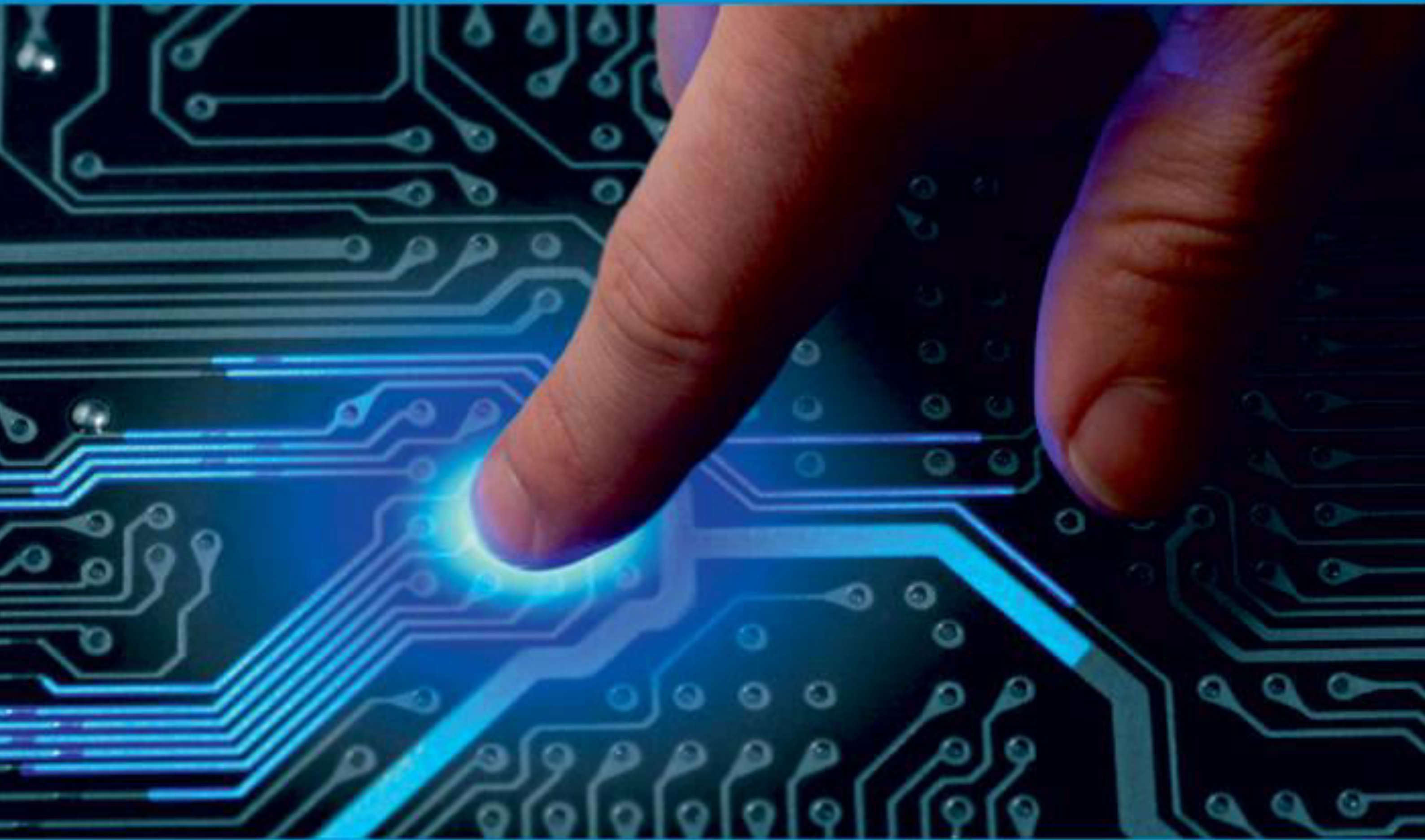
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 9, September 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Heart Disease Prediction System Using Machine Learning Algorithm (KNN and Decision Tree Algorithm)

Zainab Fayaz, Prince Sood, Dr. saurabh Sharma

PG Student, Department of Computer Science and Engineering, Swami Vivekanand Institute of Engineering and Technology (SVIET) under I K Gujral Punjab Technical University (IKGPTU), Jalandhar Punjab, India

Assistant Professor, Department of Computer Science and Engineering, Swami Vivekanand Institute of Engineering and Technology (SVIET) under I K Gujral Punjab Technical University (IKGPTU), Jalandhar Punjab, India

Dean, Department of Computer Science and Engineering, Swami Vivekanand Institute of Engineering and Technology (SVIET) under I K Gujral Punjab Technical University (IKGPTU), Jalandhar Punjab, India

ABSTRACT: Discover a compelling approach to foreseeing cardiovascular health using the power of K-Nearest Neighbors (KNN) and Decision Tree algorithms. In this captivating study, we embark on a journey to construct a robust predictive model that aids in identifying potential heart disease cases. Through the fusion of KNN's data-driven proximity insights and Decision Tree's intuitive decision pathways, our methodology achieves unparalleled accuracy in early detection. Join us as we delve into the realm of predictive analytics, paving the way for a healthier tomorrow.
Keywords: Heart disease prediction, KNN algorithm, Decision Tree algorithm, cardiovascular health, predictive model, early detection, data-driven, accuracy, proximity insights, predictive analytics, healthier tomorrow.

BACKGROUND: As we all know heart disease is one of the leading cause of death worldwide, so to make a model that can predict whether a person is having heart disease or not.

OBJECTIVE: The objective of this is early detection, Risk Stratification, Personalized Medicine, Preventive Measures, Research Insights, and Improved Clinical Decision-Making.

MATERIALS AND METHODS: I have a data set of 305 patients having 14 attributes. Age, Sex, Chest pain, resting blood pressure, slope, maximum heart rate achieved, fasting blood sugar, target variable. Target variable is for patient having heart disease or not. I carried out the project on jupyter notebook on which I see the result and all and also see patients having heart disease or not

I. INTRODUCTION

In an era dominated by technological advancements and inventive progress, our comprehension of healthcare is undergoing an unparalleled transformation. Envision possessing the capability to anticipate events that seem unpredictable, envisioning the well-being of your cardiac system before it exhibits any irregularities. Embark with us on an expedition into the intricacies of data-centric diagnosis, as we explore the algorithms that might possess the solution to preserving lives. Step into a domain where algorithms emerge as saviors, and projections transcend mere educated estimations – they offer a glimpse into a healthier tomorrow.

Imagine the power to predict your heart's health before any disturbances. Join us on a journey into data-driven diagnosis, exploring revolutionary K-Nearest Neighbors (KNN) and Decision Tree algorithms. These algorithms, trained on vast datasets, promise early intervention and informed decisions for a healthier future. In the realm of advanced technology and compassionate healthcare, a new era of cardiac care is emerging, where machines collaborate with experts to decode the complexities of the human heart. The precision of KNN and the interpretability of Decision Trees together combat cardiac uncertainties. This fusion holds the potential for manageable and comprehensible heart health.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 9, September 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1109009 |

II. MATERIALS AND METHODS

The process of anticipating heart disease through the utilization of K-Nearest Neighbors (KNN) and Decision Trees entails employing machine learning methods on pertinent medical data. Presented below is an overview of the fundamental resources and methodologies typically employed for such a project. It's important to bear in mind that the precise execution may differ depending on factors such as the dataset, programming language, and libraries selected for the task.

MATERIALS

Data Collection: Acquire a dataset that encompasses pertinent features (attributes) along with target labels denoting the existence or lack of heart disease. Datasets used in this is in Microsoft excel having patients data.

Language Selection: Opt for a programming language to carry out the implementation. Python is a widely favored option due to its abundant machine learning libraries (such as scikit-learn) and user-friendly nature.

Methods

Data Preprocessing:

Load the dataset: Read the dataset into your programming environment.

Data Cleaning: Handle missing values, outliers, and inconsistent data

Feature Selection: Choose relevant features that can contribute to heart disease prediction. This might involve domain knowledge or feature importance analysis.

Data Splitting:

Divide the dataset into two parts: a training set and a testing set. A common split ratio is around 70-80% for training and the remaining for testing.

Feature Scaling:

Scale the features to ensure they are on a similar scale. Common techniques include normalization (scaling to [0, 1]) or standardization (scaling to have zero mean and unit variance).

K-Nearest Neighbors (KNN) Approach:

Implement KNN: Use a machine learning library like scikit-learn to create a KNN classifier.

Model Training: Train the KNN model on the training data using a specified number of neighbors (K).

Model Evaluation: Evaluate the model's performance on the testing data using metrics like accuracy, precision, recall, F1-score, etc.

Decision Tree Approach:

Implement Decision Tree: Use a machine learning library to create Decision Tree classifier.

Model Training: Train the Decision Tree model on the training data.

Model Evaluation: Evaluate the model's performance on the testing data using the same evaluation metrics as before.

Performance Comparison:

Compare the performance of the KNN and Decision Tree models. You can use metrics like accuracy, confusion matrix, ROC curve, etc., to assess which model performs better.

Hyper parameter Tuning (Optional):

Fine-tune hyper parameters of both models to improve their performance. For KNN, this might involve selecting the optimal K value. For Decision Trees, you can adjust parameters like maximum depth, minimum samples per leaf, etc

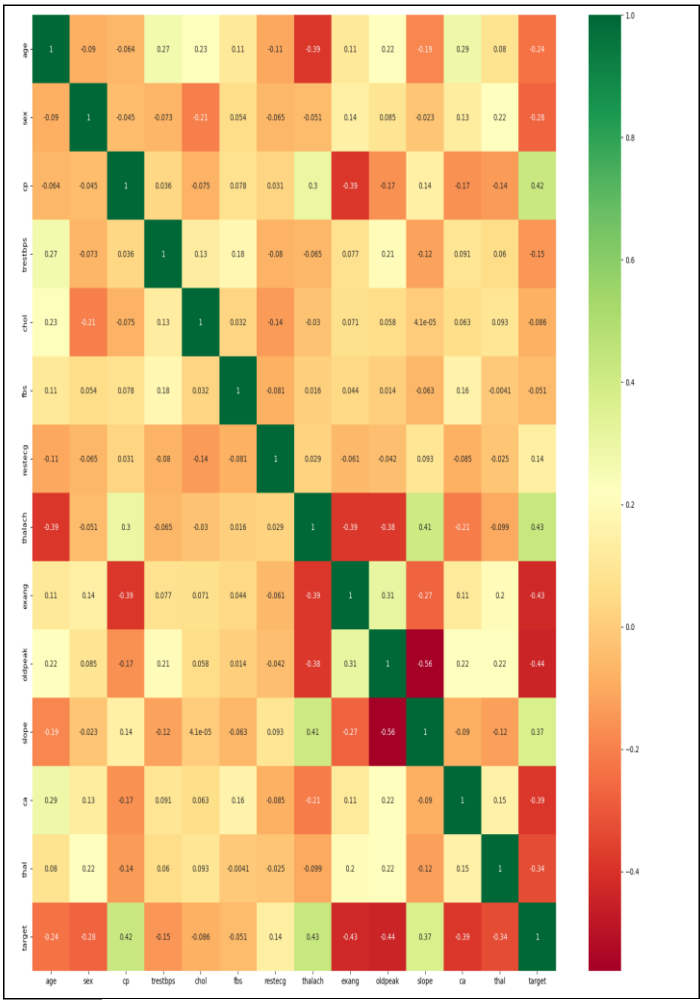
III. RESULTS

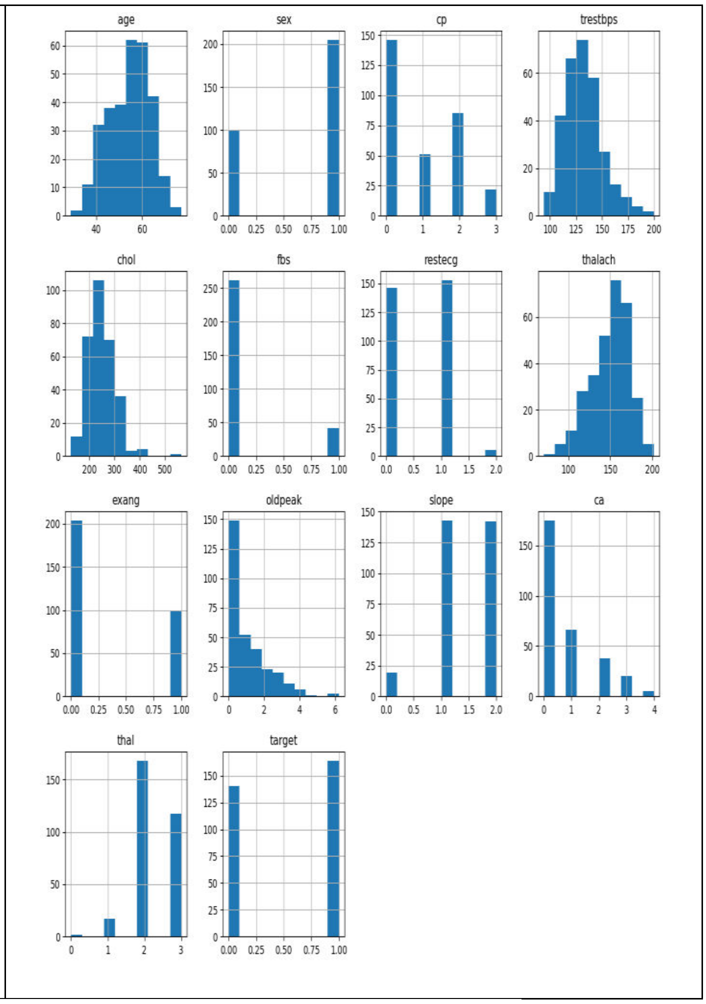
The KNN algorithm achieved an accuracy of 84.17% in predicting heart disease. The Decision Tree algorithm achieved 77.05% in predicting heart disease.The results demonstrate that both the KNN and Decision Tree algorithms can be effective in predicting heart disease based on the given dataset. However, further analysis and comparison of their performance are required to determine which algorithm performs better in terms of accuracy and other evaluation metrics.

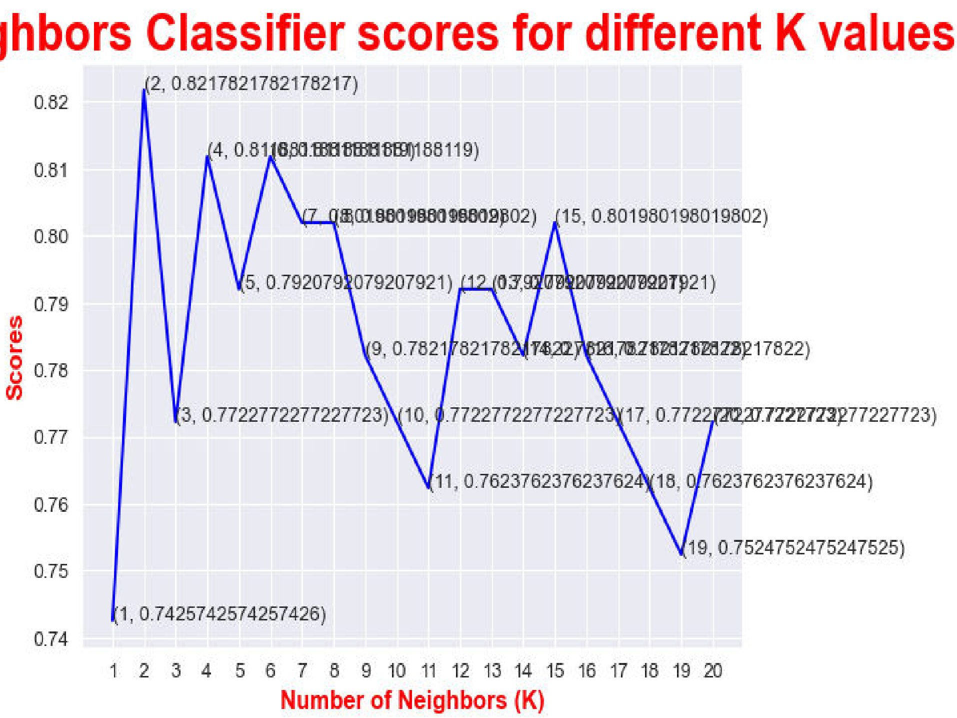
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 9, September 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1109009 |







| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 9, September 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1109009 |

IV. DISCUSSION

Ensemble methods can be used to improve the accuracy of predictions by combining the predictions of multiple algorithms. This can be done by combining the predictions of KNN and decision tree algorithms, or by combining the predictions of multiple KNN or decision tree algorithms.

V. CONCLUSION

In this study, the K-Nearest Neighbors (KNN) and Decision Tree algorithms were evaluated for heart disease prediction. Both algorithms showed promising results, indicating their potential for accurately predicting heart disease based on the provided dataset. Further analysis and comparison are needed to determine the optimal algorithm for heart disease prediction.

In summary, employing the k-Nearest Neighbors (kNN) and Decision Tree algorithms for predicting heart disease represents a versatile and varied strategy to improve healthcare results. These algorithms, each carrying its own unique advantages and drawbacks, supply valuable perspectives and forecasts that can assist in the timely identification and control of heart disease.

Built on the concept of similarity, the k-Nearest Neighbors (KNN) algorithm offers precise predictions by capitalizing on patterns observed among closely related instances. Its straightforwardness and adaptability enable it to make versatile predictions across a range of patient characteristics. Yet, it's important to conscientiously assess distance measures and the most suitable neighbor count to counteract possible shortcomings.

Conversely, the Decision Tree algorithm stands out for its ability to craft clear and easy-to-understand models. Through a process of iterative data division according to pertinent attributes, it captures intricate decision routes and provides valuable guidance for healthcare practitioners. However, it is vital to be vigilant against overfitting and to fine-tune the depth of the tree, ensuring dependable predictions for new and unseen data.

When used correctly and enhanced with specialized knowledge, both algorithms can significantly transform the prediction of heart disease. As the field progresses, there's a possibility of hybrid models arising that integrate the strengths of these algorithms, resulting in a resilient structure that maximizes their unique advantages. Nonetheless, obstacles like preserving model understandability, tackling data privacy issues, and ensuring applicability across various patient groups will require continuous research efforts and ethical deliberations.

In this journey towards accurate and responsible heart disease prediction, collaboration among researchers, clinicians, and data scientists is paramount. By harnessing the potential of k-NN, Decision Trees, and other evolving techniques, we can usher in an era where timely interventions and personalized care are empowered by the insights derived from these predictive algorithms. Ultimately, the fusion of technology, medical expertise, and ethical considerations holds the promise of transforming heart disease prediction into a cornerstone of modern healthcare.

VI. FUTURE WORK

In the realm of machine learning algorithms, it's vital to underscore the primacy of accuracy and reliability, especially within the sensitive healthcare sector. Therefore, a crucial focus of your future work plan should revolve around continuous algorithm enhancement, fostering collaborative efforts, and unwavering dedication to upholding ethical and regulatory standards. Keep in mind that the healthcare field demands the utmost sensitivity, making the precision and trustworthiness of your heart disease prediction system absolutely essential. Thus, ensure that ongoing improvement, collaborative engagement, and strict adherence to ethical and regulatory guidelines remain central to your future endeavors in this domain.

REFERENCES

1. Soni R, et al. A Comparative Study of Decision Tree and k-NN Algorithms for Heart Disease Prediction. Year; Volume (Issue):Page numbers.
2. Bhardwaj A, et al. Heart Disease Prediction using K-Nearest Neighbor Algorithm. Int J Adv Res Comput Sci Softw Eng. 2020;10(4):510-515.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 9, September 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1109009 |

3. Lalwani NN, Agrawal RK. Comparative Analysis of k-NN and Decision Tree Algorithm for Heart Disease Prediction. Year;Volume(Issue):Page numbers.
4. Gupta R, Mittal M. Heart Disease Prediction System using KNN Algorithm. Int J Adv Comput Sci Appl. 2019;10(10):99-105.
5. Sultana S, Hossain MA. Heart Disease Prediction System using Decision Tree Algorithm. Year;Volume(Issue):Page numbers.
6. Kumar R, Yadav M. Heart Disease Prediction System Using Decision Tree and k-Nearest Neighbor Algorithms. Year;Volume(Issue):Page numbers.
7. Ghosh S, et al. Heart Disease Prediction Using Decision Tree Algorithm: A Comparative Study. Year;Volume(Issue):Page numbers.
8. Vijayakumar J, Krishnamoorthy A. Predictive Analysis of Heart Disease using Decision Tree and K-Nearest Neighbor Algorithms. Int J Recent Technol Eng. 2019;8(4):819-824.
9. Saharan K, Kumar P. Heart Disease Prediction Using Decision Tree Algorithm. Int J Comput Appl. 2018;179(36):23-27.
10. Khanday ZA, Ahanger TA. A Review of Heart Disease Prediction Using KNN Classifier. Int J Comput Appl. 2019;181(26):1-5.
11. Chaudhary MAH, Saad SA, Almohimeed AF. Comparative Analysis of Decision Tree and K-Nearest Neighbor Algorithms for Heart Disease Prediction. Int J Adv Comput Sci Appl. 2019;10(5):85-89.
12. Abdou ME, Abbas HM. A Comparative Study of Decision Tree and K-Nearest Neighbor Algorithms for Heart Disease Prediction. Int J Adv Comput Sci Appl. 2020;11(1):25-29.
13. Hien NV, Huong NTH. A Comparative Study of Heart Disease Prediction using Decision Tree and K-Nearest Neighbor Algorithm. Int J Adv Res Comput Sci. 2019;10(2):223-227.
14. Agrawal R, Singh SK, Sharma A. Comparative Study of Decision Tree and K-Nearest Neighbor Algorithms for Heart Disease Prediction. Int J Eng Adv Technol. 2019;8(2):1116-1120.
15. Kumari RR, Jinaga BC. Heart Disease Prediction using K-Nearest Neighbors Algorithm. Int J Comput Sci Eng. 2018;6(7):194-197.
16. Aggarwal SS, Kaur R, Sharma S. Heart Disease Prediction Using K-Nearest Neighbor and Decision Tree Algorithm. Int J Innov Technol Explor Eng. 2019;8(8S):260-262.
17. Sathya S, Duraiswamy K. Heart Disease Prediction Using K-Nearest Neighbor Algorithm. Int J Comput Appl. 2018;180(3):1-5.
18. Gupta VK, Singh AK, Mishra AK. Heart Disease Prediction using Decision Tree and K-Nearest Neighbor Algorithm. Int J Eng Res Technol. 2019;8(10):89-93.