International Journal of Marketing Management

ISSN 2454 - 5007

www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

HEART DISEASE IDENTIFICATION METHOD USING MACHINE LEARNING CLASSIFICATION IN E-HEALTHCARE

SASHWATI ACHARYA[1], MUSUKU AMSHU[2], MANDULA SWAPNA[3], PAVAN KALYAN[4], GANDLA SAI NITHISH[5]

[1]Assistant professor, Dept.of CSE, Malla Reddy College of Engineering HYDERABAD.

[2,3,4,5]UG Students, Department of CSE, Malla Reddy College of Engineering HYDERABAD.

ABSTRACT:

In this publication, we conducted a study on cardiovascular disease using data analytics. Predicting the occurrence of cardiovascular disease is an emerging field of study, especially as more data becomes accessible. A number of researchers have looked at it using different methods. In order to identify and anticipate illness victims, we used data analytics. We started by doing some preliminary processing on data sets of varying sizes utilising three data analytics methods (Choice tree, Random forest, SVM, and KNN) to identify the most important attributes according to the connection matrix. Because of this, we were able to evaluate the consistency and accuracy of each method. Using clinical criteria, the datasets are classified. Our method use a data mining category approach to examine such factors. The optimal design showed the greatest degree of accuracy for heart disease when the datasets were examined in Python using machine learning methods.

Keywords: KNN, SVM, Random forest, Decision tree.

I INTRODUCTION

Of all the potentially fatal diseases, cardiac arrest is by far the most common. In order to learn more about heart patients, their symptoms, and the progression of their problems [1], doctors conduct several surveys on cardiac diseases. Nowadays, potentially fatal cardiac arrests are very prevalent [2]. A hint as to what lay ahead was given by a number of signs. Research in the medical field has made excellent use of technical advances to raise healthcare standards. Recent developments in

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

medical technology have made it possible to provide patients with more precise diagnoses and prognoses [3]. In order to accurately prepare for heart ailments, machine learning might be a great choice for you. As a result, three different formulae will be used. The method combines a logistic regression decision tree with an arbitrary forest. On top of that, these three methods consistently and regularly produce better outcomes [4]. Technological progress is making forecasts more comprehensible. Nowadays, people all across the globe work really hard to become famous and rich, and they live lavish lifestyles as a result. In the midst of their hectic schedules, people often neglect their health [5]. As a consequence, their diets and ways of life have changed. Conditions like high blood pressure, diabetes mellitus, and countless others are more likely to develop in young people whose lives are already fraught with stress and anxiety. The progression of cardiovascular disease may occur to any of these reasons [6].

This study used Python to run a battery of categorization and clustering algorithms on the UCI repository's cardiovascular illness dataset. The primary objective is to find all possible combinations of features and test them against various formulae. Afterwards, the best approach out of all the methods is chosen for early-stage cardiovascular disease prediction [7]. It would be much easier to identify and classify the illness if the three algorithms, Decision tree, Random forest, and Logistic Regression were applied [8]. The version is trained and classified using a dataset. The disease was predicted using the most precise and effective method once the design was trained [9].

Declaration of Issue

Some of the risk factors for the condition were found to include high blood pressure, total cholesterol, LDL cholesterol, and HDL cholesterol. Predictions for coronary heart disease in 12 years included 383 men and 227 women [9]. It was shown that using the constant variables itself brought the accuracy of this category technique up to par with CHD prediction. Because cardiac problems cannot be predicted with a higher learning rate or more accuracy in their early phases, the current algorithms can only predict them with a 93% accuracy [10].

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

II LITERATURE SURVEY

In order to provide an HDPM more effectively, several studies have confirmed an improvement in cardiac disease clinical prognosis based entirely on tool learning versions. Researchers have achieved a great deal in analysing the efficacy of forecast modifications using two publicly available coronary heart disease datasets, Statlog and Cleveland. A cardiac problem medical decision support system based on chaotic firefly approach and challenging units-based top-notch reduction (CFARS-AR) was developed by Long et al. for the Statlog dataset (2015). In order to classify the illness, units were used to reduce the shape of capacity, and health problems firefly components were used. After that, in comparison to several fantastic variants, like NB, SVM, and ANN, the installed version has become significantly different.

According to Nahato et al. (2015), RS-BPNN is the result of combining BPNN with rough devices-primarily based entirely on trends. Based on the selected characteristics, the recommended RS-BPNN achieved an accuracy of around ninety.Just 4%. A number of efficiency indicators were compared by Dwivedi (2018) across six AI variants: ANN, SVM, LR, precise enough-next-door neighbour (kNN), classification tree, and NB. According to the results, LR outperformed many other styles when it came to accuracy (85%, 89%, 81%, and 85%, respectively), phase of degree of sensitivity, location of information, and precision.

Using a combination of several AI architectures (okay-NN, DT, NB, LR, SVM, Semantic Network (NN)), and a strategy for identifying excellent traits, Amin et al. (2019) were able to complete comparative assessment. Results showed that the hybrid (combining NB and LR) with selected traits achieved an acceptable level of accuracy (87.4%). Researchers have routinely used the Cleveland cardiac health concerns dataset to provide predictive patterns.

A hybrid prediction format was developed by Verma et al. (2016) using K-method clustering, MLP, piece swam optimisation (PSO), and correlation characteristic element (CFS). As a result, the recommended hybrid model achieved an accuracy of up to 90.28 percent.

In a comparative study, Haq et al. (2018) [6] compared a hybrid model that combined various attribute alternative

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

methods with expert system styles. The model included remedy, minimal-redundancy maximal-relevance (mRMR), the very least outright contraction and alternative operator (LASSO), LR, kNN, ANN, SVM, DT, NB, and RF. According to their findings, the fundamental efficiency of the styles is impacted by the loss in capabilities. Compared to other combinations used in the studies, the study found that LR-primarily based on full device uncovering collection of guidelines (MLA) with Relief-based on a combination of the two produced the highest accuracy (up to 89%).

A style based on SVM beauty layout and inferred Fisher rating specific need gadgets (MFSFSA) was suggested by Saqlain et al. (2019). The attributes that were selected were determined by the Fisher score that was better than the desired score. Following that, SVM checked and determined the MCC using a record approach using the selected feature portion. Study after study found that combining FSFSA with SVM produces accuracy, sensitivity, and uniqueness as high as 81.19%, 79.99%, and 88.50%, respectively.

A hybrid approach combining NB, BN, RF, and MLP was proposed by Latha and Jeeva (2019). Accuracy levels as high as 85.4% were achieved by the proposed version. In order to enhance the clinical analysis technique, Ali et al. (2019) [5] advocated for loaded SVMs.

The first SVM was used to filter out irrelevant features, and the second one to predict the occurrence of cardiac problems. The results showed that compared to various variants and prior study, the supported format achieved much higher trendy average efficiency. A hybrid RF/immediate model (HRFLM) was developed by Mohan et al. (2019) to improve the HDPM's performance. In general, they found that the recommended method achieved precision, accuracy, sensitivity level, f-diploma, and strength as high as 88.1 percent, 90.1 percent, 90.8 percent, 90.1 percent, and 82.6 percent.

Gupta et al. (2020) have established a device data form using RF-based fully MLA and variable analysis of combined truths (FAMD). To determine the relevant abilities, the FAMD was used, and to guess the current situation, the RF was used. Using a degree of uniqueness, degree of sensitivity, and precision of up to 96.34%, 89.28%, and 96.76%, respectively, the recommended

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

technique outperformed other variations in terms of previous research study results.

EXISTING SYSTEM

Cardiovascular diseases (CVDs), which are heart-related infections, have become the most dangerous illness worldwide and in India during the past few years. They are the main cause of a staggering number of deaths worldwide. Therefore, a trustworthy, accurate, and attainable framework is required to analyse these diseases in time for appropriate treatment. To automate the analysis of vast and complex data, AI algorithms and methods have been used for many clinical datasets. Recently, many scientists have started using a few AI techniques to assist the healthcare sector and the experts in the investigation of heart-related ailments.

PROPOSED SYSTEM

In terms of global mortality, cardiovascular disease ranks first. Forecasting is difficult for doctors as it calls for a higher degree of predictive competence and skill. There may be a knowledge deficit, but data is abounding in the healthcare industry. Despite the abundance of data available online from healthcare systems, effective analytic tools are lacking, making it difficult to uncover hidden trends. Clinical effectiveness, together with cost and waiting time reduction, will unquestionably be enhanced by an automated method. This programme tries to foretell when a disease could strike by using data collected from Kaggle. Finding hidden patterns and estimating the visibility worth on a range are the goals of applying data mining algorithms to the dataset. It is quite difficult to assess and analyse using conventional methods because of the massive amounts of data required to forecast heart status. Our goal is to find a reliable method for predicting the occurrence of cardiovascular disease.

METHODOLOGY

Information Asset

The dataset used in this study to anticipate cardiac status was obtained from the UCI Artificial Intelligence database. It is possible to run machine learning algorithms on the databases that comprise UCI. This data collection is authentic. Including the proper 14 scientific factors, the dataset consists of 300 data instances. The dataset's scientific description is based on tests performed to diagnose cardiac issues,

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

such as the severity of hypertension, the kind of chest pain, the results of electrocardiograms, and others.

Formula Summary

Here we'll go over the two main algorithms that this system uses: i) the decision tree classification formula

the Support Vector Machine (SVM) algorithm.

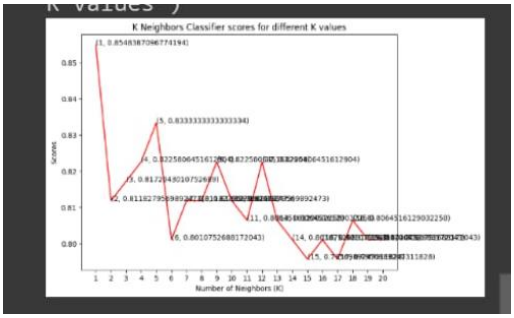
III. The K-next-door-neighbors method.

IV. The random forest.

The hope is that this will lead to a diagnosis of heart disease. Records in the datasets are divided into two groups: training and examination. Following data pre-processing, methodologies for data extraction from categories such as uninformed Bayes and decision trees were used. This section displays the final results of the categorization designs that were created using Python shows. To get the results, we need both training datasets and test sets of data.

RESULTS EXPLANATION

As part of our project, we have integrated a user interface (UI) with a database so that people can register and access their cardiac risk assessment results. Even without creating an account and logging in, users may still use the fast forecast button to anticipate a cardiac issue. The following step is for the consumer to input 13 characteristics, including personal information like age, gender, cholesterol level, etc. It will also train the dataset after delving into the data, and a design will be created according to the individual's specifics. The proportions of examination and training are 25% and75%, respectively. By choosing the develop design and heart disease threat percentage on our user interface (UI), individuals may access alternatives for symptom management and prevention when the percentage is more than 60%, indicating the presence of a potential heart issue. Users may view their past prediction data on our website by logging in using the same login details as when they predicted before. If the design anticipates a cardiac issue, the person might choose to communicate with the client directly by phone or email. Every piece of personal information is stored in a database and may be accessed at any time.



International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

Fig.1. K neighbors algorithm.

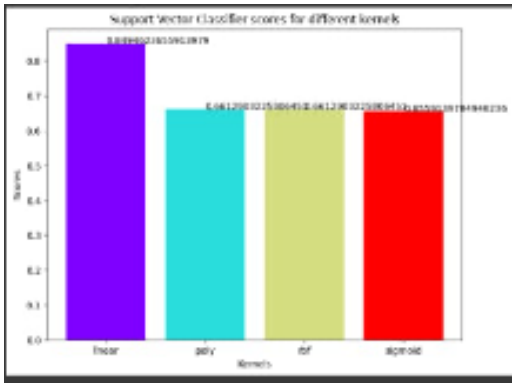


Fig.2. SVM classifier.

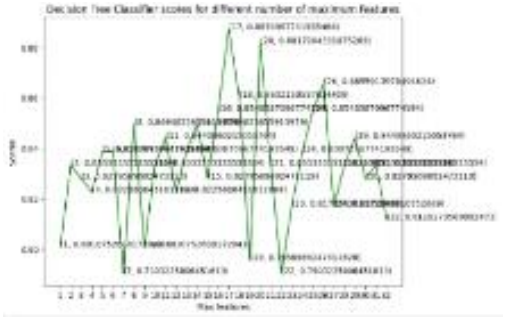


Fig.3. Decision tree algorithm.

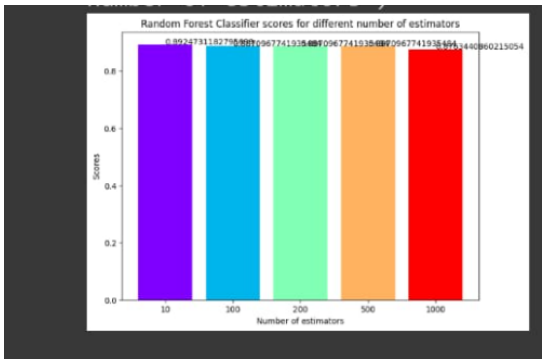


Fig.4. Random forest classifier.

The UCI repository's heart disorder dataset is processed using Python the usage of some of clustering and class algorithms. The number one objective is to discover and target all viable combos of attributes through the use of diverse algorithms. Next, the technique that plays the excellent in predicting the onset of heart disease at an early degree is selected.

CONCLUSION

As the number of fatalities caused by heart disease rises, the need of develop a system that can accurately and reliably predict heart problems grows. The primary goal of the research was to identify the most efficient ML algorithm for the detection of cardiac problems. This project uses the UCI maker learning repository dataset to compare the accuracy ratings of Logistic Regression, Random Woodland, and K Neighbours for cardiac issue prediction. This study's results show that the logistic regression algorithm is one of the most dependable algorithms for predicting cardiovascular disease, with an accuracy score of 89%. The accuracy of maker discovery algorithms is affected by the dataset that is utilised for training and screening. Cardiac disease prediction may make use of a variety of alternative devices finding approaches. Furthermore, logistic regression is an effective tool for dealing with binary classification issues, such as the prediction of cardiac problems. It is possible that decision trees may under perform randomly generated forests. It is also possible to apply set approaches and synthetic

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

semantic networks on the collected data. The outcomes may be improved by comparing and contrasting.

ACKNOWLEDGMENT

We thank CMR Technical Campus for supporting this paper titled "**HEART DISEASE IDENTIFICATION METHOD USING MACHINE LEARNING CLASSIFICATION IN E-HEALTHCARE**", which provided good facilities and support to accomplish our work. I sincerely thank our Chairman, Director, Deans, Head of the Department, Department Of Computer Science and Engineering, Guide and Teaching and Non- Teaching faculty members for giving valuable suggestions and guidance in every aspect of our work.

REFERANCES

[1]. Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.

[2]. Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.

[3]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[4]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150-154). IEEE.

[5]. Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "Analysis of data mining techniques for heart disease prediction." In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-5. IEEE, 2016.

[6].Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[7].Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 16, Issue. 2, 2024

2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[8]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer Science & Information Technology Journal (2014): 53-59.

[9]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 704-706. IEEE, 2015.