International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 06 | June 2022 www.irjet.net p-ISSN: 2395-0072

PREDICTING THE RISK OF HAVING HEART DISEASE USING MACHINE LEARNING TECHNIQUES

Chandu D[1], Ch Sivakumar[2], Darshan Vinayak S[3], Dereddy Parthasaradhi Reddy[4], Karthik M D[5]

[1,2,3,4,5]*Student, Dept. of Computer science and Engineering, Dayananda Sagar University, Karnataka, India*

Abstract - *Machine Learning is employed across many spheres round the world. The healthcare industry isn't any exception. Machine Learning can play a vital role in predicting the presence/absence of Locomotor disorders, heart diseases, and more. the target of the proposed model is to predict the danger of getting cardiopathy using machine learning technics. Machine learning is widely used nowadays in many business applications like e-commerce and plenty of more. Prediction is one amongst the areas where this machine learning is employed, our topic is about the prediction of cardiopathy by processing patient datasets and data of patients to whom we want to predict the possibility of occurrence of cardiovascular disease. Such information, if predicted well before, can provide important insights to doctors who can then adapt their diagnosis and treatment per-patient basis.*

Key Words: **Heart disease prediction, Classification, Regression, Machine learning and**

1. INTRODUCTION

The objective of the proposed model is to predict the chance of getting heart condition using machine learning techniques. Machine learning is widely used nowadays in many business applications like e-commerce and lots of more. Prediction is one in all the areas where this machine learning is employed, our topic is about the prediction of cardiopathy by processing a patient's dataset and data of patients to whom we want to predict the prospect of occurrence of heart condition.

Heart disease could be a term covering any disorder of the guts. Heart diseases became a significant concern to handle as studies show that the amount of deaths thanks to heart diseases has increased significantly over the past few decades in India, in fact, it's become the leading reason behind death in India.

A study shows that from 1990 to 2016 the death rate thanks to heart diseases increased by 34 percent from 155.7 to 209.1 deaths per one lakh population in India.

Thus, preventing heart diseases has become quite necessary. Good data-driven systems for predicting heart diseases can improve the whole research and prevention process, ensuring that more people can live healthy lives. this can be where Machine Learning comes into play. Machine Learning helps in predicting heart diseases, and therefore the predictions made are quite accurate.

We can train our prediction model by analyzing existing data because we already know whether each patient has cardiopathy. This process is additionally referred to as supervision and learning. The trained model is then accustomed predict if users suffer from heart condition.

2. PROPOSED SYSTEM

Our system is a website-based machine learning application trained on a dataset from Kaggle. The admin inputs the required attributes to get the prediction for that patient. The model will determine the probability of heart disease.

We have tested the following sixteen algorithms:

2.1 CLASSIFICATION ALGORITHMS

1. Logistic Regression
2. Naive Bayes
3. Support Vector Machine
4. K-Nearest Neighbors
5. Decision Tree
6. XG Boost

2.2 REGRESSION ALGORITHMS

1. Logistic Regression
2. Polynomial Regression
3. Support Vector Regression with RBF kernel
4. Support Vector Regression with Linear kernel
5. Support Vector Regression with Poly kernel
6. Decision Tree Regression
7. Bayesian Ridge
8. Lasso
9. Ridge
10. Random Forest Regression

The algorithms have been trained and tested using the dataset obtained from Kaggle. 80% of data is used for training, while the rest is used for testing. Additional processing is done on the dataset to make the training process efficient and make the model accurate. The algorithms were tested and judged based on their accuracy,

2. DATA SET

3.1 DATA SET - CLASSIFICATION

Feature	Variable Type	Value Type
Age	Objective Feature	int
Gender	Objective Feature	Categorical Code
Chest Pain	Subjective Feature	Categorical Code
Restingblood pressure	Examination Feature	int
Cholesterol	Examination Feature	Categorical Code
Fasting blood pressure	Examination Feature	Categorical Code
Resting ECG	Examinati onFeature	Categorical Code
Maximum heart rate achieved	Subjective Feature	int
Exercise induced angina	Subjective Feature	binary
Old peak	Examinati onFeature	Categorical Code
Number of major vessels	Examinati onFeature	Categorical Code
slope	Examinati onFeature	Categorical Code
Thalassemia	Examinati onFeature	Categorical Code

Table -1: Dataset Attributes-Classification

3.2 DATA SET - REGRESSION

Feature	Variable Type	Value Type
Age	Objective Feature	Int
Gender	Objective Feature	Categorical Code
Total Cholesterol	Examination Feature	int
High Density Lipid	Examination Feature	int
S Blood Pressure	Examination Feature	int
Smoke	Subjective Feature	binary
Blood pressure medication	Subjective Feature	binary
Diabetic	Examination Feature	binary

Table -2: Dataset Attributes-Regression

4. PROCESS

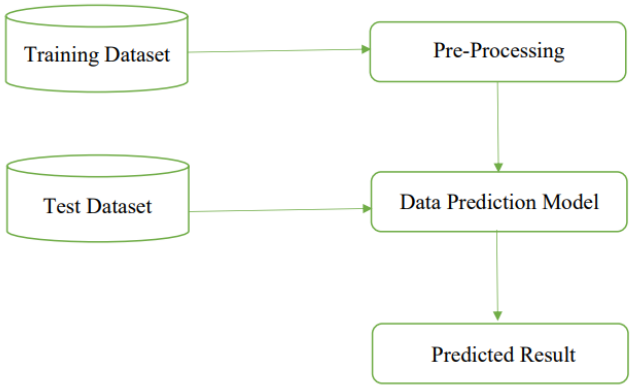

Figure -1: Flow Chart

5. RESULT

After implementing classification algorithms Linear Regression gave 86.34%, Naive Bayes gave 85.37%, Support Vector Machine 83.90%, K-Nearest Neighbors 72.20%, Decision Tree 99.50%, XG Boost 99.02% by observing these data we can say that Decision tree is the best accurate results

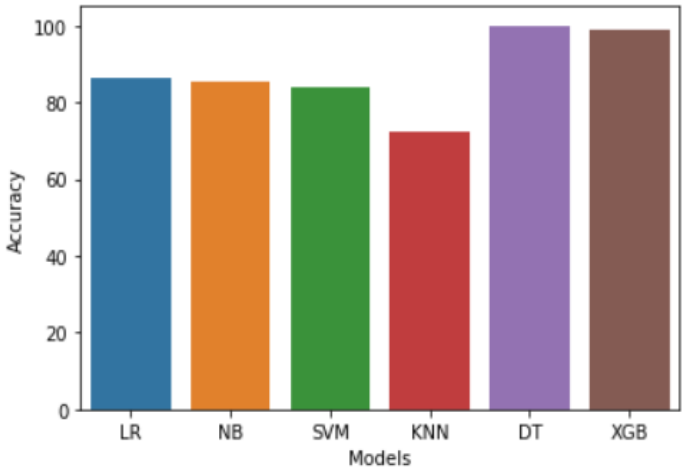

Figure -2: Classification Accuracy Chart

After implementing Regression algorithms Logistic Regression 74.54%, Polynomial Regression 85.54%, Support Vector Regression with RBF kernel 45.10%, Support Vector Regression with Linear kernel 70.93%, Support Vector Regression with Poly kernel 45.69%, Decision Tree Regression 74.32%, Bayesian Ridge 74.54%, Lasso 62.74%, Ridge 74.54%, Random Forest Regression 84.68%

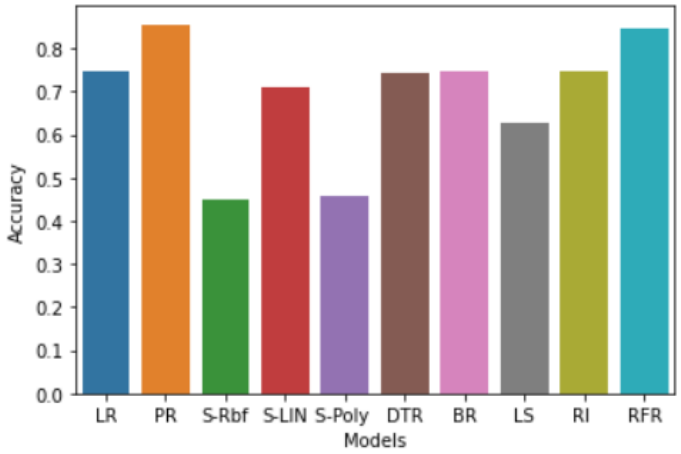

Figure -2: Regression Accuracy Chart

3. CONCLUSIONS

In this project, we proposed both classification and regression algorithms to improve the prediction of machine learning models. The goal for the regression is to predict risk percentage of patients having heart disease. For the classification algorithm we have used 13 features and for regression 8 features, since random forest algorithm has been very effective and has provided us with the most accuracy compared to all the other algorithms such as linear regression, polynomial regression, SVM with RBF kernel, SVM with linear kernel, SVM with polynomial kernel. Our project can be employed in many real-life applications or in other medical diagnoses to analyze great amounts of data and identify the risk factors involved in different diseases. Our main limitation is the difficulty to extend these findings to heart disease due to the small sample size. For future developments, we plan to apply our method to a larger dataset and perform the analysis of some other diseases with different feature selection techniques.

REFERENCES

[1] Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta, "*Heart Disease Prediction using Machine Learning Techniques*", in International Conference on Advances in Computing, Communication Control and Networking, 2020.

[2] Prathamesh Keni, Pratik Poshe, Kaustubh Latake, Prof. Rovina Dbritto, "Heart Disease Prediction using Machine Learning", in International Research Journal Of Engineering And Technology, 2022.

[3] Aditi Gavhane, Isha Pandya, Gouthami Kokkula, Prof. Kailas Devadkar, "*PREDICTION OF HEART DISEASE USING MACHINE LEARNING*", in International conference on Electronics, Communication and Aerospace Technology, 2018.

[4] M. Snehith Raja, M. Anurag, Ch. Prachetan Reddy, NageswaraRao Sirisala, "*MACHINE LEARNING BASED HEART DISEASE PREDICTION SYSTEM*", in International Conference on Computer Communication and Informatics, 2021.

[5] Dr. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. Rohith Sai, R. Sai Suraj, "*HEART DISEASE PREDICTION USING HYBRID MACHINE LEARNING*" in the international conference on inventive computation technologies, 2021.

BIOGRAPHIES

Mr. Chandu D is pursuing Bachelor of Technology degree in department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. CH Sivakumar is pursuing Bachelor of Technology degree in department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Darshan Vinayak S is pursuing Bachelor of Technology degree in department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Dereddy Parthasaradhi Reddy is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Karthik MD is pursuing Bachelor of Technology degree in department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.