## Heart Disease Prediction using Machine Learning Algorithms

A Mathematical Programing Report

Submitted in the partial fulfillment of the requirements for the award of the degree of

# Bachelor of Technology in

Department of Computer Science and Engineering

By

K.V.R.L KRISHNA AJAY(2010030077)

Medhilesh(2010030551)

Ramakrishna(2010030127)

Tejeshwar reddy(2010030169)

under the supervision of

# **Dr. Anal Paul**

# Assistant professor



Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

March, 2022

**Declaration**

The Social Internship Report entitled “Heart Disease Prediction” is a record of bonafid work of K.V.R.L KRISHNA AJAY(2010030077),Medhilesh(2010030551)

,Ramakrishna(2010030127),Tejeshwar reddy(2010030169) submitted in partial fulfillment for the award of B.Tech in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/ University/ Institute.

**Certificate**

This is to certify that the Social Internship Report entitled “Heart Disease Prediction” is being submitted K.V.R.L KRISHNA AJAY(2010030077),Medhilesh(2010030551)

,Ramakrishna(2010030127),Tejeshwar reddy(2010030169) submitted in partial fulfillment for the award of B.Tech in CSE to the K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/ University/Institute.

## Signature of the Supervisor

Dr. Anal paul

Assistant Professor

## Signature of the HOD Signature of the External Examiner

**ACKNOWLEDGEMENT**

First and foremost, we thank the lord almighty for all his grace & mercy showered upon us, for completing this Social Internship successfully.

We take grateful opportunity to thank our beloved Founder and Chairman who has given constant encouragement during our course and motivated us to do this Social Internship. We are grateful to our Principal **Dr. L. Koteswara Rao** who has been constantly bearing the torch for all the curricular activities undertaken by us.

We pay our grateful acknowledgement & sincere thanks to our Head of the Department **Dr. Chiranjeevi Manike** for his exemplary guidance, monitoring, and constant encouragement throughout the course of the Social Internship. We thank Mrs. Anuradha of our department who has supported throughout this Social Internship holding a position of supervisor.

We whole heartedly thank all the teaching and non-teaching staff of our department without whom we won’t have made this Social Internship a reality. We would like to extend our sincere thanks especially to our parent, our family members and friends who have supported us to make this Social Internship a grand success.

**ABSTRACT**

Titanic disaster occurred 100 years ago on April 15, 1912, killing about 1500 passengers and crew members. The fateful incident still compel the researchers and analysts to understand what can have led to the survival of some passengers and demise of the others. With the use of machine learning methods and a dataset consisting of 891 rows in the train set and 418 rows in the test set, the research attempts to determine the correlation between factors such as age, sex, passenger class, fare etc. to the chance of survival of the passengers. These factors may or may not have impacted the survival rates of the passengers. In this research paper, various machine learning algorithms namely Logistic Regression, Naive Bayes, Decision Tree, Random Forest have been implemented to predict the survival of passengers. In particular, this research work compares the algorithm on the basis of the percentage of accuracy on a test dataset.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| S.No | Topic |
| 1 | Introduction |
| 2 | Literature Survey |
| 3 | Methodology and Implementation |
| 4 | Output |
| 5 | Results Discussion |
| 6 | Conclusion |
| 7 | Future Work |
| 8 | Social Contribution |
| 9 | References |

**Introduction**

This is the legendary Titanic ML competition – the best, first challenge for you to dive into ML competitions and familiarize yourself with how the Kaggle platform works.

The competition is simple: use machine learning to create a model that predicts which passengers survived the Titanic shipwreck.

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

In this challenge, we ask you to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc).

In this competition, you’ll gain access to two similar datasets that include passenger information like name, age, gender, socio-economic class, etc. One dataset is titled `train.csv` and the other is titled `test.csv`.

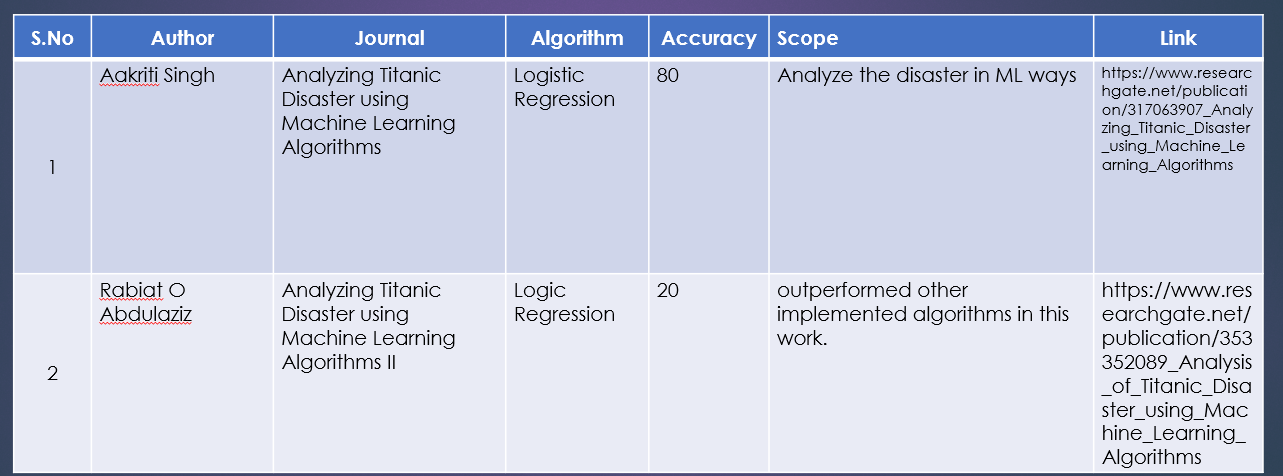
Train.csv will contain the details of a subset of the passengers on board (891 to be exact) and importantly, will reveal whether they survived or not, also known as the “ground truth”.

The `test.csv` dataset contains similar information but does not disclose the “ground truth” for each passenger. It’s your job to predict these outcomes.

Using the patterns you find in the train.csv data, predict whether the other 418 passengers on board (found in test.csv) survived.

Check out the [“Data” tab](https://www.kaggle.com/c/titanic/data) to explore the datasets even further. Once you feel you’ve created a competitive model, submit it to Kaggle to see where your model stands on our leaderboard against other Kagglers.

**Literature Survey**



**Methodology and Implementation**

**Random Forest:**

Random Forest is a supervised learning algorithm. Random forest can be used for both

classification and regression problems, by using random forest regressor we can use random

forest on regression problems. But we have used random forest on classification in this project

so, we will only consider the classification part.

Random Forest pseudocode

1. Randomly select “k” features from total “m” features.

Where k << m

2. Among the “k” features, calculate the node “d” using the best split point.

3. Split the node into daughter nodes using the best split.

4. Repeat 1 to 3 steps until “l” number of nodes has been reached.

5. Build Forest by repeating steps 1 to 4 for “n” number times to create “n” number of trees.

Random forest prediction pseudocode 1. Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target) 2. Calculate the votes for each predicted target. 3. Consider the high voted predicted target as the final prediction from the random forest algorithm.

Graphical user interface, text, application

Description automatically generated

Pruning the depth of the trees

Text

Description automatically generated

Accuracy score of Random Forest is 86.9%

Confusion Matrix

Chart

Description automatically generated

Precision score

Graphical user interface, text, application

Description automatically generated

Recall

Graphical user interface, application

Description automatically generated

F-Score

Graphical user interface, text, application

Description automatically generated

False negative rate

Graphical user interface, text

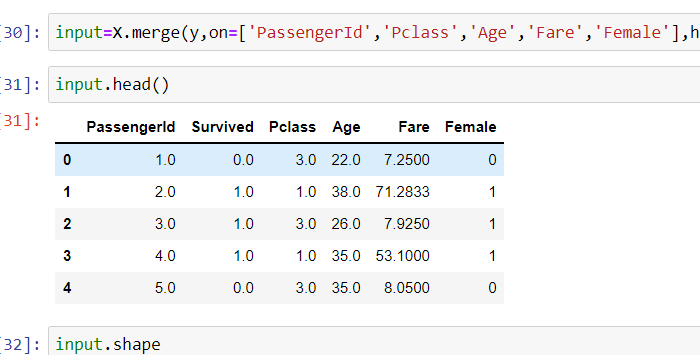
Description automatically generated with medium confidence

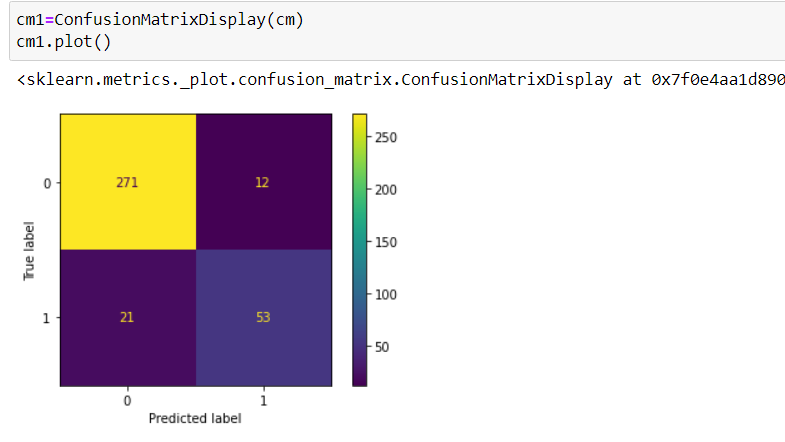
Diagram

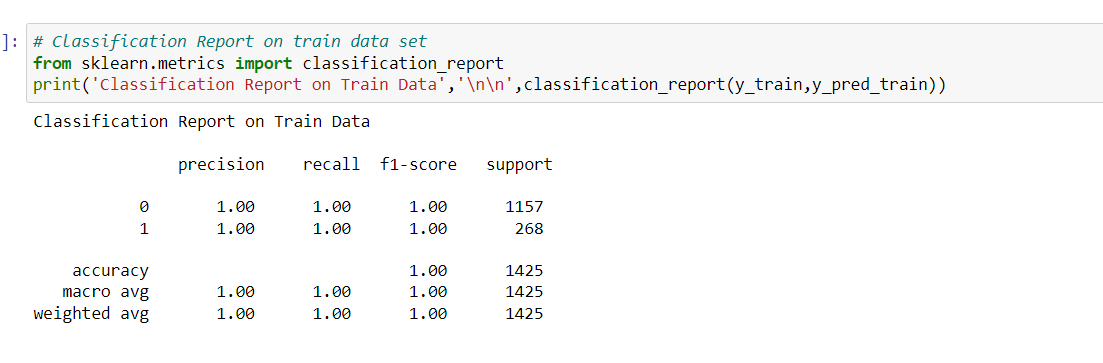
Description automatically generated

Random Forest algorithms are used for classification as well as regression. It creates a tree for the data and makes prediction based on that. Random Forest algorithm can be used on large datasets and can produce the same result even when large sets record values are missing. The generated samples from the decision tree can be saved so that it can be used on other data. In random forest there are two stages, firstly create a random forest then make a prediction using a random forest classifier created in the first stage.

**Output:**







**Results Discussion**

This project aims to know whether the patient has heart disease or not. The records in dataset are divided into training and testing sets. After preprocessing data, the data classification techniques namely Naïve Bayes, KNN, Decision Tree, Linear Regression and Random Forest Algorithms were applied. The project involved analysis of the heart disease patient dataset with proper data processing. Then the 5 models were trained and tested, out of these 5 models Random Forest Algorithms were giving maximum scores.

**Conclusion**

The overall objective of our project is to predict accurately with a smaller number of tests and attributes the presence of heart disease. In this project, fourteen attributes are considered which form the primary basis for tests and give accurate results. Many more

input attributes can be taken but our goal is to predict with a smaller number of attributes and faster efficiency to predict the risk of having heart disease at a particular age span. Five data mining classification techniques were applied namely K-Nearest Neighbor, Naive Bayes, Decision Tree, Random Forest & Logistic Regression. It is shown that Random Forest has better accuracy than the other techniques.

**Future Scope**

We are planning to introduce an efficient disease prediction system to predict the heart

disease with better accuracy using Support Vector Machine (SVM). Our project aims to

provide a web platform to predict the occurrences of disease based on various symptoms. The user can select various symptoms and can find the diseases with their probabilistic figures. Our project can be improved by implementing medicine suggestion to the patient along with the results. We can implement feedback from the experienced doctors who can give their views and opinions about certain medicines /practices done by the doctor on the patient. We can implement a live chat option where the patient can chat with a doctor available regarding medication for the respective result for their symptoms. Our project could be used as a training tool for Nurses and Doctors who are freshly introduced in the field related to heart diseases. The patient can have a choice in choosing the medicines he/she should take to have a healthier life. Moreover, if implemented on a large scale it can be used in medical facilities like hospital, clinics where a patient wouldn’t have to wait in long queues for treatment if he is feeling symptoms related to heart disease

**Social Contribution**

* We have implemented a friendly website through flask.
* Through the details that have entered it predicts weather he had heart disease or not .
* It helps him/her to contact the doctor through our prediction.
* In this way people’s lives can be saved through our project.

**References**

* M. K. Awang and F. Siraj, “Utilization of an artificial neural network in the prediction of heart disease,” Int. J. Bio-Science Biotechnology, vol. 5, no. 4, pp. 159–165, 2013.
* P. Selvakumar and S. P. Rajagopalan, “A survey on neural network models for heart disease prediction,” J. Theor. Appl. Inf. Technol., vol. 67, no. 2, pp. 485–497, 2014.
* Methaila, P. Kansan, H. Arya, and P. Kumar, “Early Heart Disease Prediction Using Data Mining Techniques,” vol. 6956, no. October, pp. 53–59, 2014.
* S. F. Dessai, “Intelligent Heart Disease Prediction System Using Probabilistic Neural Network,” no. 5, pp. 38–44, 2013.
* T. Karayilan and Ö. Kiliç, “Prediction of Heart disease using neural network,” in 2nd International Conference on Computer Science and Engineering, UBMK 2017, 2017, pp. 719–723.
* M. Mardiyono, R. Suryanita, and A. Adnan, “Intelligent Monitoring System on Prediction of Building Damage Index using Neural-Network,” TELKOMNIKA (Telecommunication Comput. Electron. Control., vol. 10, no. 1, p. 155, 2015.
* N. Guru, A. Dahiya, and N. Rajpal, “Decision support system for heart disease diagnosis using Neural Network,” 2007.
* J. S. Singh Navdeep, “No Title,” Intell. Hear. Dis. Predict. Syst. using CANFIS Genet. Algorithm, Int. J. Biol. Med. Sci., no. International Journal of Advance Research, Ideas and Innovations in Technology© 2018, www.IJARIIT.comAll Rights, p. Reserved Page | 987, 2008.
* R. G. S. Rajkumar Asha, “No Title,” Diagnosis Hear. Dis. Using Datamining Algorithm, vol. Issue 10 V, no. Global Journal of Computer Science and Technology, p. Page 38, 2010.
* S. Radhimeenakshi and G. M. Nasira, “Prediction of Heart Disease using Neural Network with Back Propagation,” Int. J. Data Min. Tech. Appl., vol. 4, no. 1, pp. 19– 22, 2016. “Prediction of Heart Disease using Artificial Neural Network,” VFAST Trans. Softw. Eng., pp. 102–112, 2019.