Heart Failure Prediction

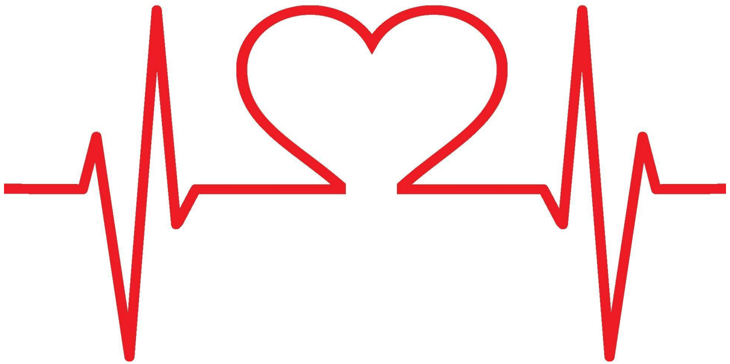
Meelad Doroodchi
ITCS 5154 Final Project

Title Slide

Title: Heart Failure Prediction Using Machine Learning
Author: Andrew Mvd (user on Kaggle to upload the dataset)
Year: Dataset year (2020)
Conference/Journal Name: Kaggle Datasets
My Name: Meelad Doroodchi
Link to Dataset:
https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data

Problems & Challenges

- Problem Statement
  - Predicting heart failure outcomes is critical for improving patient care.
  - Mortality rates are high, and early prediction can help intervene effectively.
    - Mortality rates are 1 in 5
- Challenges
  - Class imbalance in the dataset (more survivors than deaths).
  - Identifying relevant features that contribute significantly to mortality prediction.



Motivation

- Why This Problem?
    - Heart failure affects millions globally, with high mortality rates.
    - Accurate predictions can help in clinical decision-making.
    - Real life effect and be able to help people predict and understand their health
    - Healthcare is key for certain outcomes such as false negatives since telling a patient they do not have heart failure when they do is a major problem.
- Why This Dataset?
    - Contains clinical data that is readily available in real-world scenarios.
    - A mix of demographic, clinical, and lab-related features makes it practical.

Related Work(Primary Paper being Reproduced)

Title: *Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm*
This paper uses a stacked ensemble of machine learning models, including Random Forest, XGBoost, and Decision Trees, to predict heart failure survival. It addresses class imbalance using SMOTE and achieves high accuracy (around 99.98%). This would be a great paper to reproduce and the one that I want to because it offers a clear methodology and publicly available data.
Dataset: 299 heart failure patients from the Faisalabad Institute of Cardiology.
Method: Ensemble methods combining supervised classifiers.
Source: arXiv.org

Why This Paper?

- Paper Chosen: *Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm*
- Key Methodology
    - Ensemble learning (Stacked Ensemble) for robust predictions.
    - Combined the strengths of multiple models: Decision Trees, XGBoost, and Logistic Regression.
    - The stacked ensemble approach leverages multiple models to improve accuracy.
    - It aligns with the challenges of heart failure prediction.
    - Paper also describes a dataset with 299 patients in it just like this current dataset.

Other Related Works

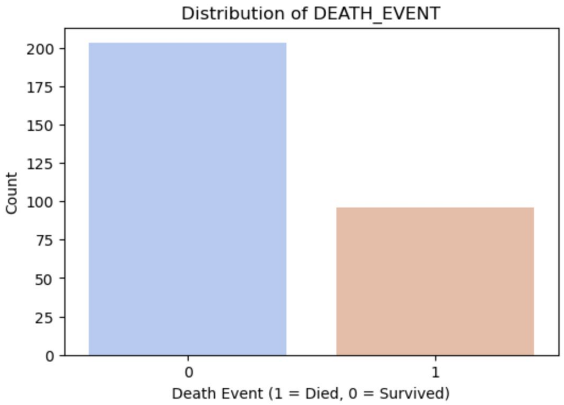
Title: ***Heart Failure Survival Prediction using Various Machine Learning Models***
This study compares several models, including Random Forest and XGBoost, applied to the UCI heart failure dataset. The paper provides useful insights into the performance of various machine learning models for heart failure prediction, achieving 88% accuracy using LightGBM.
Method: LightGBM, Random Forest, decision trees.
Source: PeerJ (PeerJ).
Title: ***Survival Prediction of Heart Failure Patients using Serum Creatinine and Ejection Fraction***
This paper focuses on predicting heart failure survival using only two critical clinical features—serum creatinine and ejection fraction—achieving strong predictive power using logistic regression and decision trees. It demonstrates the value of simpler, interpretable models in clinical settings.
Dataset: 299 patients with heart failure.
Method: Logistic regression, Random Forest.
Source:(BioMed Central).

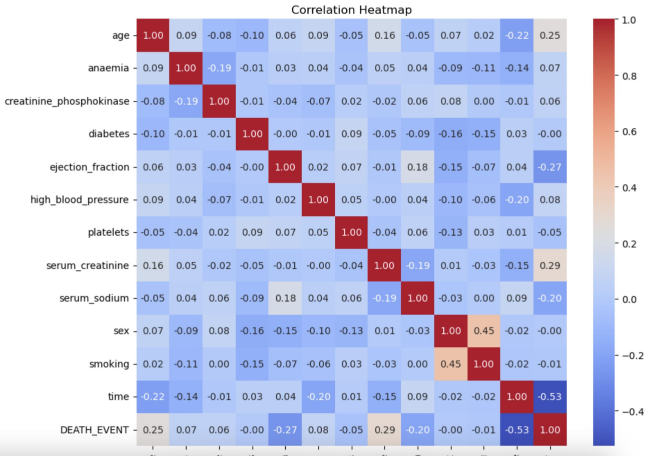
Methodology

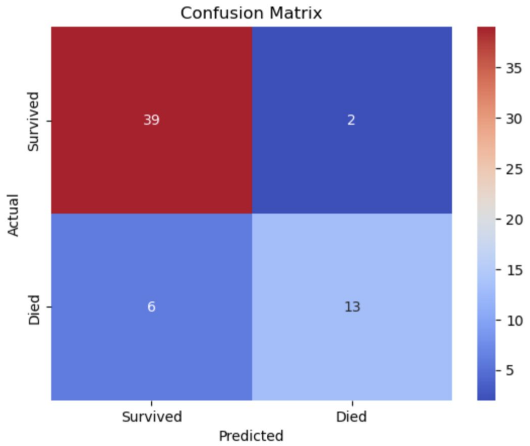
- Steps Followed
  - Preprocessed the dataset (handled missing values, scaled features, addressed class imbalance using SMOTE).
  - Explored the dataset with visualizations (correlation heatmap, boxplots, target distribution).
  - Implemented a stacked ensemble method combining Random Forest, XGBoost, and Logistic Regression.
  - Evaluated the model using metrics like confusion matrix, ROC curve, and AUC.
- Preprocessing Steps
  - Handling missing values.
  - Normalization of continuous features.
  - SMOTE for class imbalance.
- Stacked Ensemble
  - Base models: Decision Tree, XGBoost, Logistic Regression.
  - Meta-model: Logistic Regression.
  - Evaluation metrics: AUC, ROC, confusion matrix.
- Workflow
  - Dataset → Base Models → Meta-model → Final Predictions

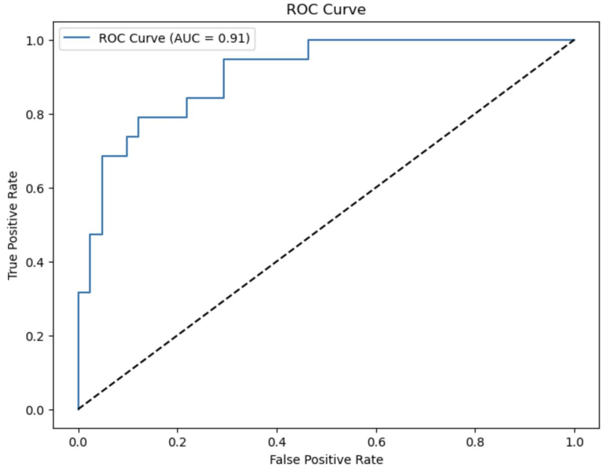
Results and Observations

- Observations
    - The stacked ensemble model performed well, achieving a high AUC.
    - It effectively handled class imbalance after using SMOTE.
    - ROC curve with AUC = 0.91.
    - Showing the stacked ensemble model performs well in distinguishing between the two classes.
- EDA visualizations also here









Conclusion & Future Works

- Serum creatinine and ejection fraction emerged as the most important predictors of mortality.
- Age showed a weaker, but still relevant, correlation with the outcome.
- The stacked ensemble method is effective for predicting heart failure outcomes since it allows me to see the false positive and negative rate which in healthcare are the most important features.
- Key predictors include serum creatinine, ejection fraction, and age.
- Test on larger datasets.
- Explore deep learning methods.
- Develop a clinical application (e.g., a predictive dashboard)

References/Citations

1. Zaman, S. M. Mehedi, et al. "Survival Prediction of Heart Failure Patients Using Stacked Ensemble Machine Learning Algorithm." arXiv, 30 Aug. 2021, https://arxiv.org/abs/2108.13367.Accessed 15 Sept. 2024.
2. Mamun, Mamdud, et al. "Heart Failure Survival Prediction Using Various Machine Learning Models." *PeerJ Computer Science*, vol. 7, 2021, https://doi.org/10.7717/peerj-cs.1894.
Accessed 15 Sept. 2024.
3. Chicco, Davide, et al. "Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone." BMC Medical Informatics and Decision Making, vol. 20, no. 1, 2020, https://doi.org/10.1186/s12911-020-1023-5.
Accessed 15 Sept. 2024.
4. Andrew Mvd. "Heart Failure Clinical Records." *Kaggle*, 2020, https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data. Accessed 15 Sept. 2024.