# Predicting Heart Health

Isha Saldhi

# Background

Heart disease is a leading cause of death in the United States. Some commonly considered risk factors are high blood pressure, high cholesterol, and smoking.

The CDC conducts telephone surveys to collect the health status of U.S. residents as part of the Behavioral Risk Factor Surveillance System (BRFSS).

There are many possible factors that could contribute to heart disease such as BMI, diabetes status,

# Methods

Creating a Neural Network to use someone’s health features to predict if someone is at risk for heart disease.

Creating a decision tree in order to determine which factors are most important in predicting heart health.

# Deep Learning Model

**Deep Learning Model**

```
# Define the deep learning model
nn_model = tf.keras.models.Sequential()
nn_model.add(tf.keras.layers.Dense(units=16, activation="relu", input_dim=30))  # Adjusted input_dim to match the nu
nn_model.add(tf.keras.layers.Dense(units=16, activation="relu"))
nn_model.add(tf.keras.layers.Dense(units=1, activation="sigmoid"))

# Compile the Sequential model together and customize metrics
nn_model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"])

# Train the model
fit_model = nn_model.fit(X_train_scaled, y_train, epochs=50)

# Evaluate the model using the test data
model_loss, model_accuracy = nn_model.evaluate(X_test_scaled, y_test, verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

The Deep Learning Model had an accuracy in training of 94.64%.

Using this model would let us predict one's risk of heart disease.

To find which features are most important to watch, I created a decision tree.

```
Epoch 50/50
5767/5767 [==============================] - 14s 2ms/step - loss: 0.1453 - accuracy: 0.9487
1923/1923 - 3s - loss: 0.1525 - accuracy: 0.9464 - 3s/epoch - 2ms/step
Loss: 0.15249761939048767, Accuracy: 0.9464117288589478
```

# Decision Tree

```
# Calculating the confusion matrix
cm_rf = confusion_matrix(y_test, predictions_rf)
cm_df_rf = pd.DataFrame(
    cm_rf, index=["Actual 0", "Actual 1"], columns=["Predicted 0", "Predicted 1"]
)

# Calculating the accuracy score
acc_score_rf = accuracy_score(y_test, predictions_rf)

# Displaying results
print("Confusion Matrix")
display(cm_df_rf)
print(f"Accuracy Score : {acc_score_rf}")
print("Classification Report")
print(classification_report(y_test, predictions_rf))
```

The Deep Learning Model had an accuracy in training of 94.64%.

Using this model would let us predict one's risk of heart disease.

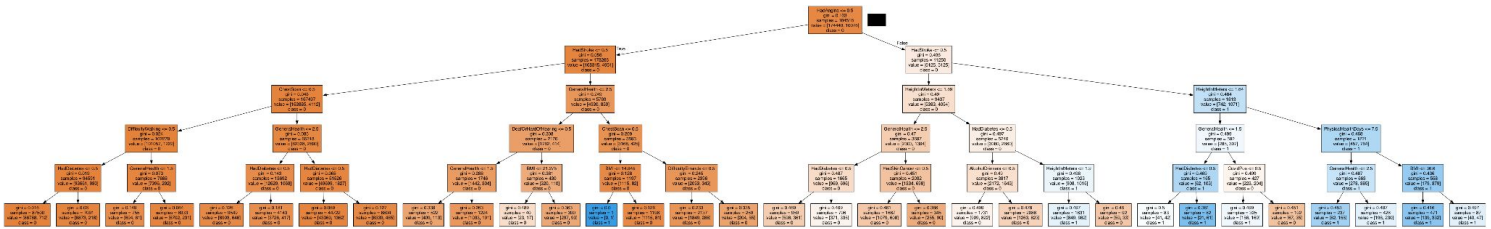
Confusion Matrix

| | Predicted 0 | Predicted 1 |
|---|---|---|
| **Actual 0** | 57717 | 430 |
| **Actual 1** | 2888 | 471 |

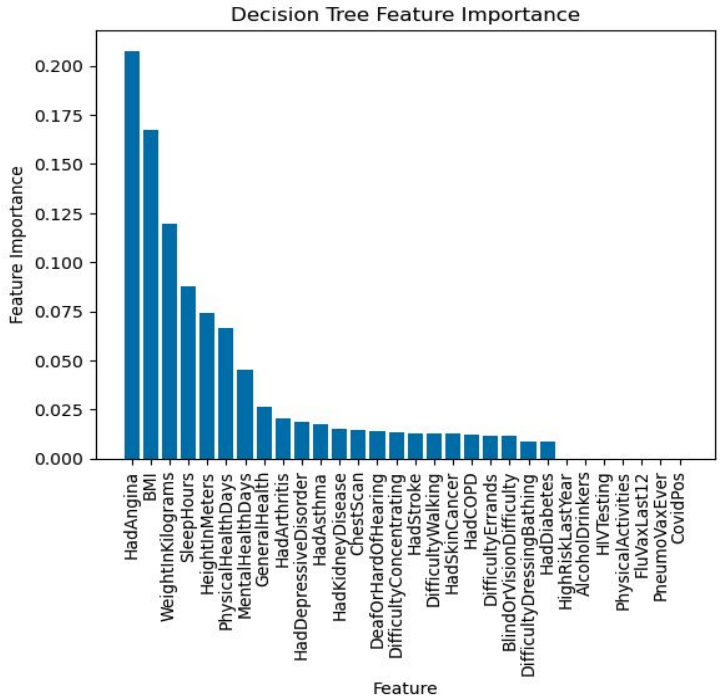
Accuracy Score : 0.9460540435079504
Classification Report

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.95 | 0.99 | 0.97 | 58147 |
| 1 | 0.52 | 0.14 | 0.22 | 3359 |
| accuracy | | | 0.95 | 61506 |
| macro avg | 0.74 | 0.57 | 0.60 | 61506 |
| weighted avg | 0.93 | 0.95 | 0.93 | 61506 |

To find which features are most important to watch, I created a decision tree.



# Decision Tree Feature Analysis



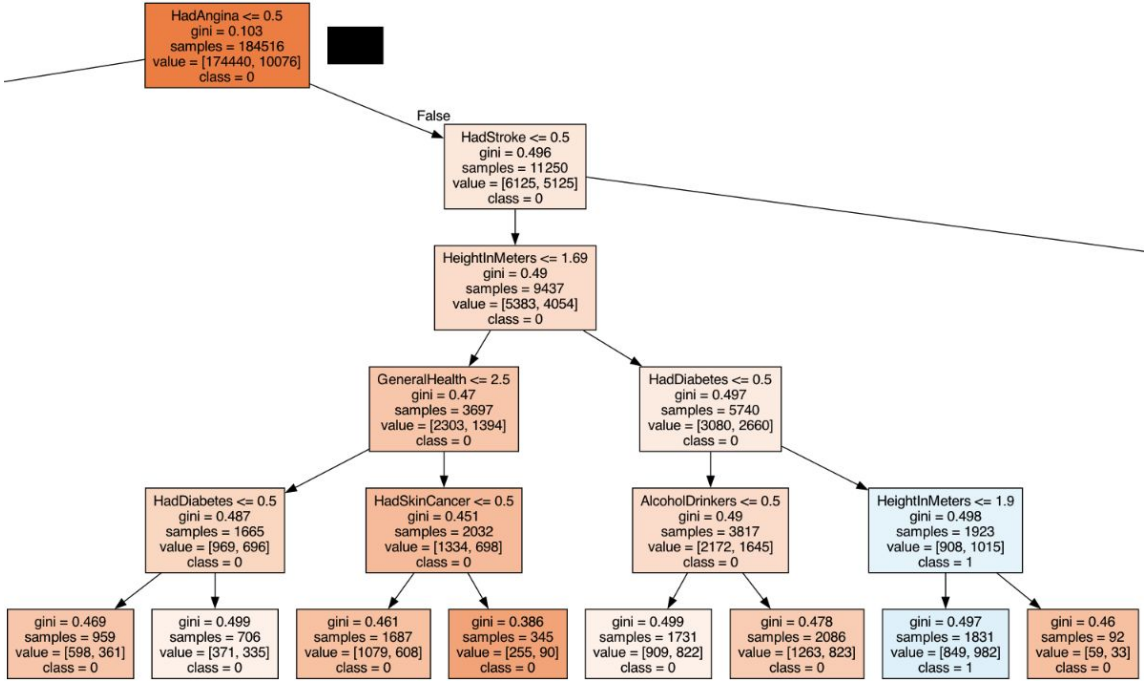

Most important factors identified are HadAngina, BMI, WeightInKilograms, and SleepingHours.

But this doesn't show us which features together create the highest risk.

# Sub-Branch Analysis Pt 1

HadAngina is a big deciding factor.

More at risk HadDiabetes and between 1.69 and 1.9 m

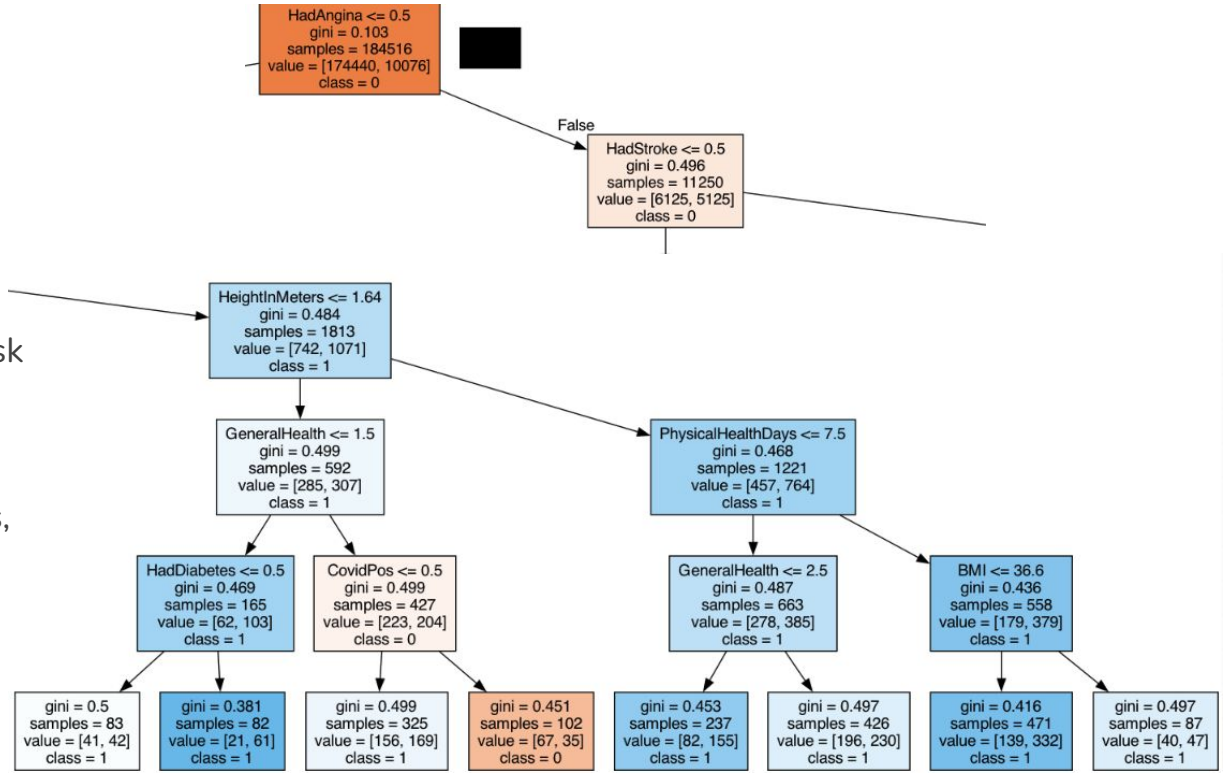


# Sub-Branch Analysis Pt 2

HadAngina is a big deciding factor, and if someone also has had a stroke they are at much higher risk of heart disease.

Unexpected exceptions: lower risk if CovidPos; higher risk if bigger HeightInMeters

Expected High risk: HadDiabetes, BMI, GeneralHealth



# Conclusion

Machine learning can be very helpful in the healthcare industry.

Having access to technology and algorithms like this could help with catching health risks earlier and is a valuable preventative tool.

This sort of tool could also help people at home to assess their own health and know which symptoms might be an indicator of a bigger problem.