillment score.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

II. PROPOSED WORK

In this review, we will take a gander at the solutions to the difficulties confronting the current framework by fostering the attributes of three illnesses called coronary illness, bosom malignant growth, and diabetes, and tracking down normal infections in people. We will examine the proof gathered in the Patients' Guide by anticipating three dangerous infections that are remembered for a similar program, get positive and negative input from patients and assess emergency clinics and doctors from great to awful. Parental figure input is vital and will give criticism on the most proficient method to treat their patients. Regardless of whether it is agreeable or significant, and assuming the clinic the executives is spotless and inviting. By giving web-based criticism, patients and parental figures can give positive and negative input, and we are prepared to give constant guidance to clinics and local area doctors and to foresee future results in view of infection advising. Look for exhortation from the fitting emergency clinic and specialist and get ready for the emergency clinic.

III. ARCHITECTURE DIAGRAM

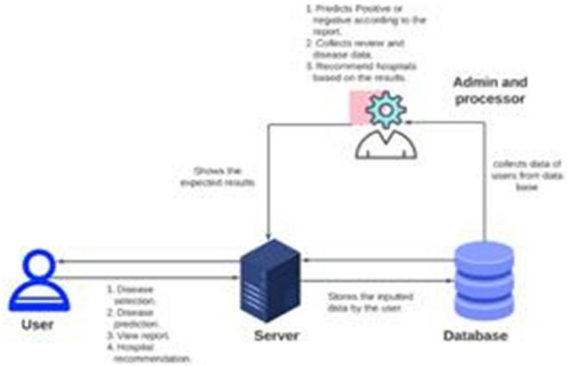


Fig:1 System Architecture.

IV. EXPLANATION

Building framework plan WebApp mirrors the overall qualities of hypermedia. This plan is in accordance with the objectives set for WebApp, the substance it gives, the clients it visits and the movement theory it has made. Content construction is intended to center, reflect, and guide things. WebApp is underlying, which chooses how to design applications like client utilization, inner handling, and traffic stream. The WebApp engineering is characterized by the climate where the material is executed.

A. Data pre-processing

Pre-handling of data is a significant stage in AI, in light of the fact that the nature of data accessible and the data required straightforwardly influence the nature of our learning model; Therefore, we must deal with our data first prior to remembering it for our model.

B. Logistic Regression

Strategic relapse is one more incredible method for dealing with a ML calculation utilized for two issues (while focusing on). The simplest thing to contemplate as far as relapse is the relapse of issues other than issue arranging. Calculated relapse primarily utilizes the strategies work depicted beneath to show twofold result factors. The main contrast among basic and obsolete is that the coordinated operations distance is restricted to 0 to 1. What's more, on account of fixing, the slack doesn't need a relationship among's feedback and result. This should be possible involving nonlinear changes in private connections. Conceptive wellbeing is utilized by MSE (Root Mean Squared Error) on the grounds that MSE (Root Mean Squared Error) can be switched utilizing a misfortune called "Likelihood Comparison (MLE)". Assuming that the likelihood is more noteworthy than 0.5, the speculation is delegated 0. In any case, give the initial step. Prior to doing any exploration on strategies, we should initially clarify the activity of logit. Since normal rationale is 100, the Logit capacities are clarified. Likelihood 0.5 compares to a rationale of 0, likelihood under 0.5 relates to a positive worth, and likelihood more noteworthy than 0.5 relates to a positive worth. The calculated capacity is somewhere in the range of 0 and 1 ($P \in [0,1]$) when the consistent tasks are inconsistent numbers that are hard to acquire from endlessness ($P \in [-\infty, \infty]$).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

C. Random Forest

Normal woods can be a gathering of trees. Here, the independence is partitioned into vectors, and each tree gives an underlying stage division called a x distribution. Customary timberlands give a gathering of guaranteed trees to make a fundamental variety of trees, and Breiman picked the best strategy, the technique for cooking or grouping each tree in one of the Random Forests, and Breiman followed the accompanying advances: Randomly organized N archives, yet additionally supplanted, as should be visible from the first numbers, this is a boot test. An illustration of this is tree establishing preparing. In the event that there is another M info, m << M chooses something similar for every hub, and m is a variable chosen from M, so a positive detachment from m addresses the property to be utilized for separation. The consistent worth of m during woods improvement. Each tree develops as large as could be expected. try not to cut. In this manner many trees are brought into the woods; The quantity of trees anticipated by the ntree boundary. The greatest number of factors (m) chose for every hub is again called "mtry" or k. The profundity of the tree can be constrained by hub boundaries (for instance, the quantity of leaves), and now and then by something like one. As referenced above, it streams from every one of the trees that fill in the backwoods to decide the degree of substitution in the wake of preparing or catching the woodland. Each tree gives another example class to casting a ballot. All tree ideas were merged and the greater part (larger part vote) grouping was affirmed at another level. Going on here, the woodland characterizes a tree backwoods assembled utilizing the RI timberland. In the ranger service area, each tree was chosen and a freight test was made for substitution, yet around 1/3 of the first material was absent. This rundown of models is called OOB (Out of pocket) data. Each tree has its own OOB data, which is utilized to look at the breaks in each tree in the timberland, and is known as the OOB break estimation.

V. WORKFLOW DIAGRAM

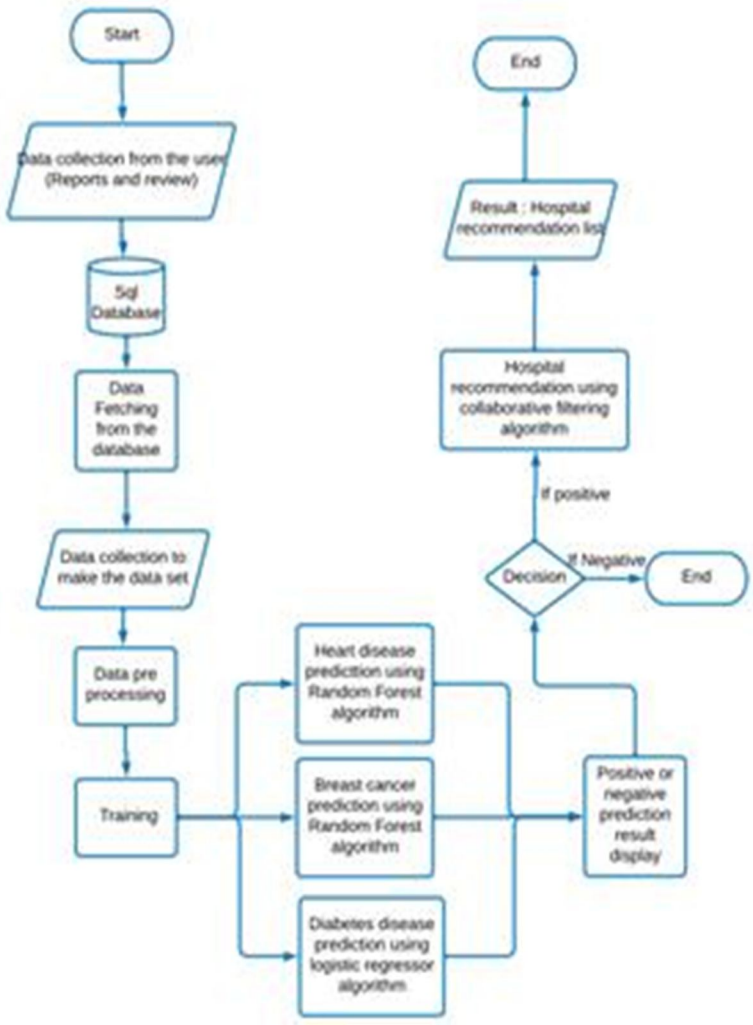


Fig:2 Flowchart Diagram.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

VI. RESULTS AND DISCUSSION

When we see around there are many patients that does not get the right treatment at the right time because of their lack of decision taking about the choice of hospital and doctors, they don't know what you do now and end up very serious at the end. The objective of the project is to provide the service to patients by suggesting them the best hospital to find their cure for their existing disease . The project is to provide a very easy solution for the patients to get recommendation to what doctor or hospital they need to go after diagnosed with a severe disease. This web application can find the solution to that, no need of thinking about what should be done after diagnosed with a severe disease. This web application handles reports to make predictions and give results accordingly to that, a best hospitals can be selected more for their treatment and more lives can be saved. After easy login or registering into the app the patient can predict their disease after inputting certain reports from their medical diagnosis report which will display accurately that the patient has the particular disease or not it will show in form of positive or negative. After the Prediction they will be having an option to get recommended hospital which are best for the treatment of their disease nearby. By this way the app can save many more lives more before its too late to get the treatment.

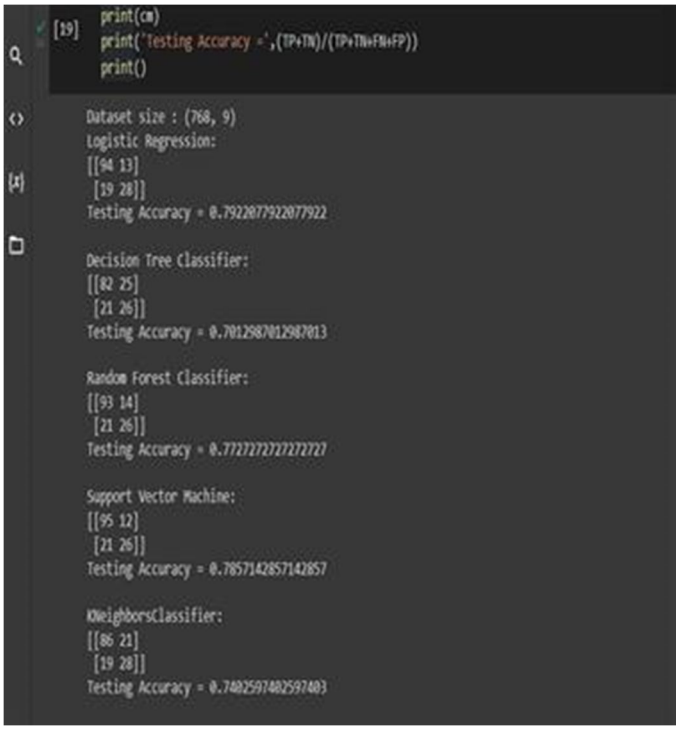


Fig:3.1 - Diabetes disease prediction algorithm selection (Logistic regression showing the best accuracy).



Fig:3.2- Performance analysis of diabetes disease prediction.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

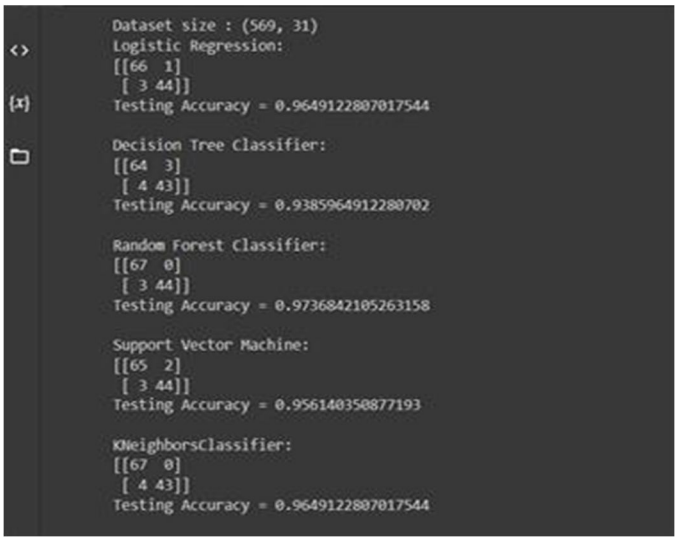


Fig:3.3 – Breast cancer prediction algorithm selection (Random Forest classifier showing the best accuracy).



Fig:3.4- Performance analysis of Breast cancer disease prediction.

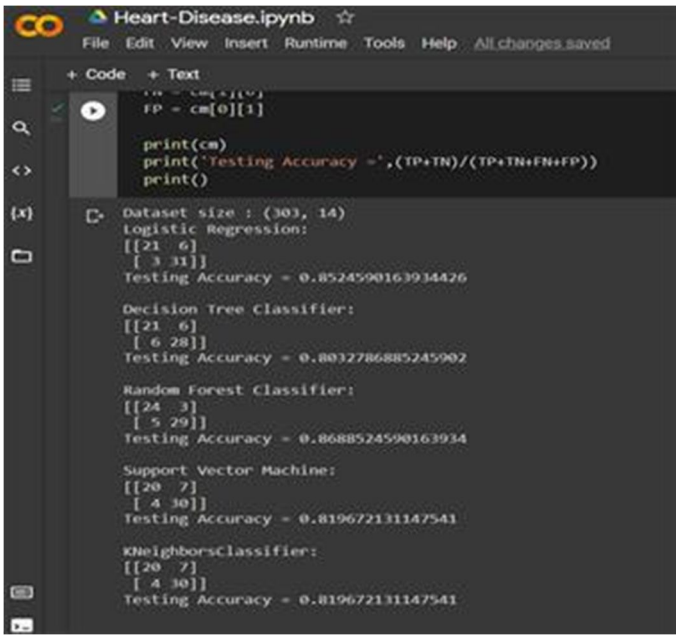


Fig:3.5 – Heart Disease prediction algorithm selection (Random Forest classifier showing the best accuracy).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

```
print("RandomForestClassifier for Heart Disease:")
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test,y_pred))
```

```
RandomForestClassifier for Heart Disease:
              precision    recall  f1-score   support

           0       0.81      0.81      0.81        27
           1       0.85      0.85      0.85        34

    accuracy                           0.84        61
   macro avg       0.83      0.83      0.83        61
weighted avg       0.84      0.84      0.84        61

0.8360655737704918
```

Fig:3.6 -Performance analysis of heart disease prediction.

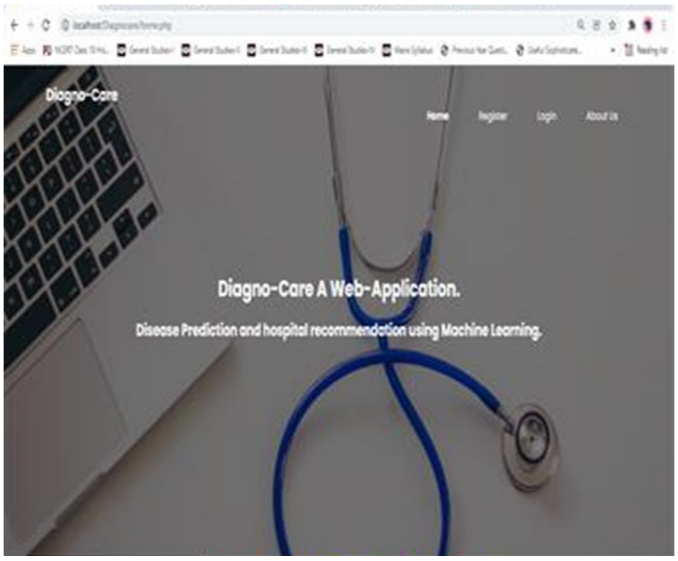


Fig:3.7 – Application Homepage (There are options for the user to register, login and to know about the application)

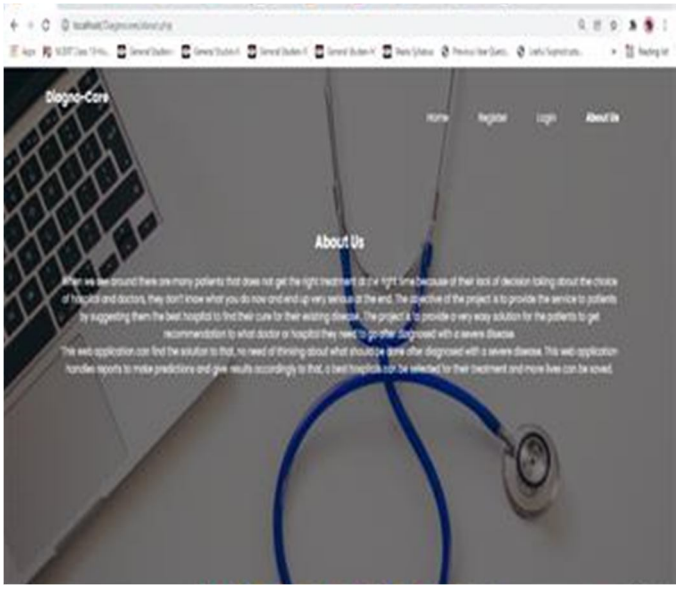


Fig:3.8 – Application About Us page (To know about the application what is the purpose and what it does).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

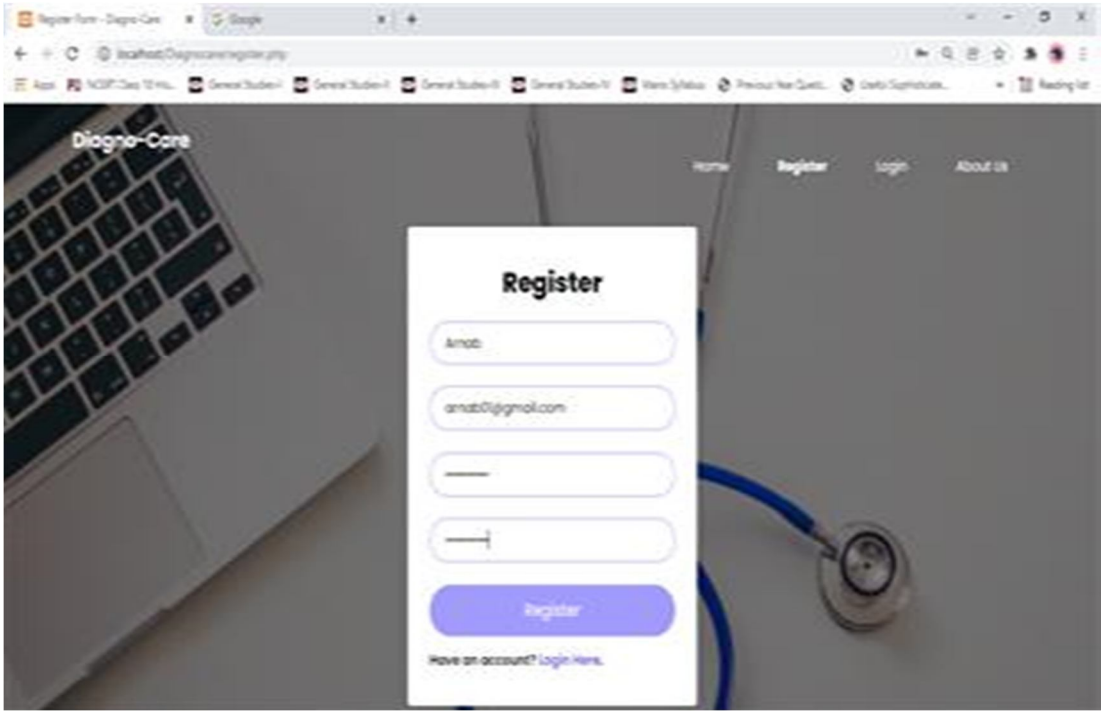


Fig:3.9 – Application Register page (new users can register here into the application with id and password to safeguard their reports and reviews)

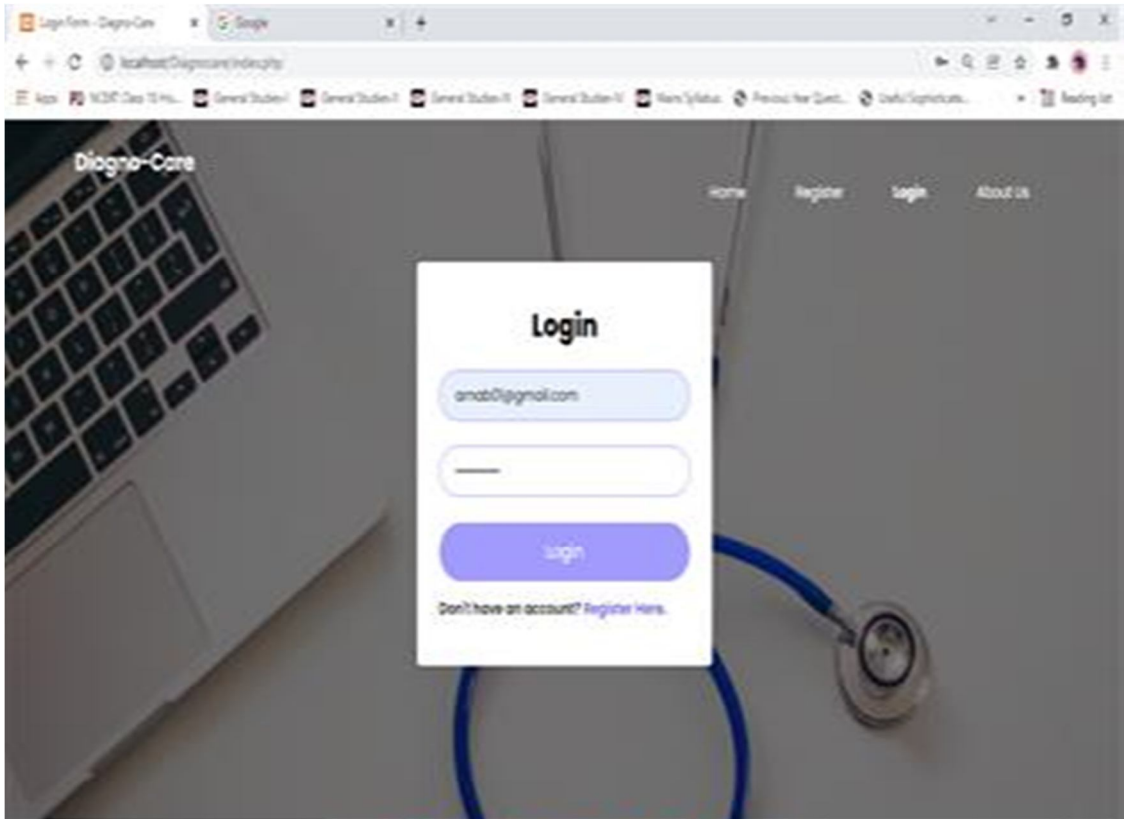


Fig:3.10 – Application Login page (Already registered users can use user id and password to login to the application and use the application for their benefits which is very user friendly to use.)

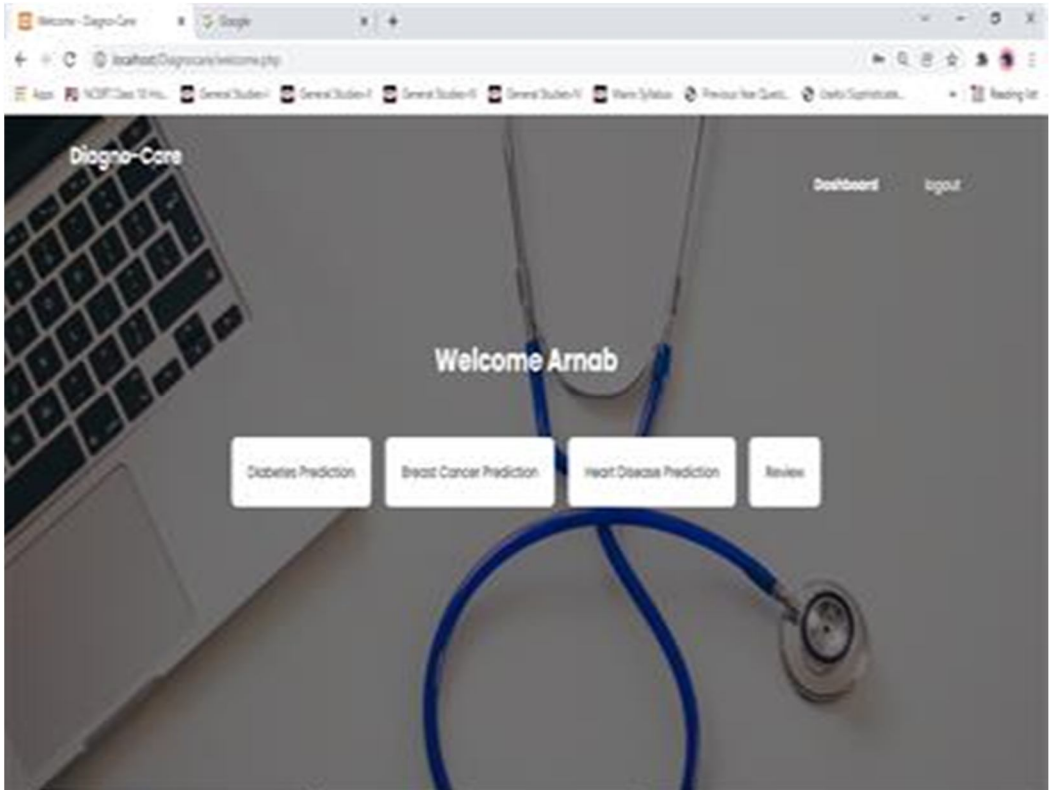


Fig:3.11 – Application Dashboard page (where users are given different options for their disease prediction and to get recommendation and give reviews)

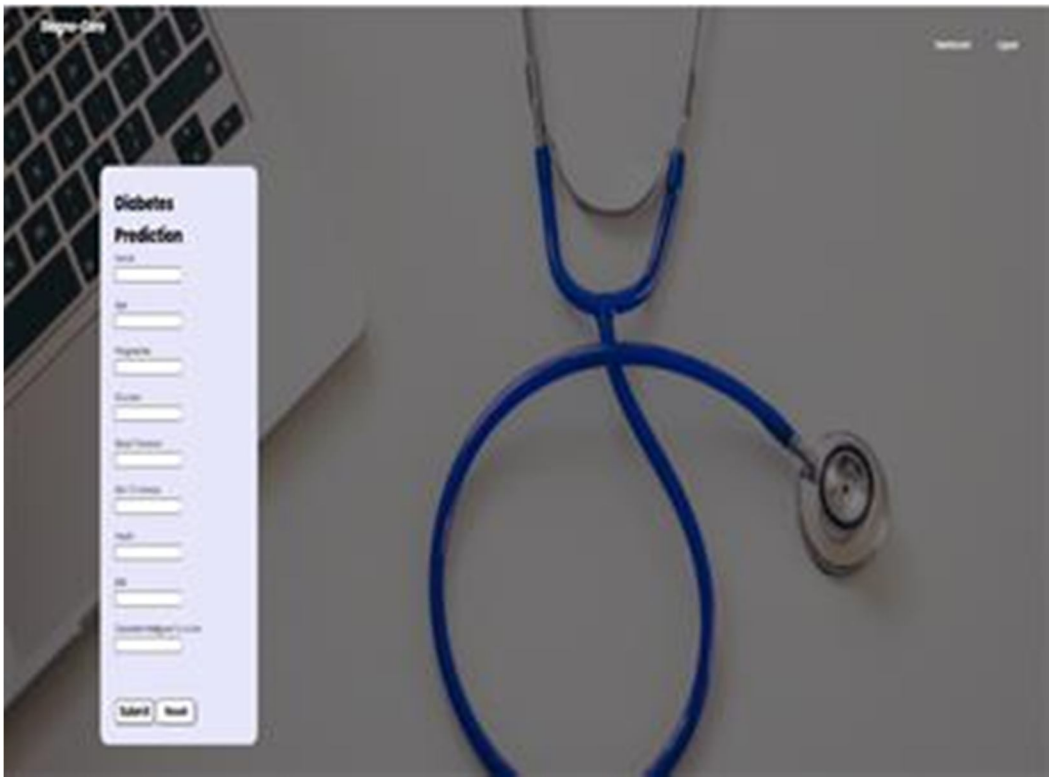


Fig:3.12 – Report entry page (where users can give enter according to their diagonised report)

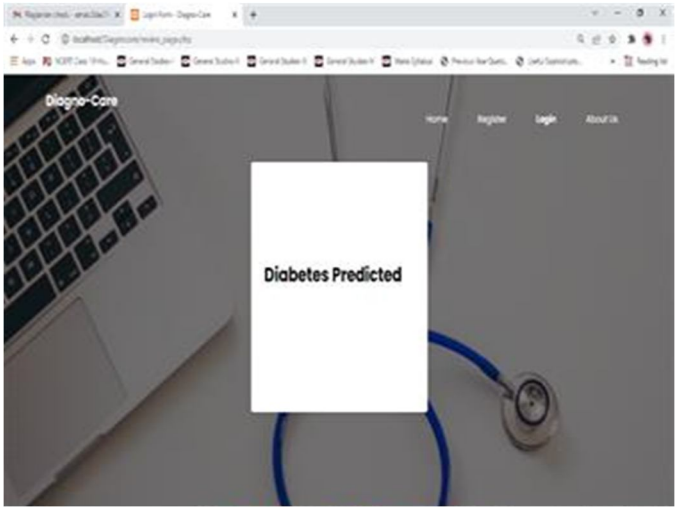


Fig:3.13 – Result page (where User get their report after the prediction whether he is suffering from the disease or not.)

VII. FUTURE IMPLEMENTATIONS

Future implementation is to recommend hospitals based on users review with the algorithm Collaborative Filtering: The motivation behind the CF calculation is to ascertain the benefits of a specific item for another item that is offered or for a chose client in view of the client's related knowledge and afterward the thoughts of different clients. In view of authenticity and effortlessness, we expect to pay attention to what different clients share for all intents and purpose and love comparable preferences. Consolidating the inclinations of the two clients is considered by the orientation correspondence of the past. All CF techniques share the capacity to anticipate or give groundbreaking plans to individual clients who will appreciate utilizing past clients. Central issues depend on the possibility of connecting customers or items, and network is characterized as the demonstration of contracting between the first or the best. The two most significant CF modes are typically executed as client based objects, while the joined solicitation technique is separated into two gatherings: memory-based and model-based. A memory-based approach is a heuristic in view of an assortment of things that clients have recently esteemed and remarked on. This expertise requires all scores, things, and clients to retain. The system depends on the utilization of an evaluation group to track down a model and make speculations. This innovation takes into account a normalized web-based evaluation technique. The CF strategy utilizes the thoughts of different networks to draw in clients. By and large, thoughts for the individuals who use them are utilized to gather the flavor of different clients. Hence, the CF expects that purchasers who have concurred in the past will probably settle on what's to come. The CF framework requires a lot of information handling, including broadband, for example, web-based business and web facilitating. Throughout the course of recent years, CF has advanced and has at long last become perhaps the most well-known method for significantly impacting the manner in which you approach directing. Today, PCs, as well as the Internet, assist us with contemplating the thoughts of an extraordinary spot with numerous individuals. People can profit from local gatherings, permitting them to acquire information from different clients and gaining from an assortment of items. Also, data can assist clients with making their own thoughts or check significant items out. Specifically, CF methods are utilized to assist clients with observing new items they might like, get guidance on explicit items, and associate with different clients who have comparative issues.

VIII. CONCLUSION

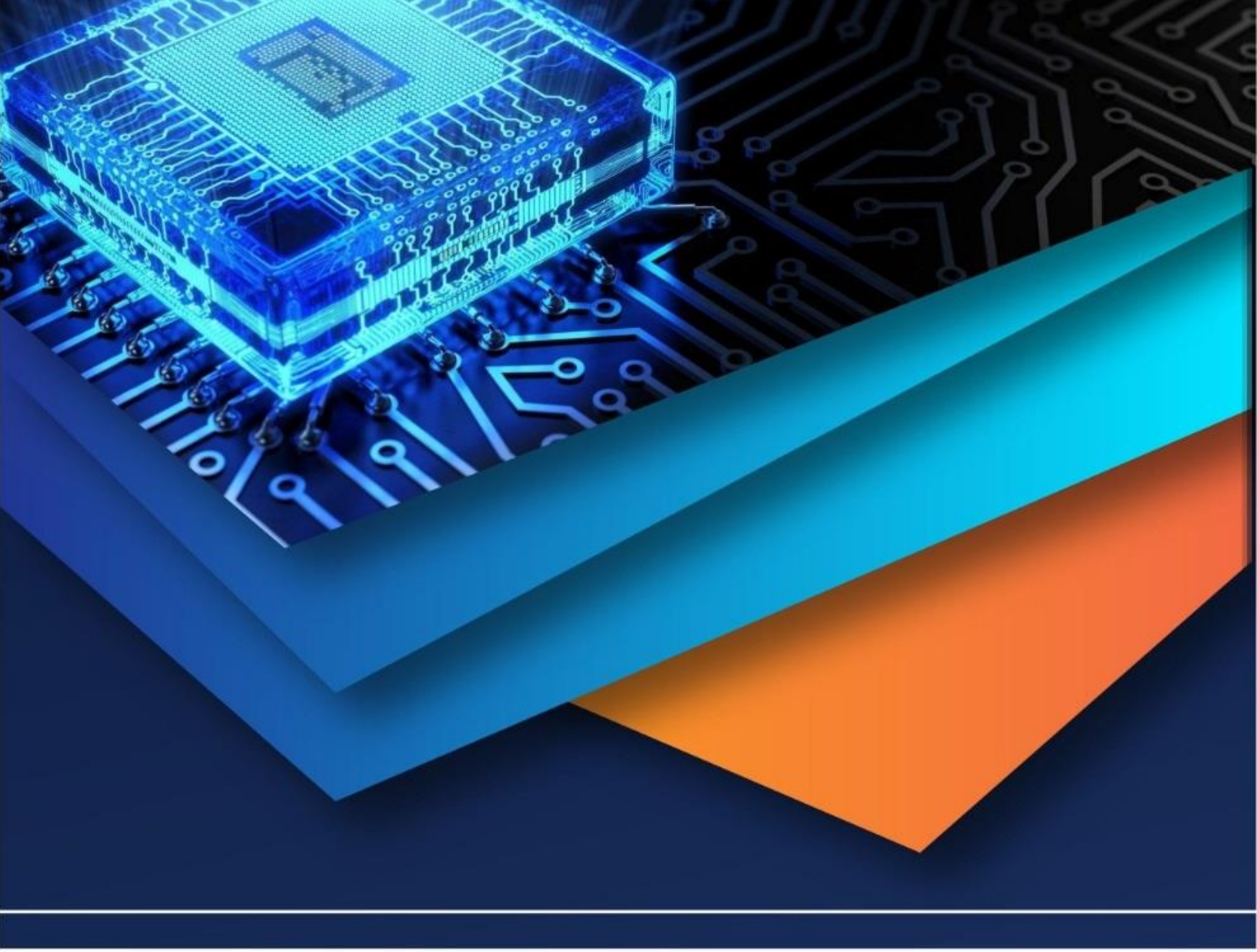
Earlier days in hospitals they lack in technological aspects for testing and issuing the reports which might take one day or may be more than that to issue the report for the lab related work that are being executed manually to predict the disease also they lack in efficiency and accuracy. But nowadays we have ample amount of data to show that these similar aspects or components can lead to this disease (exception may occur), so with the help of machine learning we have tried to implement similar system to predict the above stated disease which are most commonly found in person these days. In this application we have tried to implement a similar system which focuses on the three most deadly disease heart disease, breast cancer and diabetes diseases. We have implemented an effective way to reduce the dimensionality, reducing and eliminating the irrelevant data and increasing the accuracy. After the prediction of the disease a positive and negative report will be displayed according to which the patients can get best and nearby hospitals recommendations.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

By making it simpler for patients to get to medical services emergency clinics. As a general rule, we can utilize this program to execute groundbreaking thoughts that will help individuals with medical conditions and simultaneously track down the most ideal choice for one test. The assessments of individuals in the medical clinic and the specialists assume a significant part and they can simply decide. These offices were expected to be utilized to engage patients in light of the clinical application framework.

REFERENCES

[1] UCI Machine Learning Repository." [Online]. Available: https://archive.ics.uci.edu/ml/index.php. [Accessed: 21-Apr2018].
[2] W. Bergerud, "Introduction to logistic regression models with worked forestry examples: biometrics information handbook no. 7," no. 7, p. 147, 1996.
[3] S. Sperandei, "Lessons in biostatistics Understanding logistic statistical method," Biochem. Medica, vol. 24, no. 1, pp. 12– 18, 2014.
[4] J. R. Quinlan, "Induction of Decision Trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.
[5] T. M. Mitchell, "Decision Tree Learning," Machine Learning. pp. 52–80, 1997.
[6] L. Breiman, "Random Forest," pp. 1–33, 2001.
[7] M. Denil, D. Matheson, and N. De Freitas, "Narrowing the Gap: Random Forests In Therein, M., Matheson, D., & De Freitas, N. (2014). Narrowing the Gap: Random Forests in Theory and In Practice. Proceedings of The 31st International Conference on Machine Learning, (1998), 665–673. Retrieved from ht," Proc. 31st Int. Conf. Mach. Learn., no. 1998, pp. 665–673, 2014.
[8] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," Sch. EECS, Washington. State Univ., pp. 1–13, 2006.
[9] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," vol. 22, no. 2, pp. 103–104, 2000.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)