Computer Science Department

Final Project: Heart Attack Prediction

Course: Introduction to AI (CS156)

Course Instructor: Yulia Newton, Ph.D.

Student: Preet LNU

Introduction/Background

A heart attack is a medical emergency. A heart attack usually occurs when a blood clot blocks blood flow to the heart. Without blood, tissue loses oxygen and dies. Symptoms include tightness or pain in the chest, neck, back, or arms, as well as fatigue, lightheadedness, abnormal heartbeat, and anxiety. I will be using AI to predict heart attack for the patient by analyzing data like cholesterol, heart beats, blood pressure, etc. I will be using a dataset which has observations from different patients.

Dataset description

This dataset has data about many factors that leads to a heart attack. I will be using a .csv file from this link which I was able to download. This dataset has 303 observations with 14 attributes.

All the independent variables are numeric.

a. Age : Age of the patient
b. Sex : Sex of the patient
c. exang: exercise induced angina (1 = yes; 0 = no)
d. ca: number of major vessels (0-3)
e. cp : Chest Pain type chest pain type
f. trtbps : resting blood pressure (in mm Hg)
g. chol : cholestoral in mg/dl fetched via BMI sensor
h. fbs : (fasting blood sugar > 120 mg/dl) (
i. rest_ecg : resting electrocardiographic results
j. thalach : maximum heart rate achieved
k. thalachh: maximum heart rate achieved
l. exng: exercise induced angina (1 = yes; 0 = no)
m. OldPeak: previous peak

Dependent variable:

a. Output(Binary Categorical) : the result that will determine if the patient is at risk for heart attack or not.(yes = 1, no = 0). We will be aiming to get this variable and based on the variable this problem seems like a classification problem. The number of classes will be 2(Class 0 and Class1).

Methodology

Algorithm

The machine learning model I used in this project was Decision Trees.Decision Tree algorithm is a hierarchical tree which can be traversed to make a decisions, they make decisions top to bottom, by splitting nodes into sub-nodes based on some criteria. Most decision trees are binary (bifurcating splits) but not a requirement.

I decided to use a decision tree to predict stroke risk because decision trees work well in making complex decisions, and stroke risk can be a very complex problem with various deciding factors.

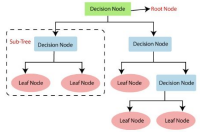

Figure 1: This figure shows the basic structure of a decision tree.
https://addepto.com/decision-tree-machine-learning-model

Application to the project

With the use of training data, I was able to plot a decision tree which contains all the decision taken to provide the output and to predict the heart attack.

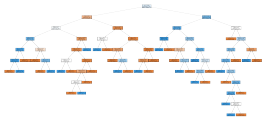

Figure 2: This is the decision tree I got.

Analysis and Results

Validation/cross-validation results

The model has an average heart attack prediction accuracy of roughly 78% across 5 folds of cross-validation. This value might have been greater with an actual validation set or more data entries.

```
Individual cross-validation accuracies: [0.88 0.78 0.75 0.67 0.83]
Mean cross validation accuracy: 0.7806122448979591
```

Figure 3: accuracy for 5 fold cross-validation.

Training results

The training accuracy of the model was 100%.

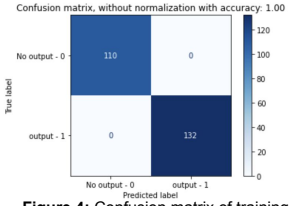

Figure 4: Confusion matrix of training result in decision tree.

Test results

Test Data accuracy for the model was 77%.

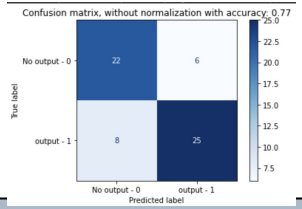

Figure 5: confusion matrix of test result in decision tree.

Summary/Conclusions

My project's overall outcomes could definitely be improved. Given that it is a binary classification problem, the current issue is not necessarily one that is difficult to forecast. 78 % cross-validation accuracy and 77% test data accuracy aren't as great as they could be when compared to comparable datasets I've utilized in the past. Despite this, there aren't many modifications that could materially alter the project's outcomes.

If I had more time, I would like to look into the potential contribution of each independent variable to the likelihood of suffering a heart attack. For instance, I would like to know how cp status is relative to another independent variable, such as age. This might be displayed in a graph or table with values for each independent

Key References

Dev@kodzilla.pl. (2022, October 10). *Decision tree machine learning model*. Addepto. Retrieved December 2, 2022, from https://addepto.com/decision-tree-machine-learning-model

Acknowledgements

I would like to thank Professor Newton for giving me this opportunity and to introduce me to the topic of machine learning and artificial intelligence.