Heart Disease Risk Prediction

Leo Dai & Lupin Cai

MOTIVATION

- **Public Health Burden**: Heart attacks are a leading cause of death worldwide. They contribute significantly to the burden of cardiovascular diseases, impacting public health systems. The high prevalence of heart attacks necessitates extensive medical services, from emergency care to long-term rehabilitation.
- **Enhanced Risk Prediction**: Integrate a wide range of factors (demographic, clinical, lifestyle) for a more accurate, personalized risk assessment.
- **Preventive Healthcare**: Identify high-risk individuals early, allowing for targeted preventive measures and lifestyle modifications, potentially reducing heart attack occurrences.
- **Resource Optimization**: Efficiently allocate healthcare resources by focusing on those at highest risk, improving healthcare system efficiency and reducing costs.

Related Works

- **AI and Deep Learning Models for Cardiovascular Risk Prediction**: The American Heart Association highlighted several key findings from their 2023 scientific sessions. One study demonstrated the effectiveness of AI in analyzing heart sound data from a digital stethoscope for detecting heart valve disease, outperforming traditional methods. Another study used AI and deep learning to analyze eye images of individuals with prediabetes and Type 2 diabetes to predict their risk of cardiovascular events, such as heart attacks and strokes
- **Machine Learning in Cardiovascular Disease Prediction**: A meta-analysis published in "Scientific Reports" reviewed various machine learning (ML) algorithms used for predicting cardiovascular diseases, including coronary artery disease (CAD). The study found that boosting algorithms showed promising results for CAD prediction, with area under the curve (AUC) of 0.88.

Methodologies

- Preprocessing & Data Cleaning
- Feature Selection
- Model Selection
- Hyperparameter Tuning
- Model Assessment

Preprocessing & Data Cleaning

- **Drop NaNs in the data with .dropna**
- **Encode categorical data into numerical:**
  - **Binary: Yes = 1, No = 0**
  - **Non-Binary: Apply one-hot encoding that create a new binary column for each categorical value in the column with numerical values, and replace the original column with the concatenated new columns**

Feature Selection

- Lasso coefficient:

```
# Feature selection with Lasso
y = data['HadHeartAttack']
X = data.drop('HadHeartAttack', axis=1)
lasso = LassoCV().fit(X, y)
important_features = lasso.coef_ != 0
X_selected = X.loc[:, important_features]

selected_feature_names = X.columns[important_features]
print("Selected Features:", selected_feature_names.tolist())
```

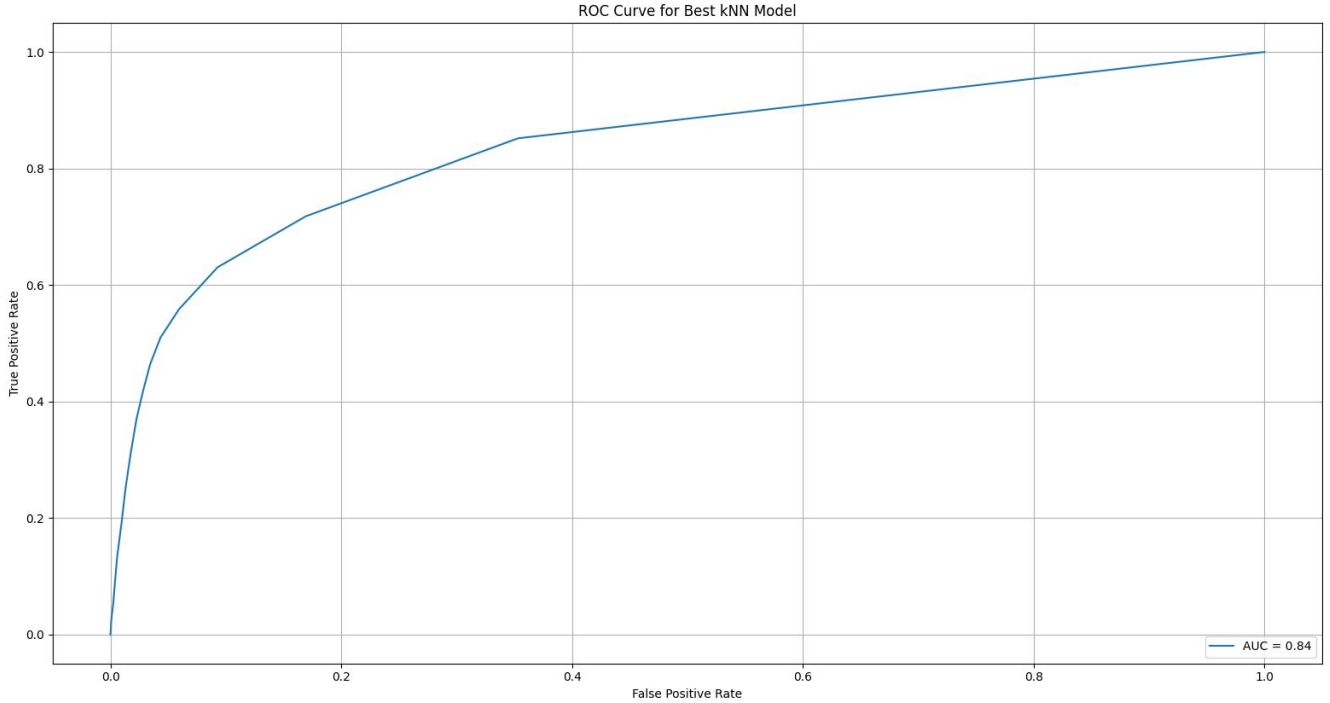
Model Selection

- Knn
- Decision Tree
- Neural Network

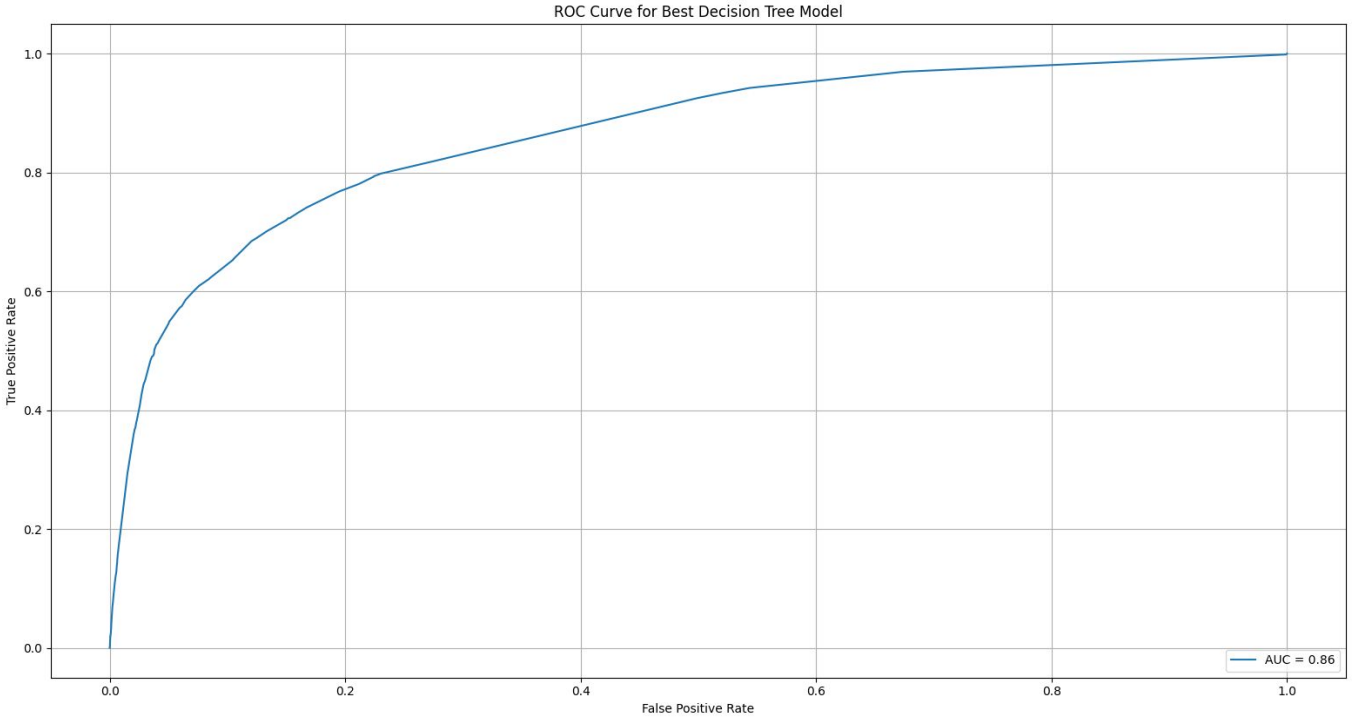
Hyperparameter Tuning using K-fold Grid Search

- Knn: metric = Euclidean, k = 23
- Decision Tree: criterion = gini, max tree depth = 7, min leaf sample = 5
- Neural Network: activation = relu, alpha = 0.05, hidden layer size = (10,), learning rate = constant, solver = adam

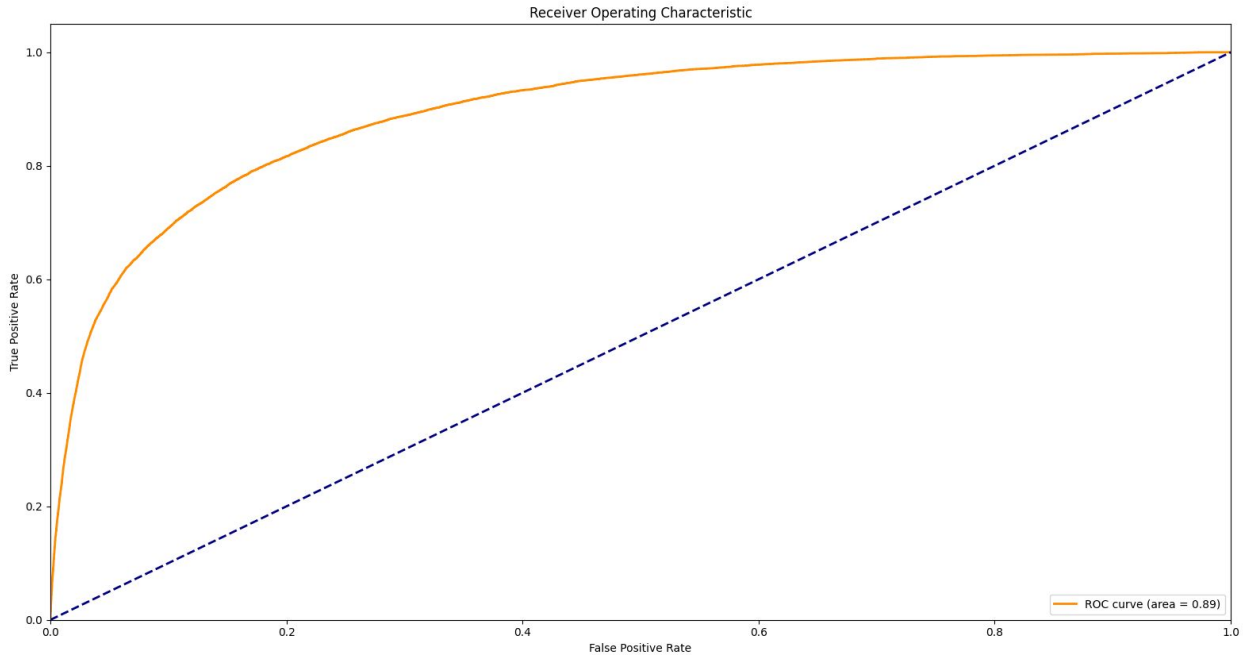
Model Assessment: Knn



Model Assessment: Decision Tree



Model assessment: Neural Network



Summarized Results

	Accuracy	AUC
Knn	0.947	0.84
Decision Tree	0.9479	0.86
Neural Network	0.949	0.89

DEMO & Project link

Link to demo: http://100.25.153.185/

Link to github: https://github.com/LupinC/CS334Project

Thanks for Watching