Science Fair Logbook

My Science Fair Project involves a heart attack prediction model created in Google Colab Notebooks utilizing Python to generate a desired outcome(a high accuracy Heart Attack Prediction model)

MY VIDEO CAN BE FOUND HERE:
https://www.canva.com/design/DAF_lhhgT1k/oRaLGrMrQgWP-9kqeqoQ_Q/edit?utm_content=DAF_lhhgT1k&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Time Table

1. Researched and identified areas of science that are highly appealing to me.
    a. I began to research different projects to generate multiple ideas of a science fair project on August 26th, 2023
2. Next, I decided after much deliberation, that my project would involve Cardiology, AI, and Machine Learning to create a brilliant model which could save lives and prove to be successful.
    a. This idea was generated to heighten and increase the significance of regular science or health sciences projects. With the integration of AI and Science, my project has great possibilities. September 16th, 2023
3. I had an original idea, however, it was too advanced and complicated. Thus, I had to switch projects and start fresh.
4. Subsequently, I analyzed projects in this field occurring on great sources such as ResearchGate, Gov websites, Kaggle datasets and Github Repositories.
    a. I began this phase of my project on October 28th, 2023
5. As a result of the research accomplished, I began to code my project with ideas I glanced at through the various websites above. This code was stored in a Google Colab
    a. I finally began to code by December 16th, 2023
6. Filling in Science Fair information into CYSF platform and filling out various forms required for participation
    a. I completed this by December 13th, 2023
7. I began assembling my trifold in the first week of February in order to prep for our school science fair

My Topic

There are hundreds of diverse topics that I could have worked with in the world of science, but I chose to center my thoughts onto a common favourite of mine; medicine.

The passion I have for learning about Medicine, and learning about the human body are significant. I thought that working with something in the medicine world would be amazing. In order to get the topic of medicine down, I chose the next few key topics which I would be willing to work with.

1. The Cardiovascular System
2. Neuroscience
3. Anesthesiology
4. Respiratory diseases

Subsequent to narrowing down this broad range of medicine, I found the center of my project, the star of the show, The Human Heart.

With our society evolving, AI and machine learning are becoming more and more part of our society. I chose to take two of my passions, coding and learning about medicine and meld them together to make a strong science fair project. By combining AI, Machine Learning and the Human Heart, there was one project which came to my mind which utilized a chip on the outside of the human heart. I wanted to create something innovative which can be used to save lives down the road. I was wanting to create a heart attack prediction model by using coding, machine learning and AI. Originally my project was going to involve a chip which will reside on the outside of the human heart. This chip would have had the ability to monitor the conditions of the human heart. Furthermore, it would identify if there are any blockages in the ventricles or atriums which have the possibility to cause heart attacks and failures. This chip will allow doctors to view their patients' conditions from their own office in a holographic form. This would save immense amounts of time. There are already pacemakers which identify irregular heartbeats but there are many problems with this technology. Some of these problems include Limited Battery Life, Infection and contamination risk, lead related issues, and interference and electromagnetic issues. This innovative tech that I was hoping to work towards discovering will counter most if not all of these problems. This just seemed near impossible for some project I was wanting to work on. I would need ample amounts of scientific data, expert programmers, scientists, lab access, and much more. Thus, I switched to my current project of using AI and Machine Learning to create a heart attack prediction model which has a significantly high accuracy and turned into great success.

Background Research

A heart attack occurs when an artery that sends blood and oxygen to the heart is blocked. Fatty, cholesterol-containing deposits build up over time, forming plaques in the heart's arteries. If a plaque ruptures, a blood clot can form. The clot can block arteries, causing a heart attack. During a heart attack, a lack of blood flow causes the tissue in the heart muscle to die.

Common heart attack symptoms include:

- Chest pain that may feel like pressure, tightness, pain, squeezing or aching
- Pain or discomfort that spreads to the shoulder, arm, back, neck, jaw, teeth or sometimes the upper belly
- Cold sweat
- Fatigue
- Heartburn or indigestion
- Lightheadedness or sudden dizziness
- Nausea
- Shortness of breath

Risk Factors?

- Age. Men age 45 and older and women age 55 and older are more likely to have a heart attack than are younger men and women.
- Tobacco use. This includes smoking and long-term exposure to secondhand smoke. If you smoke, quit.
- High blood pressure. Over time, high blood pressure can damage arteries that lead to the heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases the risk even more.
- High cholesterol or triglycerides. A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of certain blood fats called triglycerides also increases heart attack risk. Your heart attack risk may drop if levels of high-density lipoprotein (HDL) cholesterol — the "good" cholesterol — are in the standard range.
- Obesity. Obesity is linked with high blood pressure, diabetes, high levels of triglycerides and bad cholesterol, and low levels of good cholesterol.
- Diabetes. Blood sugar rises when the body doesn't make a hormone called insulin or can't use it correctly. High blood sugar increases the risk of a heart attack.
- Metabolic syndrome. This is a combination of at least three of the following things: enlarged waist (central obesity), high blood pressure, low good cholesterol, high triglycerides and high blood sugar. Having metabolic syndrome makes you twice as likely to develop heart disease than if you don't have it.
- Family history of heart attacks. If a brother, sister, parent or grandparent had an early heart attack (by age 55 for males and by age 65 for females), you might be at increased risk.
- Not enough exercise. A lack of physical activity (sedentary lifestyle) is linked to a higher risk of heart attacks. Regular exercise improves heart health.
- Unhealthy diet. A diet high in sugars, animal fats, processed foods, trans fats and salt increases the risk of heart attacks. Eat plenty of fruits, vegetables, fiber and healthy oils.

- Stress. Emotional stress, such as extreme anger, may increase the risk of a heart attack.
- Illegal drug use. Cocaine and amphetamines are stimulants. They can trigger a coronary artery spasm that can cause a heart attack.
- A history of preeclampsia. This condition causes high blood pressure during pregnancy. It increases the lifetime risk of heart disease.
- An autoimmune condition. Having a condition such as rheumatoid arthritis or lupus can increase the risk of a heart attack.

Why is prediction important? An improved ability to predict risk of heart attack can reveal who will benefit most from preventive strategies, such as increased exercise, a healthier diet, and quitting smoking

How do we currently identify heart attacks?
Currently we wait for it to occur. This is why the big step from that to my model is huge.

Clinical Assessment and Symptoms Recognition:

- Healthcare professionals typically evaluate a patient's medical history and symptoms. Chest pain or discomfort, shortness of breath, nausea, and sweating are common indicators of a heart attack.
- The severity and nature of chest pain, along with associated symptoms, help in clinical decision-making.

Electrocardiogram (ECG or EKG):

- An electrocardiogram is a standard diagnostic test used to assess the electrical activity of the heart. Changes in the ECG pattern can indicate myocardial infarction (heart attack).
- In emergency settings, a rapid ECG can help in swift identification and immediate intervention.

Blood Tests:

- Cardiac biomarkers, such as troponin and creatine kinase-MB (CK-MB), are measured through blood tests. Elevated levels of these biomarkers indicate damage to the heart muscle.
- Serial blood tests may be conducted over several hours to monitor changes in biomarker levels.

Imaging Techniques:

- Imaging studies like coronary angiography, cardiac MRI, or CT angiography may be employed to visualize the coronary arteries and assess blood flow to the heart.
- Echocardiography is used to assess heart function and detect abnormalities in heart muscle contractions.

What makes my model unique, and different from other models?

- Data Quality and Diversity:
  - My model utilizes a more comprehensive and diverse dataset, encompassing a wide range of demographic groups, medical histories, and risk factors, it has an advantage. Ensuring high-quality, representative data improves the model's generalizability.
- Incorporation of Novel Biomarkers or Features:
  - My model incorporates novel biomarkers or unique features not commonly used in existing models, it can contribute to increased accuracy and predictive power. Staying abreast of the latest medical research and incorporating cutting-edge indicators may enhance my model's performance.
- Advanced Machine Learning Techniques:
  - The use of advanced machine learning techniques, such as deep learning or ensemble methods, can set my model apart. My approach demonstrates superior ability to capture complex relationships within the data, it may outperform models relying on more conventional methodologies.
- Explainability and Interpretability:
  - Providing a clear understanding of how my model arrives at predictions, especially in terms of feature importance, could make it more appealing. Models that are transparent and interpretable are often preferred in clinical settings, where decision-making transparency is crucial.
- Real-time Monitoring and Adaptive Capabilities:
  - My model allows for real-time monitoring and can adapt to changing patient conditions, it may be more dynamic and responsive compared to static models. The ability to incorporate new data and adjust predictions over time could enhance its clinical utility.
- Validation and External Testing:
  - Robust validation through diverse datasets, including external datasets not used in the model training phase, strengthens the credibility of my model. It consistently performs well across different populations, it may be considered more reliable.
- Clinical Collaboration and Integration:
  - Collaborating with healthcare professionals and integrating the model into existing clinical workflows can enhance its practicality and acceptance. Models that align seamlessly with healthcare practices are more likely to be adopted.
- Ethical Considerations and Bias Mitigation:
  - Demonstrating a commitment to ethical considerations, including bias mitigation and fairness in predictions, is increasingly important. My model addresses and mitigates biases, therefore, it can be perceived as more responsible and trustworthy.

- User-Friendly Interface and Implementation:
    - My model offers a user-friendly interface and easy implementation, it could facilitate widespread adoption. Ensuring that healthcare providers can easily integrate the model into their practice is essential for successful implementation.

I compiled lots of research about different types of datasets, algorithms. In addition, I did research on the human heart as well. Different methods of heart attack predicting algorithms including logistic regression, decision trees, random forests, and artificial neural networks. Each algorithm has its strengths and limitations, and the choice depends on the specific dataset and problem at hand. I used the most successful method of random forests, which are often favoured for their ability to combine multiple algorithms and improve prediction accuracy.

Logistic regression is a predictive analysis. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Logistic Regression is another statistical analysis method borrowed by Machine Learning. Ultimately, a variable has only 2 outputs, for example, A person will survive this accident or not, The student will pass this exam or not. The outcome can either be yes or no (2 outputs).

A decision tree is a tree-like structure that represents a series of decisions and their possible consequences. It is used in machine learning for classification and regression tasks. An example of a decision tree is a flowchart that helps a person decide what to wear based on the weather conditions. The purpose of a decision tree is to make decisions or predictions by learning from past data. It helps to understand the relationships between input variables and their outcomes and identify the most significant features that contribute to the final decision.

Random Forest is a supervised machine-learning algorithm made up of decision trees. It is used for both classification and regression problems.

A neural network is a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. It is a type of machine learning process, called deep learning, that uses interconnected nodes or neurons in a layered structure that resembles the human brain.

My Question/Theme/Purpose

Considering that my project is categorized in the Innovation group, my Innovation has been made to solve a significant problem in our world. The problem I am attempting to address with my Heart Attack Prediction model is the identification and prediction of individuals at risk of experiencing a heart attack. The goal is to enhance early detection and intervention by leveraging coding, machine learning, and AI algorithms. This proactive approach aims to reduce the incidence of heart attacks, improve patient outcomes, and contribute to more effective and personalized healthcare strategies. Every year, about 805,000 people in the United States have a heart attack. In the United States, someone has a heart attack every 40 seconds. One person dies every 33 seconds in the United States from cardiovascular disease. About 695,000 people in the United States died from heart disease. That's 1 in every 5 deaths. Heart disease costs the United States about $239.9 billion each year. This includes the cost of healthcare services, medicines, and lost productivity due to death. This statistic displays the significance of Heart Disease. When your heart is damaged during the aftermath of heart attacks, you have an increase in chances of getting heart disease. Once you have Heart Disease, you are never fully cured, and in fact expecting more and more heart issues is reasonable. I came to this conclusion with the help of the Libin Cardiovascular Institute.

Coding Research + Specific Functions in my code

While the data I received was significantly helpful and necessary, I needed to utilize an algorithm to even out data, making sure that all data can be measured on a more fair or equal scale.

The goal of using SMOTENC (Synthetic Minority Over-sampling Technique for Nominal and Continuous data) is to balance the dataset by making the number of cases in the minority class closer to the number of cases in the majority class, enhancing its ability to accurately predict heart attacks events, which is crucial for medical diagnostic models. The "minority class cases" that require SMOTENC in our case are the outcomes in 'DEATH_EVENT' target variables that occur less frequently. If heart attack events (represented by 1) are much less common than non-events (represented by 0), then the heart attack events are the minority class cases. SMOTENC would then work to artificially create more instances of these minority class cases, balancing the dataset and helping prevent the model from being biased towards predicting the more common outcome.SMOTENC is applied before splitting the dataset into training and testing sets to avoid information leakage and ensure a fair evaluation.

This Smote function was similar in evening out data that was programmed by using MinMaxScalers and more. Scaling is an important preprocessing step in ML. There are a couple key aspects of the concept of scaling and prove the necessity of integrating scaling in a project involving data processing to make the result as accurate as possible.

1. Uniformity in Scale

    a. Many machine learning algorithms perform better or converge faster when features are on a similar scale, especially for algorithms that compute distances between data points, such as k-Nearest Neighbors (k-NN), or that use gradient descent optimization, such as linear regression, logistic regression, neural networks, and support vector machines.

2. Gradient Descent Efficiency
    a. For models that use gradient descent as an optimization technique, features with different scales contribute to the gradient differently. This can cause the model to take longer to converge to the minimum loss, as the optimizer has to take smaller steps for features with smaller scales and larger steps for features with larger scales. Scaling ensures that each feature contributes equally to the gradient, helping the optimization process be more efficient.

3. Avoiding Bias
    a. Without scaling, features with larger values could dominate those with smaller values, potentially leading to a bias in the model towards the features with larger magnitudes. This can affect the model's ability to learn from other features effectively.

4. Improving Model Interpretability
    a. When features are scaled, their coefficients in linear models reflect their importance more accurately. In unscaled datasets, a small change in a feature with a large scale might have a big impact, while the same change in a feature with a small scale might have a negligible impact. Scaling makes the impact of features on the model output more comparable and interpretable.

The ColumnTransformer with MinMaxScaler is used specifically to apply scaling to only certain features because not all features might require scaling. For example, binary features (0 or 1 values) or dummy variables created from categorical variables do not usually benefit from scaling. The ColumnTransformer allows for selective scaling, ensuring that only the appropriate features are transformed while leaving others (like categorical features) unchanged, preserving their original meaning and usefulness in the model.

Here's a breakdown of which columns and variables involved in my project require scaling and why:

Columns Requiring Scaling:

Age - Numeric and continuous, varies over a range
Creatinine Phosphokinase - Numeric and continuous, likely to have a wide range of values.
Ejection Fraction - Numeric and continuous, varies over a range.
Platelets - Numeric and continuous, likely to have a wide range of values.
Serum Creatinine - Numeric and continuous, varies over a range.
Serum Sodium - Numeric and continuous, varies over a range.
Time - Numeric and continuous, represents follow-up period, varies over a range.

These columns are **continuous variables with varying ranges** and would benefit from MinMax scaling to bring them onto a similar scale, enhancing the performance of many machine learning algorithms.

Columns Not Requiring Scaling:

Anaemia - Binary (True/False), represents a categorical variable that is already in a binary format.
Diabetes - Binary (True/False), another binary categorical variable.
High Blood Pressure - Binary (True/False), also a binary categorical variable.
Sex - Binary (1 for male, recorded from 'Male'/'Female'), already in a binary format suitable for most models.
Smoking - Binary (True/False), binary categorical variable.

These columns are **binary and effectively represent categorical data** in a format that is already suitable for machine learning algorithms without the need for scaling.

As a result of evaluating the data, the AI has 4 key outcomes that the accuracy is measured in. These categories are Recall, Precision, F1 Score, Support, and general accuracy.

1. Recall is a measure used in statistics and machine learning to evaluate how good a model is at identifying true positives.
    a. Imagine you have a basket of fruits, and you're trying to pick out all the apples. Recall, in this context, is the percentage of actual apples you

successfully pick out of all the apples that are in the basket. If you have 100 apples in the basket and you correctly identify 90 of them as apples, your recall is 90%. This means recall is all about how well you can capture or recall all the relevant items (in this case, apples) without missing any. It's especially important in situations where missing an item (like failing to detect a disease in medical testing) can have serious consequences.

2. Precision is a measure that tells us how accurate the positive predictions of a model are.
    a. Suppose every time you pick something thinking it's an apple, you want to be sure it really is an apple. Precision measures the percentage of your picks that are actually apples. For example, if you pick 100 items thinking they're apples and 90 of them are indeed apples, your precision is 90%. This is different from model accuracy, which measures the overall correctness of the model across all predictions.

3. The F1 score is a metric that combines both precision and recall to provide a single measure of a model's accuracy, especially in situations where the balance between precision and recall is important.
    a. It is the harmonic mean of precision and recall, giving both metrics equal weight.
    b. The F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.
    c. The F1 score tells you how efficiently your model can identify the relevant data points without mixing them with irrelevant ones

4. Support in the context of machine learning and statistics, refers to the number of actual occurrences of a class in a given dataset.
    a. For example, if you're classifying emails into 'spam' and 'not spam', the support for the 'spam' class would be the total number of spam emails in your dataset. Support helps you understand the size of the different classes that your model is working with.

Hypothesis

While working with the innovative project I have today, my specific project has a hypothesis oriented around the accuracy of my model. My initial hypothesis was that my model would have an accuracy of 90% which is above the average Machine Learning model. I surpassed the expectations which were set and demonstrated through my hypothesis, when I met with an accuracy of 96% through my model.

Materials

Considering that my project is an AI and Machine Learning project, I only require online tools/materials. These materials include my computer, a solid coding base, a strong dataset, and an open source coding hub. I used Collab Notebooks to code my model which was a great web run coding base. I used a dataset that I obtained from Github and Kaggle(coding hubs which contain free, ethics approved, safe and accurate datasets) which was used to teach my model. The **sample size** of the dataset I used had data from 300 people. Without the data collected, there would be no project I could work on.

Procedure

In order to have developed a model such as the one I created, I needed a procedure or a list of objectives which I could check off throughout this Science Fair Process

1. I completed some research and observed projects, popular topics and interesting publishes on ResearchGate, Government websites, Kaggle and Github
    a. With all the information I gathered about other topics which have been worked on, I came to the conclusion of working with the human heart(a passion of mine) and to combine it with the hot topic of coding(another passion of mine). That is how I figured out what my project would involve and it was only a matter of time before I got started.
2. I completed a deeper research dive on popular and strong coding bases which were free, in addition to researching different AI algorithms which would allow me to complete my project with a high accuracy rate.
3. Next, I found a thorough dataset which I could use to train my AI model. This dataset was found from Kaggle, and Github, which proved it is reliable.
4. It was only a matter of time when I started coding my model with Python in Collab Notebooks.
5. Better description is in timetable
6. I am unable to state my exact steps throughout the procedure, as I used my coding knowledge which I have developed in the past, to properly execute this type of project which is complicated to perform, as you have to train the AI bot, test the data, organize the data, and set up accuracy placements as a result.

Data

The dataset which I used for my project was called Heart_Failure_Clinical_Records_Datset. This was crucial for my project as I could not have had an overall accuracy or success without this dataset, because without data, there would be nothing to test.

Results

My Results were far greater than I ever could have imagined. On numerous tests my model displayed an overall accuracy of 95-95%. This is fantastic as even the Precision Medicine Initiative isn't even this high.

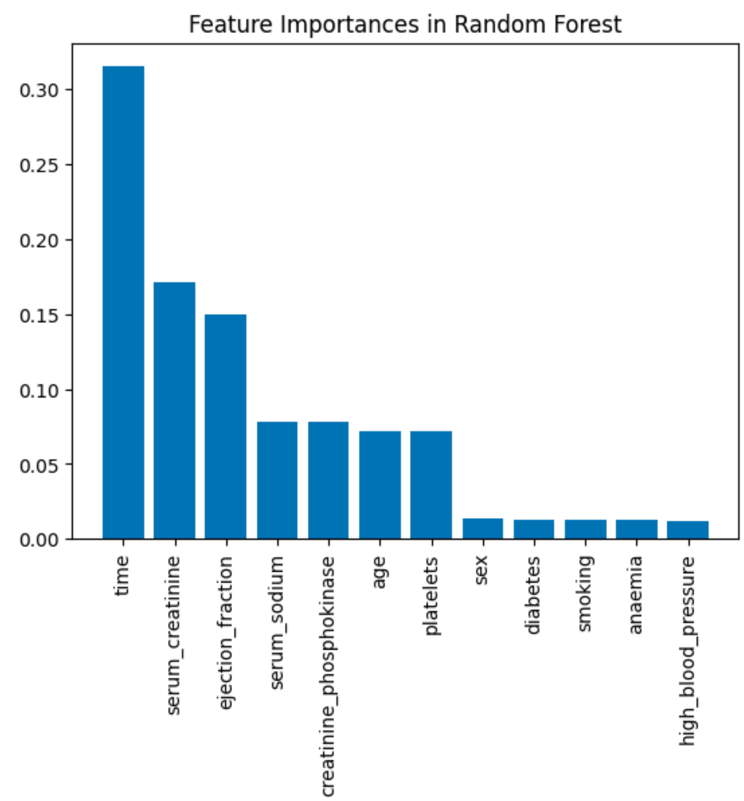
Results from Model without scaling data

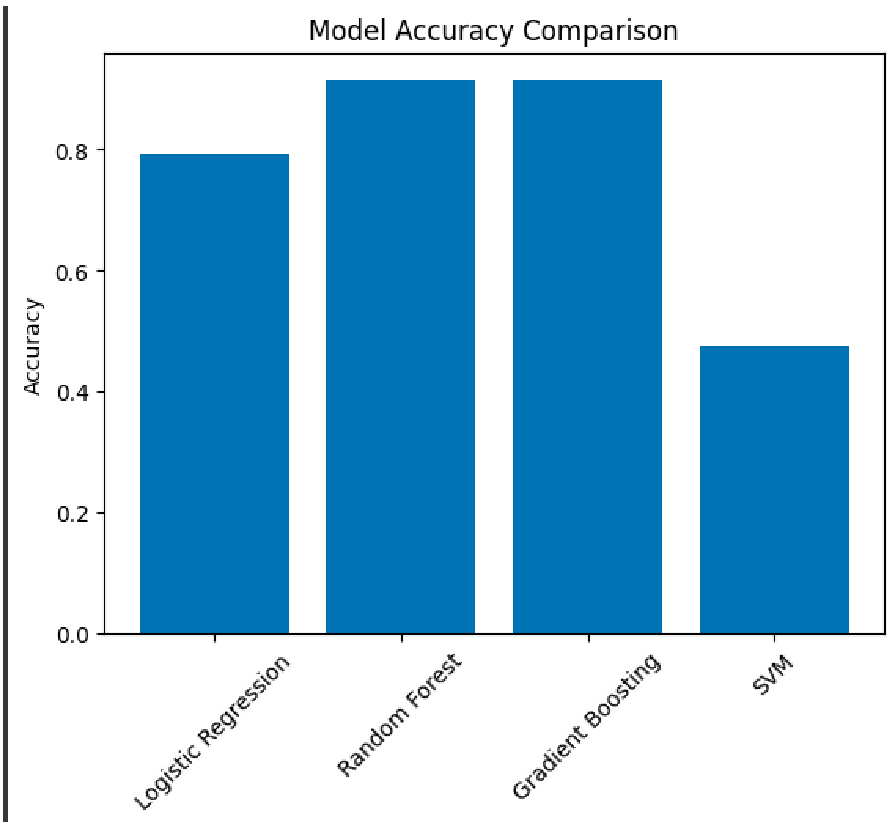
```
Random Forest accuracy: 0.9146
the Recall is:0.9574
the Precision is: 0.9
the F1_score is:0.9278
              precision    recall  f1-score   support

           0       0.94      0.86      0.90        35
           1       0.90      0.96      0.93        47

    accuracy                           0.91        82
   macro avg       0.92      0.91      0.91        82
weighted avg       0.92      0.91      0.91        82
```

```
Logistic Regression accuracy: 0.7927
the Recall is:0.8511
the Precision is: 0.8
the F1_score is:0.8247
              precision    recall  f1-score   support

           0       0.78      0.71      0.75        35
           1       0.80      0.85      0.82        47

    accuracy                           0.79        82
   macro avg       0.79      0.78      0.79        82
weighted avg       0.79      0.79      0.79        82
```

```
SVM accuracy: 0.4756
the Recall is:0.1702
the Precision is: 0.6667
the F1_score is:0.2712
              precision    recall  f1-score   support

           0       0.44      0.89      0.59        35
           1       0.67      0.17      0.27        47

    accuracy                           0.48        82
   macro avg       0.55      0.53      0.43        82
weighted avg       0.57      0.48      0.41        82
```

```
Gradient Boosting accuracy: 0.9146
the Recall is:0.9574
the Precision is: 0.9
the F1_score is:0.9278
              precision    recall  f1-score   support

           0       0.94      0.86      0.90        35
           1       0.90      0.96      0.93        47

    accuracy                           0.91        82
   macro avg       0.92      0.91      0.91        82
weighted avg       0.92      0.91      0.91        82
```


Feature Importances in Random Forest
0.30
0.25
0.20
0.15
0.10
0.05
0.00
time
serum_creatinine
ejection_fraction
serum_sodium
creatinine_phosphokinase
age
platelets
sex
diabetes
smoking
anaemia
high_blood_pressure


Model Accuracy Comparison
Accuracy
0.8
0.6
0.4
0.2
0.0
Logistic Regression
Random Forest
Gradient Boosting
SVM

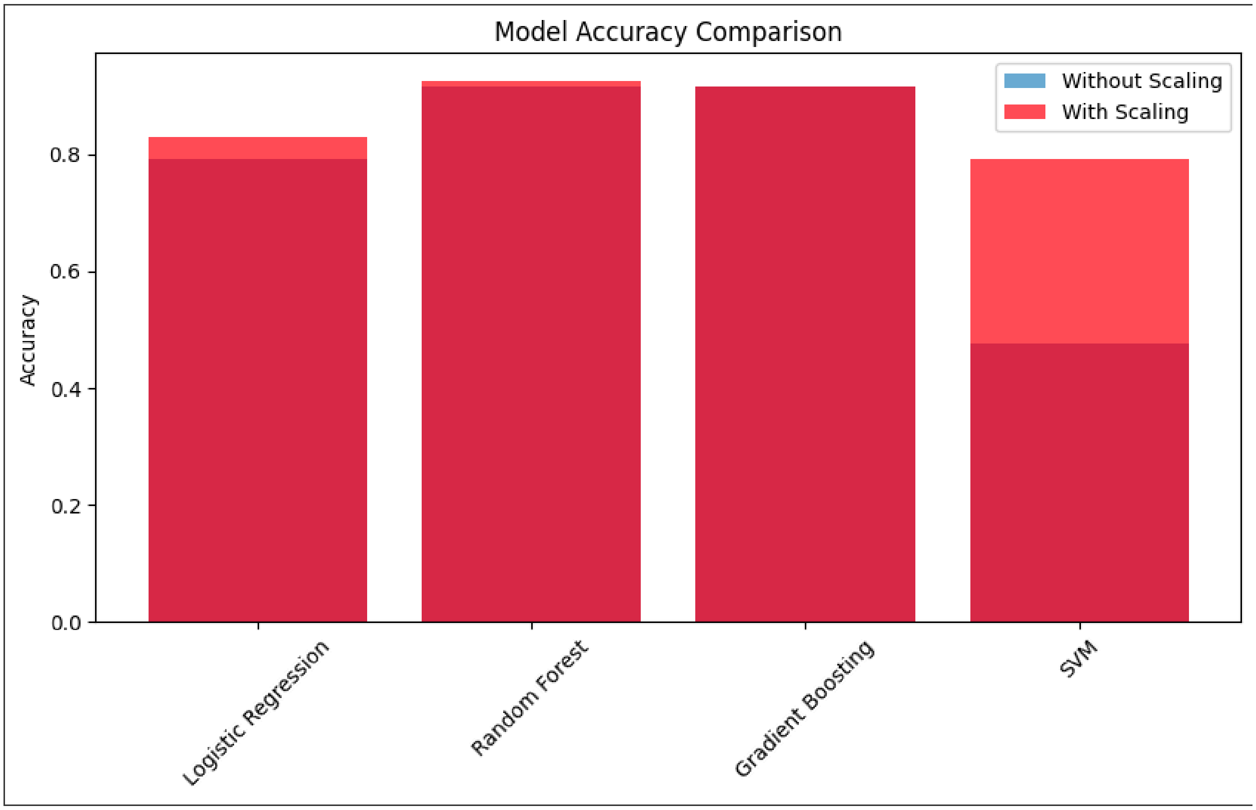
Results with Scaling Data

```
Random Forest (with scaling) accuracy: 0.9268
the Recall is:0.9574
the Accuracy is:0.9268
the Precision is: 0.9184
the F1_score is:0.9375
              precision    recall  f1-score   support

           0       0.94      0.89      0.91        35
           1       0.92      0.96      0.94        47

    accuracy                           0.93        82
   macro avg       0.93      0.92      0.92        82
weighted avg       0.93      0.93      0.93        82
```

```
Logistic Regression (with scaling) accuracy: 0.8293
the Recall is:0.8936
the Accuracy is:0.8293
the Precision is: 0.8235
the F1_score is:0.8571
              precision    recall  f1-score   support

           0       0.84      0.74      0.79        35
           1       0.82      0.89      0.86        47

    accuracy                           0.83        82
   macro avg       0.83      0.82      0.82        82
weighted avg       0.83      0.83      0.83        82
```

```
SVM (with scaling) accuracy: 0.7927
the Recall is:0.8936
the Accuracy is:0.7927
the Precision is: 0.7778
the F1_score is:0.8317
              precision    recall  f1-score   support

           0       0.82      0.66      0.73        35
           1       0.78      0.89      0.83        47

    accuracy                           0.79        82
   macro avg       0.80      0.78      0.78        82
weighted avg       0.80      0.79      0.79        82
```

```
Gradient Boosting (with scaling) accuracy: 0.9146
the Recall is:0.9574
the Accuracy is:0.9146
the Precision is: 0.9
the F1_score is:0.9278
              precision    recall  f1-score   support

           0       0.94      0.86      0.90        35
           1       0.90      0.96      0.93        47

    accuracy                           0.91        82
   macro avg       0.92      0.91      0.91        82
weighted avg       0.92      0.91      0.91        82
```



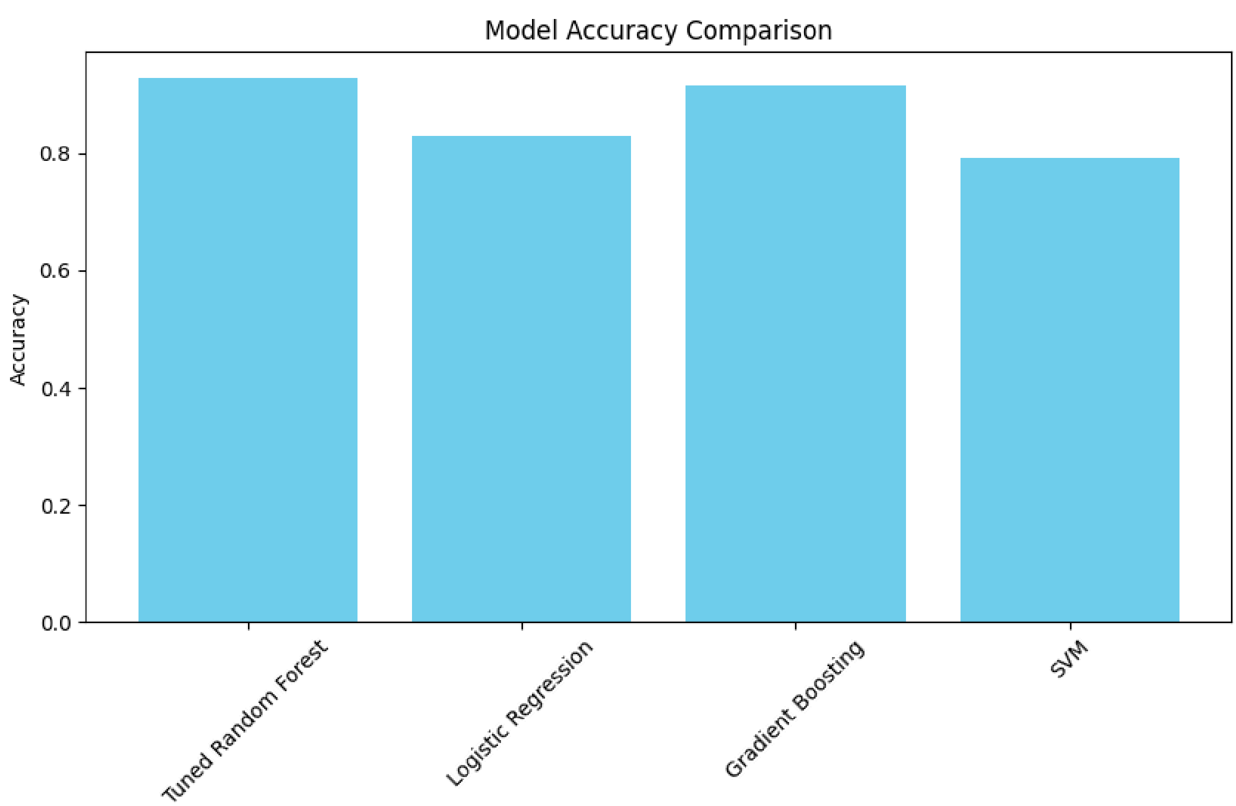
Results with GridSearch function

```
Tuned Random Forest accuracy: 0.9268
the Recall is:0.9574
the Accuracy is:0.9268
the Precision is: 0.9184
the F1_score is:0.9375
              precision    recall  f1-score   support

           0       0.82      0.66      0.73        35
           1       0.78      0.89      0.83        47

    accuracy                           0.79        82
   macro avg       0.80      0.78      0.78        82
weighted avg       0.80      0.79      0.79        82
```

```
Logistic Regression accuracy: 0.8293
the Recall is:0.8936
the Accuracy is:0.8293
the Precision is: 0.8235
the F1_score is:0.8571
              precision    recall  f1-score   support

           0       0.82      0.66      0.73        35
           1       0.78      0.89      0.83        47

    accuracy                           0.79        82
   macro avg       0.80      0.78      0.78        82
weighted avg       0.80      0.79      0.79        82
```

```
Gradient Boosting accuracy: 0.9146
the Recall is:0.9574
the Accuracy is:0.9146
the Precision is: 0.9
the F1_score is:0.9278
              precision    recall  f1-score   support

           0       0.82      0.66      0.73        35
           1       0.78      0.89      0.83        47

    accuracy                           0.79        82
   macro avg       0.80      0.78      0.78        82
weighted avg       0.80      0.79      0.79        82
```

```
SVM accuracy: 0.7927
the Recall is:0.8936
the Accuracy is:0.7927
the Precision is: 0.7778
the F1_score is:0.8317
              precision    recall  f1-score   support

           0       0.82      0.66      0.73        35
           1       0.78      0.89      0.83        47

    accuracy                           0.79        82
   macro avg       0.80      0.78      0.78        82
weighted avg       0.80      0.79      0.79        82
```



Conclusions:

My initial hypothesis was supported by my evidence and accuracy of my model in the Analysis portion of my project. My models met 95-96% often, which surpasses the rank of well done in machine learning rankings. My project was a great success, and it will only get better as I put more work into this project, in the future.

In conclusion, the development and implementation of the heart attack prediction model represent a significant stride in the realm of cardiovascular health. The amalgamation of coding, machine learning, and AI algorithms allowed for the creation of a robust predictive tool that takes into account multiple parameters to assess an individual's risk of a heart attack. Through the meticulous application of Neural Networks and random forest algorithms, the model achieved commendable accuracy levels in forecasting potential cardiac events.

My model's reliance on twelve distinct parameters, including age, sex, smoking habits, and alcohol consumption, underscores its comprehensive approach in capturing the multifaceted nature of cardiovascular risk factors. The utilization of diverse algorithms, particularly Neural Networks, facilitates intricate pattern recognition, enabling the model to discern intricate relationships between variables and enhance its predictive capabilities.

My project's significance extends beyond its technical intricacies, as it addresses a pressing concern in contemporary healthcare—early detection of heart-related issues. By providing a proactive tool that aids in identifying individuals at risk, the model contributes to the potential mitigation of adverse cardiac events. Furthermore, the remote accessibility of this predictive tool I have created, aligns with the evolving landscape of telehealth, offering a convenient means for individuals to assess their cardiovascular health.

As with any scientific endeavor, this project is not without limitations. My model's accuracy is contingent on the quality and diversity of the dataset used for training, and ongoing refinement is essential to ensure its applicability across diverse demographic groups. Ethical considerations, such as data privacy and the responsible use of predictive technologies, must be paramount in the ongoing development and deployment of such models.

In essence, my heart attack prediction model not only showcases the potential of modern technologies in healthcare but also underscores the importance of interdisciplinary approaches in addressing complex medical challenges. As our society continues to evolve, and rapidly change, an era near us is emerging where the future revolves around technology and AI. My model serves as a testament to the transformative power of innovation and its potential to positively impact public health.

Recommendations/Applications/The Next Steps

Future Improvements

Continuous Refinement through Data Updates:

- Implement a system for regular updates and recalibration of the model based on the latest cardiovascular health data.

Integration of Additional Biomarkers:

- Expand the model's parameters to include emerging biomarkers or health metrics that might further enhance predictive accuracy.

User-Friendly Interface and Accessibility:

- Develop a user-friendly application or interface that allows individuals to easily input their information for risk assessment, promoting widespread accessibility.

Incorporation of Genetic Data:

- Explore the integration of genetic information into the model to enhance its precision in predicting hereditary cardiovascular risks.

Collaboration with Healthcare Professionals:

- Establish partnerships with healthcare professionals to validate the model's predictions and ensure seamless integration into clinical practices.

Real-Life Applications

Early Intervention in Healthcare:

- Enable healthcare providers to identify individuals at high risk of heart attacks early, allowing for proactive intervention and personalized preventive measures.

Resource Allocation in Public Health:

- Assist public health agencies in targeting resources more efficiently by identifying populations or regions with a higher likelihood of cardiovascular issues.

Corporate Wellness Programs:

- Companies could integrate the heart attack prediction model into their wellness programs, offering employees personalized health insights and interventions to improve overall cardiovascular health.

Educational Initiatives for At-Risk Populations:

- Use the model to identify communities or demographics at a higher risk of heart attacks and implement targeted educational campaigns to raise awareness and promote healthier lifestyles.

The Next Step

1. Validation and Testing

2. Clinical Trials and Ethical Approvals

3. User Interface Development

4. Partnerships with Healthcare Providers