Heart attack prediction

Abstract

This paper attempts to build and train two different types of classifiers, Random Forest and XGBoost classifiers, on medical data from the Heart Attack Analysis & Prediction Dataset found on Kaggle, to be able to predict the risk group a patient belongs to (high risk or low risk), based on their medical data. Seven classifiers are built and trained on different combinations of features, and finally a comparison of model performance is done to evaluate which model is the best.

The best performing model, XGBmodel1, was trained on all 13 features in the dataset and had an accuracy of 0.787 (78,7%). In general, both Random Forest and XGBoost performed well when trained and tested on all features in the dataset, however, XGBoost classifier models performed a little better in almost all conditions.

Introduction

Heart attack and cardiovascular disease

Cardiovascular disease is the cause of many deaths, especially in the US. Cardiovascular disease is a common term for all disease that affect the heart or blood vessels.

Coronary heart disease, sometimes referred to as coronary artery disease, is a common type of heart disease where build-up of fat damages the coronary arteries, which supply blood flow to the heart muscle. At such damaged areas, the cells in the blood that helps clotting to stop wounds from bleeding (called platelets) can stick to the damaged areas inside the coronary arteries, which causes a blockage of blood flow. This can then lead to ischemia (lack of oxygen to the cells of the heart muscle) or myocardial infarction (commonly known as heart attack) (Torpy et al., 2009). Thus, myocardial infarction, or a heart attack, occurs when the blood flow to the heart muscle is suddenly blocked.

According to a report from the American Heart Association, published on *Circulation*, there are annually 805,000 heart attacks in the American population, where 200,000 are recurring and 605,000 are first attacks (Tsao et al., 2023). Furthermore, about 12% of patients who experience heart attacks will die from it (Krumholz et al., 2019). Those are quite high numbers, however, those who do not die immediately from a heart attack will likely experience the after-effects, as the heart has already taken some damage. According to Mehta et al. (2016) there are some differences between men and women in the occurrence of death and other heart problems, after experiencing a heart attack. Regardless of age, 26% of women die within a year of a first heart attack, while only 19% of men will die in the same time span. Furthermore, within 5 years of a first heart attack, 47% of women die, develop heart failure, or have a stroke, while the same happens to only 36% of men (Mehta et al., 2016).

Some of the most common risk factors of coronary heart disease are: age (being older than age 40 for men and 45 for women), family history of coronary heart disease, smoking, hypertension (high blood pressure), diabetes, obesity, unhealthy cholesterol levels, low physical activity, and accumulation of abdominal fat (Torpy et al., 2009).

Heart attack prediction through classifier models

Knowing the risk factors and premonitory symptoms of myocardial infarction, there is a possibility to attempt prediction of the risk levels of a patient using a classifier that is trained on medical data. A classifier is a programmed model which can be trained on a number of features (variables) in a dataset (including the outcome variable) and learn the patterns in the data. The model can then be tested on a subset of the data (excluding the outcome variable), to assess the accuracy of the model.

Decision trees

There are different types of classifiers, and they work in different ways. The Decision tree classifier is, much like the name suggests, structured like a tree. Figure 1 is an example of a Decision tree. As illustrated in the figure below, each internal node represents a feature (variable), each branch represents a decision rule, and each leaf node represents an outcome. The decision tree is grown by adding question nodes using labelled training. Once the trees are constructed, they classify new instances by sorting each item into a class by following the path through the tree, starting at the root node in the top of the hierarchy. Decision trees are flexible and can make accurate predictions if trained on data of high quality (Kingsford & Salzberg, 2008).

Figure 1

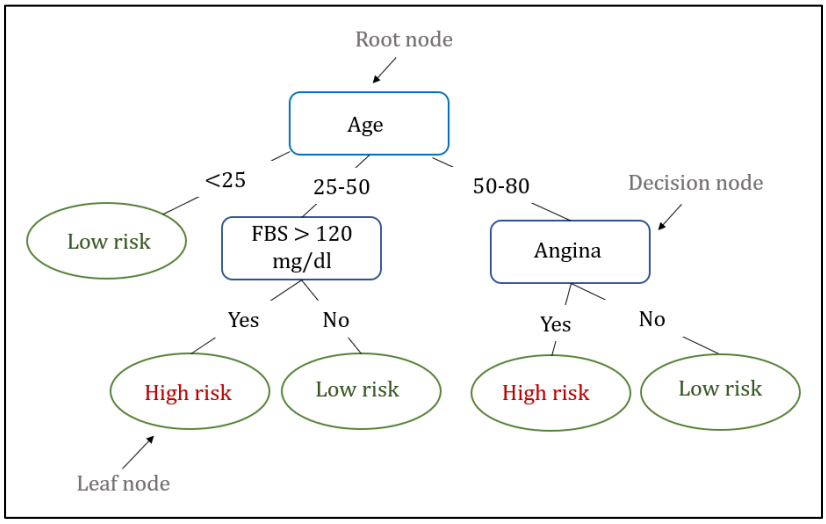


Figure 1: Example of a Decision tree. FBS : Fasting blood sugar.

Random Forest

Random Forest is an ensemble method that uses multiple decision trees that are fit to various sub-samples of the dataset and is therefore often more accurate than a single decision tree (*Sklearn.Ensemble.RandomForestClassifier*, n.d.). In this approach, many decision trees are grown by a randomized tree-building algorithm, and each tree is different as only a small random subset of features are included when the question at each node is chosen. In addition, the training set is sampled to produce a modified version of the original training set, with the same size, but with some items included more than once. This allows for slightly different trees for every run (Kingsford & Salzberg, 2008).

When a trained Random Forest classifier is introduced to new data, each decision tree in the model makes a prediction, and the outcome that occurs most frequently will be the final prediction of the Random Forest model. This ensures that each prediction is more robust and that the model is controlling for overfitting, as individual decision trees tend to overfit (*1.11. Ensemble Methods*, n.d.).

XGBoost classifier

Another classifier model is the XGBoost classifier. XGB stands for extreme gradient boosting. Gradient boosting decision trees (GBDT) is a decision tree ensemble learning algorithm, much like Random Forest. The XGBoost classifier also uses a decision tree ensemble learning algorithm, but the decision trees are built and combined in a different way than Random Forest. “Gradient boosting” is the idea of boosting or improving a weak model by combining it with other weak models, to create a strong model, by using the gradient descent method (*What Is XGBoost?*, n.d.). Gradient descent is a mathematical method for finding the maximum or minimum of a differentiable function, by taking steps toward or away from the gradient of the function at the given point(‘Gradient Descent’, 2023). A GBDT model trains an ensemble of shallow decision trees, and with each iteration the error residuals of the previous model is used to fit the next model, thus, the gradient descent method is used to build and train the model to have minimal error. Furthermore, with the XGBoost model, the trees are built in parallel instead of sequentially. So while Random Forest minimizes variance and overfitting, XGBoost minimizes bias and underfitting (*What Is XGBoost?*, n.d.).

In this paper I will attempt to build, train and test both Random Forest and XGBoost classifier models, and assess model performance, in order to create the best performing classifier which can predict the heart attack risk level of a patient, based on medical data.

Methods

All computations for this paper were performed using Python 3 (Rossum & Drake, 2010), and can be found in GitHub repository: https://github.com/GacildaAnne/DatascienceExam

Data

The dataset used to train and test the classifier models in this paper are from the *Heart Attack Analysis & Prediction Dataset* (found on Kaggle: *Heart Attack Analysis & Prediction Dataset*, n.d.), which is a subset of the original dataset available on The University of California Irvine Machine Learning Repository (*UCI Machine Learning Repository: Heart Disease Data Set*, n.d.).

The dataset consists of medical data from 303 patients, 207 males and 96 females, ranging from age 29 to 77 (M = 54.36, SD = 9.08). See Appendix A for graph of age distribution across genders. The original dataset contains 76 variables, but all published experiments focus on only 14 variables, which are also the same 14 variables in the Heart Attack Analysis & Prediction Dataset available on Kaggle, see Appendix B. Some of the variable names are explained in Table 1 below:

Table 1

Age	Age of patient.
Sex	Sex of patient. 0: Female 1: Male
Cp	Chest pain type.

	1: Typical angina 2: Atypical angina 3: Non-anginal pain 4: (0) Asymptomatic
Trtbps	Resting blood pressure (mmHg) - Systolic blood pressure.
Chol	Cholesterol (mg/dl)
Fbs	Fasting blood sugar. 0: <120 mg/dl 1: >120 mg/dl
Restecg	Resting ECG results. 0: Normal 1: having ST-T wave abnormality, T-wave inversions and/or ST elevation or depression of > 0.05 mV 2: Showing probable or definite left ventricular hypertrophy by Este's criteria.
Thalachh	Maximum heart rate achieved.
Exng	Exercised induced angina.
Oldpeak	ST depression induced exercise relative to rest. The ST-segment is a flat isoelectric segment of an ECG (electrocardiogram).
Slp	Slope of the peak exercise ST segment. 0: unsloping (vertical position) 1: Flat 2: Downsloping
Caa	Number of major vessels.
Thall	Thalassemia. Inherited blood disorder characterized by absence or decreased accumulation of one of the globin subunits (Shang & Xu, 2017).
Output	Diagnosis of heart disease. Angiographic heart disease status (narrowing of any of the major blood vessels (Balashankar et al., n.d.). 0 = < 50% diameter narrowing. Lower risk of heart disease. 1 = > 50% diameter narrowing. Higher risk of heart disease.

Table 1: Descriptions of variables in Heart Attack Analysis & Prediction Dataset.

Modelling

Seven models were built, trained, and tested. The Random Forest classifier was called using the *RandomForestClassifier*() function imported from the *sklearn.ensemble* module, and the XGBoost classifier was called using the *XGBClassifier()* function imported from the *xgboost* library.

The two models, RFmodel1 and XGBmodel1, were trained on all 13 features in the dataset to compare models with fewer features to these models which include many (all possible) features. The fitted attribute *.feature_importances_* were used on both models to find the features that were evaluated to be most important. Feature importance outputs were slightly different for the two classifier models (see figure 4).

RFmodel2 was trained on the top three features and the 5th feature of the feature importance output for the Random Forest model:

RFmodel2 : cp, oldpeak, caa, chol

The 4th feature in the feature importance output were to be included in the model, however, due to *thalachh* (fourth feature in .feature_importance_ output for RFmodel1) being negatively correlated with *oldpeak* and positively correlated with *cp* (see correlation matrix in Appendix C), it was excluded from the model. Similarly, the 6th feature, *exng*, is correlated with *cp*, and was thus also excluded from the model.

As can be seen on Figure 3, the top three features in the feature importance output for XGBmodel1 are *cp*, *exang*, and *caa*. However, *exng* and *cp* are highly correlated and so is *exng* and the 4th feature in the output, *oldpeak* (see correlation matrix in Appendix C). Therefore, two models are made, one including *exng* and one including *oldpeak*. The 5th and 6th features (*sex* and *thall*) do not have any such issues with correlation and are thus included in both:

XGBmodel2 : cp, exng, caa, sex, thall

XGBmodel3 : cp, oldpeak, caa, sex, thall

Finally, two models (RFmodel3 and XGBmodel4) were trained on features according to some of the risk factors found in medical literature. Among the aforementioned risk factors (see page 2), only some are available in the dataset. There is no data concerning family history of coronary heart disease, smoking habits, or obesity and physical activity levels, and one of the variables that are present in the dataset is not entirely applicable, namely, *fbs*. Patients with type II diabetes (diabetes mellitus) have higher risk of cardiovascular disease (The Emerging Risk Factors Collaboration, 2010), thus the *fbs* variable may seem like a relevant feature. The *fbs* (fasting blood sugar) variable is a binary variable, where the value is 1 if the fasting blood sugar level of the patient is above 120 mg/dl and 0 if the value is below 120 mg/dl. According to Tominaga (1999) the diagnostic criteria of type II diabetes was fbs > 126 mg/dl, however, the diagnostic criteria has been updated to thresholds of FPG (fasting plasma glucose, also referred to as fasting blood sugar), 2-hPG (2-hour glucose concentration), and HbA1c (amount of glucose or blood sugar attached to hemoglobin) (Kumar et al., 2016). The fasting blood sugar level is therefore no longer enough to evaluate whether a patient is diabetic and the *fbs* variable is thus excluded from RFmodel3 and XGBmodel 4.

The features included in the models are the following:

RFmodel3 : age, sex, cp, trtbps, chol

XGBmodel4 : age, sex, cp, trtbps, chol

Choice of features in RFmodel3 and XGBmodel4

Age & sex : According to Mehta et al. (2016), both mortality rate, risk factors, and after-effects of heart attacks are different depending on both the sex and age of a person (Mehta et al., 2016). Additionally, Roeters van Lennep et al. (2002) found that cardiovascular risk factors differ for men and women (Roeters van Lennep et al., 2002), thus the risk of having a heart attack is influenced by the sex of the patient.

Chest pain type: Angina is a term referring to chest pain due to reduced blood flow to the heart muscle, thus the type of chest pain is relevant for the diagnosis of a heart attack.

Blood pressure: Blood pressure is measured in two numbers: systolic blood pressure and diastolic blood pressure. Systolic blood pressure measures the pressure in the arteries when

the heart beats, while the diastolic blood pressure measures the pressure in the arteries when the heart is resting between heartbeats (*High Blood Pressure Symptoms, Causes, and Problems | Cdc.Gov*, 2021). Only the systolic blood pressure is included in the dataset, however, a study by Bundy et al. (2017) found that reducing systolic blood pressure may significantly reduce the risk of cardiovascular disease (Bundy et al., 2017), therefore *trtbps* (systolic blood pressure) is included as a feature in the model.

Cholesterol: The total cholesterol level should be less than 200 mg/dl for a person to be healthy, as a person with higher total cholesterol levels has a higher risk of heart disease. However, it should be noted that there are two types of cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL). LDL is the bad cholesterol with little protein, and is the main source artery clogging plaque, while HDL is the good cholesterol that works to clear cholesterol from the blood, and has a lot of protein (Ma, 2004). Thus it may be more efficient to include the separate measures of HDL and LDL cholesterol.

Model performance was assessed and compared using the *classification_report()* function from the *sklearn.metrics* module.

Results

Figure 2

Model name	Model performance output
RFmodel1	precision recall f1-score support Low risk 0.786 0.733 0.759 30 High risk 0.758 0.806 0.781 31 accuracy 0.770 61 macro avg 0.772 0.770 0.770 61 weighted avg 0.771 0.770 0.770 61
RFmodel2	precision recall f1-score support Low risk 0.656 0.700 0.677 30 High risk 0.690 0.645 0.667 31 accuracy 0.672 61 macro avg 0.673 0.673 0.672 61 weighted avg 0.673 0.672 0.672 61
RFmodel3	precision recall f1-score support Low risk 0.710 0.733 0.721 30 High risk 0.733 0.710 0.721 31 accuracy 0.721 61 macro avg 0.722 0.722 0.721 61 weighted avg 0.722 0.721 0.721 61

Figure 2: Model performance outputs for each Random Forest classifier model.

Figure 3

Model name	Model performance output
XGBmodel1	precision recall f1-score support Low risk 0.840 0.700 0.764 30 High risk 0.750 0.871 0.806 31 accuracy 0.787 61 macro avg 0.795 0.785 0.785 61 weighted avg 0.794 0.787 0.785 61
XGBmodel2	precision recall f1-score support Low risk 0.778 0.700 0.737 30 High risk 0.735 0.806 0.769 31 accuracy 0.754 61 macro avg 0.757 0.753 0.753 61 weighted avg 0.756 0.754 0.753 61
XGBmodel3	precision recall f1-score support Low risk 0.850 0.567 0.680 30 High risk 0.683 0.903 0.778 31 accuracy 0.738 61 macro avg 0.766 0.735 0.729 61 weighted avg 0.765 0.738 0.730 61
XGBmodel4	precision recall f1-score support Low risk 0.700 0.700 0.700 30 High risk 0.710 0.710 0.710 31 accuracy 0.705 61 macro avg 0.705 0.705 0.705 61 weighted avg 0.705 0.705 0.705 61

Figure 3: Model performance outputs for each XGBoost classifier model.

Figure 4

Feature importance output RFmodel1		Feature importance output XGBmodel1	
cp	0.280941	cp	0.279542
oldpeak	0.121452	exng	0.166955
caa	0.105862	caa	0.119970
thalachh	0.090848	oldpeak	0.074536
exng	0.084189	sex	0.070020
chol	0.077227	thall	0.061043
age	0.071903	chol	0.037126
trtbps	0.060302	restecg	0.036797
thall	0.041626	thalachh	0.034991
sex	0.030960	fbs	0.034373
slp	0.016503	age	0.032912
restecg	0.011975	trtbps	0.027728
fbs	0.006213	slp	0.024007

Figure 4: Feature importance output for RFmodel1 (Random Forest classifier) and XGBmodel1 (XGBoost classifier) which are both trained on all features in dataset.

Discussion

The performance of a model is measured in accuracy, precision, recall, and F1-scores. These measures are calculated based on the number of true/false positives and true/false negatives a model makes. The classification_report() function calculates performance measures for each class. For instance, to calculate performance measures for the high risk class, a patient being high risk is 'positive' and not being high risk is 'negative'. A true positive would be a high risk patient being classified as high risk, where a false positive would be a low risk person being classified as high risk. Similarly, a true negative would be a low-risk person being classified as low risk, and a false negative would be a high-risk patient being classified as low risk. On the contrary, when the performance measures are being calculated for the low risk class, a person being low risk would be 'positive' and a person being high risk would be 'negative'. Thus, a true positive would be a low-risk person classified as low risk, a false positive would be a high-risk patient being classified as low risk, a true negative would be a high-risk patient being classified as high risk, and a false negative would be a low-risk person being classified as high risk.

The accuracy of a model is defined as the percentage of correctly predicted instances, out of all predictions. Thus accuracy is calculated by taking the sum of true positives and true negatives and dividing by the number of samples. The precision of a model is defined as the model's ability to avoid false positives, that is, i.e. to avoid classifying low-risk patients as high risk. In other words, precision is the ratio of true positives, to the sum of true and false positives. Recall is the model's ability to detect all positive samples and is thus the ratio of true positives to the sum of true positives and false negatives.
It is particularly important that the models have high recall measures in this situation, as the consequence of misclassifying a high-risk patient could be death or serious heart damage. Finally, the F1-score is a weighted 'harmonic' mean of precision and recall. The F1-score can, like the other performance measures, be a value between 0.00 and 1.00, where 0 is the worst and 1 is the best (*Sklearn.Metrics.Precision_recall_fscore_support*, n.d.).

RFmodel1 was trained on all features in the dataset. In RFmodel1 model performance, precision is higher for low risk (precision = 0.786) than for high risk (precision = 0.758), which means the model is slightly better at avoiding false positives in the low risk class.

Recall is higher for the high-risk class (recall = 0.806) than the low risk class (recall = 0.733), which is good, as the model needs to be better at picking up all high risk patients as mentioned earlier. The F1-score is also higher for the high-risk class, and the accuracy of the model is 0.770, meaning 77% of the predictions made by RFmodel1 were correct.

In comparison, XGBmodel1 which was also trained on all features in the dataset performed remarkably, better than RFmodel1. Even though the precision score is higher for the low-risk class (precision = 0.840) than the high-risk class (precision = 0.750), the model is showing good recall measures (recall high risk = 0.871, low risk = 0.700) and good F1-scores (high risk = 0.806, low risk = 0.764) as well as accuracy (accuracy = 0.787). Overall a quite impressive performance, compared to the Random Forest model trained on the same features.

RFmodel2 is the worst performing model. All performance measures are the absolute lowest for this model, compared to all the other model performances. Interestingly, this model is performing the opposite of the trend that all the models trained on all data, and on subsets according to feature importance do. Here, the high-risk class is having higher precision and lower recall and F1-score (precision = 0.690, recall = 0.645, F1-score = 0.667) and the low-risk class has lower precision but higher recall and F1-score (precision = 0.656, recall = 0.700, F1-score = 0.677) which is the opposite trend of what is seen in the other models. This model is doing a bad job at picking up the high risk patients, and with an accuracy of 0.672 it is generally performing worse than the RFmodel1, suggesting that more features may help the model to perform better predictions.

XGBmodel2 (accuracy = 0.754) performs a little worse than the XGBmodel1, supporting the implication that fewer features (and focusing only on the most important features) do not help the model performance. Overall performance measures are somewhat fine, however not the best. It seems that strong correlation between *exang* and *cp* are not too problematic for model predictions.

XGBmodel3 is showing rather high performance precision and recall, however, they are quite skewed/unbalanced (low risk precision = 0.850, high risk precision = 0.683 ; low risk recall = 0.567, high risk recall = 0.903). This exposes an underlying problem; the model is very good at detecting positive samples when the class is high risk, however, the ratio of true positives to the sum of true and false positives (precision) are lower, indicating that the model is more inclined to classify a sample as high risk (positive). In other words, the model makes more ‘high risk’ predictions, thus capturing more positives (high recall value) but consequently

also misclassifying more low risk samples, which explains the lower recall value and higher precision value for the low risk class; as a consequence of the model making fewer 'low risk' predictions, a larger ratio of them are correct, causing a high precision value. As a consequence for this unbalance, the F1-scores and accuracy measure do not best XGBmodel1.

XGBmodel4 was trained on features according to risk factors in medical literature. XGBmodel4 is not performing as good as the other XGBoost models (accuracy = 0.705). However, the performance measures are more balanced (high risk precision = 0.710, low risk precision = 0.700 ; high risk recall = 0.710, low risk recall = 0.700 ; high risk) than XGBmodel3, likely indicating a more equal distribution of predictions or weaker bias in the model.

RFmodel3 is the final Random Forest model trained on the same features as XGBmodel4. RFmodel3 performs slightly better than XGBmodel4 (accuracy = 0.721). The difference in performance is very small and occur as a small difference in precision and recall (high risk precision = 0.733, low risk precision = 0.710 ; high risk recall = 0.710, low risk recall = 0.733).

Conclusion

Of the seven classifier models created in this project, the best performing model is XGBmodel1, an XGBoost classifier which was trained on all 13 features in the dataset. RFmodel1, a Random Forest classifier model, was trained on the same features, and performed a little less good. Models trained on only the most important features were not better than models with all features. Especially with Random Forest models, the most important features only made the worst model, but made a somewhat decent model when features were chosen by risk factors in medical literature. Thus, it cannot be concluded that the crucial factor for a good classifier model may necessarily be the number of features a model was trained on. Additionally, it may be beneficial to be aware that the feature importance attribute may not be as accurate as expected. Finally, I have found that the XGBoost classifier models generally perform better than the Random Forest models.

Bibliography

1.11. Ensemble methods. (n.d.). Scikit-Learn. Retrieved 25 May 2023, from https://scikit-learn/stable/modules/ensemble.html

Balashankar, A., Lees, A., & Jagabathula, S. (n.d.). *Predicting Angiographic Disease Status: Drawing the line between demographically decoupled and jointly trained models*.

Bundy, J. D., Li, C., Stuchlik, P., Bu, X., Kelly, T. N., Mills, K. T., He, H., Chen, J., Whelton, P. K., & He, J. (2017). Systolic Blood Pressure Reduction and Risk of Cardiovascular Disease and Mortality: A Systematic Review and Network Meta-analysis. *JAMA Cardiology*, *2*(7), 775–781. https://doi.org/10.1001/jamacardio.2017.1421

Gradient descent. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Gradient_descent&oldid=1156803153

Heart Attack Analysis & Prediction Dataset. (n.d.). Retrieved 8 May 2023, from https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

High Blood Pressure Symptoms, Causes, and Problems | cdc.gov. (2021, May 18). Centers for Disease Control and Prevention. https://www.cdc.gov/bloodpressure/about.htm

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, *26*(9), Article 9. https://doi.org/10.1038/nbt0908-1011

Krumholz, H. M., Normand, S.-L. T., & Wang, Y. (2019). Twenty-Year Trends in Outcomes for Older Adults With Acute Myocardial Infarction in the United States. *JAMA Network Open*, *2*(3), e191938. https://doi.org/10.1001/jamanetworkopen.2019.1938

Kumar, R., Nandhini, L. P., Kamalanathan, S., Sahoo, J., & Vivekanadan, M. (2016). Evidence for current diagnostic criteria of diabetes mellitus. *World Journal of Diabetes*, *7*(17), 396–405. https://doi.org/10.4239/wjd.v7.i17.396

Ma, H. (2004). *Cholesterol and Human Health*.

Mehta, L. S., Beckie, T. M., DeVon, H. A., Grines, C. L., Krumholz, H. M., Johnson, M. N., Lindley, K. J., Vaccarino, V., Wang, T. Y., Watson, K. E., & Wenger, N. K. (2016). Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association. *Circulation*, *133*(9), 916–947. https://doi.org/10.1161/CIR.0000000000000351

Roeters van Lennep, J. E., Westerveld, H. T., Erkelens, D. W., & van der Wall, E. E. (2002). Risk factors for coronary heart disease: Implications of gender. *Cardiovascular Research*, *53*(3), 538–549. https://doi.org/10.1016/S0008-6363(01)00388-1

Rossum, G. van, & Drake, F. L. (2010). *The Python language reference* (Release 3.0.1 [Repr.]). Python Software Foundation.

Shang, X., & Xu, X. (2017). Update in the genetics of thalassemia: What clinicians need to know. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *39*, 3–15. https://doi.org/10.1016/j.bpobgyn.2016.10.012

Sklearn.ensemble.RandomForestClassifier. (n.d.). Scikit-Learn. Retrieved 24 May 2023, from https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Sklearn.metrics.precision_recall_fscore_support. (n.d.). Scikit-Learn. Retrieved 3 January 2023, from https://scikit-learn/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

The Emerging Risk Factors Collaboration. (2010). Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102

prospective studies. *The Lancet*, *375*(9733), 2215–2222. https://doi.org/10.1016/S0140-6736(10)60484-9

Torpy, J. M., Burke, A. E., & Glass, R. M. (2009). Coronary Heart Disease Risk Factors. *JAMA*, *302*(21), 2388. https://doi.org/10.1001/jama.302.21.2388

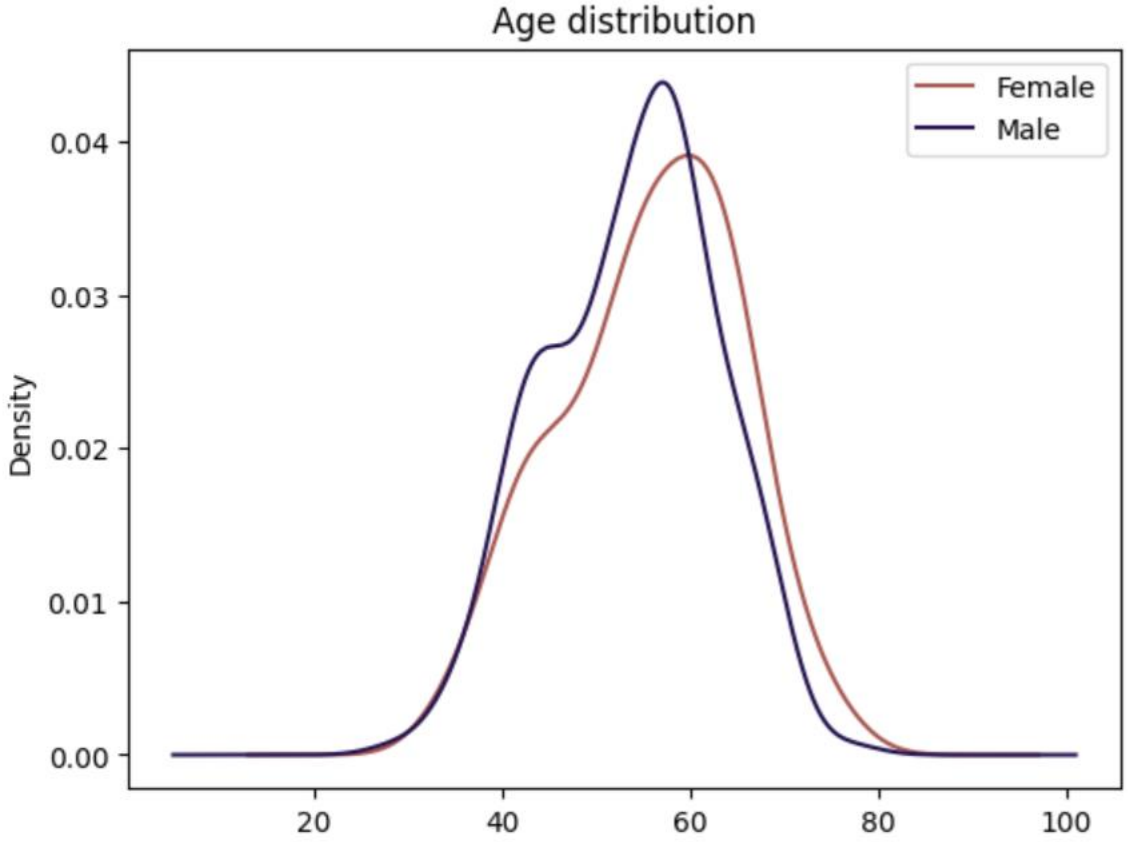
Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Anderson, C. A. M., Arora, P., Avery, C. L., Baker-Smith, C. M., Beaton, A. Z., Boehme, A. K., Buxton, A. E., Commodore-Mensah, Y., Elkind, M. S. V., Evenson, K. R., Eze-Nliam, C., Fugar, S., Generoso, G., Heard, D. G., Hiremath, S., Ho, J. E., … null, null. (2023). Heart Disease and Stroke Statistics—2023 Update: A Report From the American Heart Association. *Circulation*, *147*(8), e93–e621. https://doi.org/10.1161/CIR.0000000000001123

UCI Machine Learning Repository: Heart Disease Data Set. (n.d.). Retrieved 19 May 2023, from https://archive.ics.uci.edu/ml/datasets/Heart+Disease

What is XGBoost? (n.d.). NVIDIA Data Science Glossary. Retrieved 25 May 2023, from https://www.nvidia.com/en-us/glossary/data-science/xgboost/

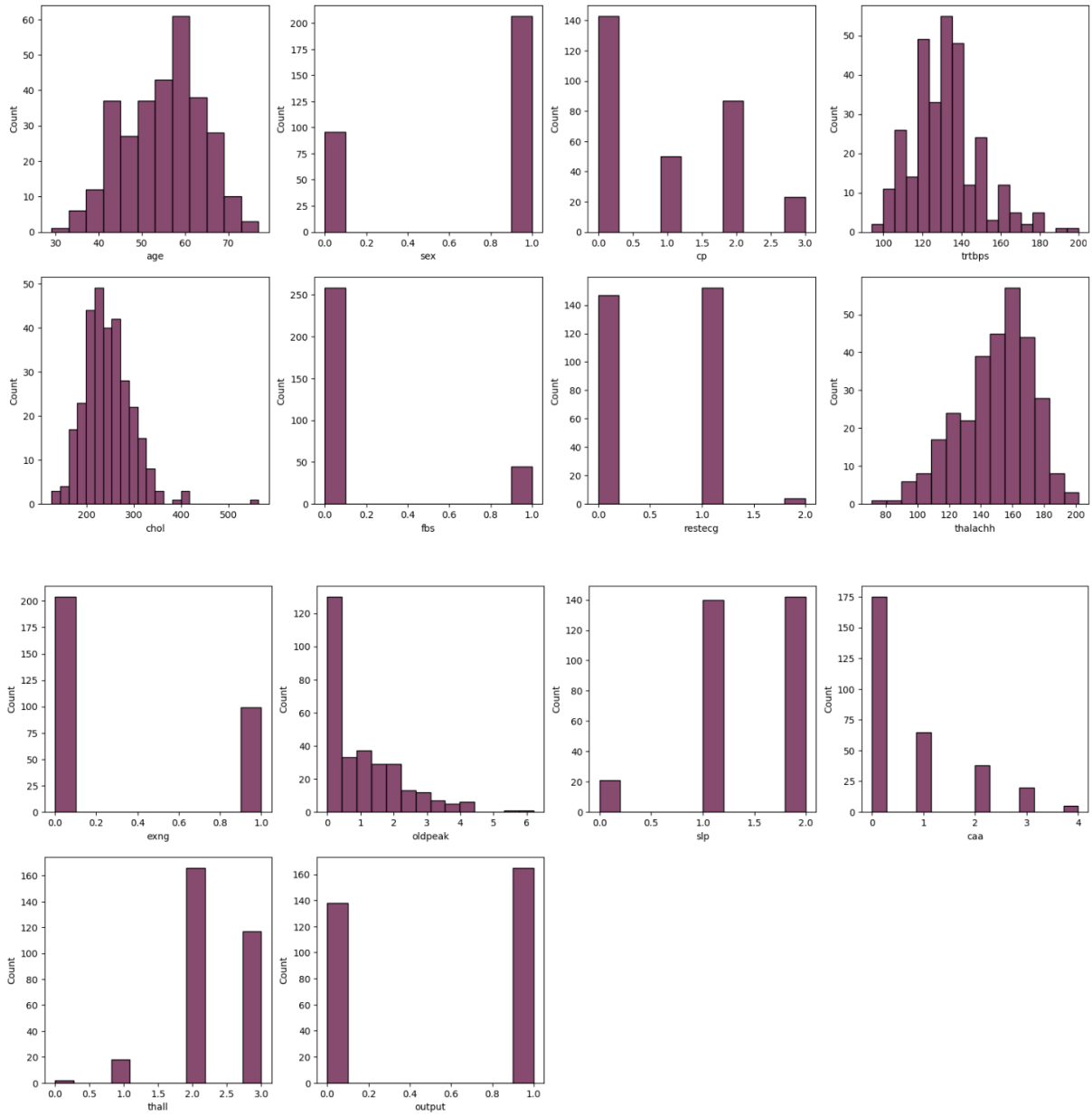
Appendix A

Age distribution across genders.



Appendix B

Bar graphs of each variable in the dataset.



Appendix C

Correlation matrix