## Predicting Heart Disease Severity using Machine Learning

By Yashas Raman, Raghav Sriram Group 14

#### Why Predict Heart Disease?

- Heart disease is a leading cause of death in humans
- Improve detection methods
- Project Goal: Detect presence and severity of heart disease from clinical features
- Tasks:
  - Binary classification: disease or no disease
  - Multiclass classification: severity (0–4 scale)



### Research Methods

#### Dataset & Preprocessing

- UCI Heart Disease dataset (303 patients)
- 13 Features
  - Age, sex, chest pain type, cholesterol, max heart rate, etc.
- Preprocessing:
  - Missing value imputation
  - Feature scaling

|                 |       |       |       |       | Fe    | eature | Correl | ation H | leatma | р     |       |       |       |       | _ | - 1.00 |
|-----------------|-------|-------|-------|-------|-------|--------|--------|---------|--------|-------|-------|-------|-------|-------|---|--------|
| age             | 1.00  | -0.10 | 0.10  | 0.28  | 0.21  | 0.12   | 0.15   | -0.39   | 0.09   | 0.20  | 0.16  | 0.36  | 0.13  | 0.22  |   | 1.00   |
| sex -           | -0.10 | 1.00  | 0.01  | -0.06 | -0.20 | 0.05   | 0.02   | -0.05   | 0.15   | 0.10  | 0.04  | 0.09  | 0.38  | 0.22  |   | - 0.75 |
| 8-              | 0.10  | 0.01  | 1.00  | -0.04 | 0.07  | -0.04  | 0.07   | -0.33   | 0.38   | 0.20  | 0.15  | 0.23  | 0.27  | 0.41  |   |        |
| trestbps        | 0.28  | -0.06 | -0.04 | 1.00  | 0.13  | 0.18   | 0.15   | -0.05   | 0.06   | 0.19  | 0.12  | 0.10  | 0.13  | 0.16  |   | - 0.50 |
| chol -          | 0.21  | -0.20 | 0.07  | 0.13  | 1.00  | 0.01   | 0.17   | -0.00   | 0.06   | 0.05  | -0.00 | 0.12  | 0.01  | 0.07  |   |        |
| tps<br>-        | 0.12  | 0.05  | -0.04 | 0.18  | 0.01  | 1.00   | 0.07   | -0.01   | 0.03   | 0.01  | 0.06  | 0.15  | 0.07  | 0.06  |   | - 0.25 |
| thalach restecg | 0.15  | 0.02  | 0.07  | 0.15  | 0.17  | 0.07   | 1.00   | -0.08   | 0.08   | 0.11  | 0.13  | 0.13  | 0.02  | 0.18  |   | - 0.00 |
| thalach         | -0.39 | -0.05 | -0.33 | -0.05 | -0.00 | -0.01  | -0.08  | 1.00    | -0.38  | -0.34 | -0.39 | -0.26 | -0.28 | -0.42 |   | 0.00   |
| exang           | 0.09  | 0.15  | 0.38  | 0.06  | 0.06  | 0.03   | 0.08   | -0.38   | 1.00   | 0.29  | 0.26  | 0.15  | 0.33  | 0.40  |   | 0.25   |
| oldpeak         | 0.20  | 0.10  | 0.20  | 0.19  | 0.05  | 0.01   | 0.11   | -0.34   | 0.29   | 1.00  | 0.58  | 0.30  | 0.34  | 0.50  |   |        |
| slope           | 0.16  | 0.04  | 0.15  | 0.12  | -0.00 | 0.06   | 0.13   | -0.39   | 0.26   |       | 1.00  | 0.11  | 0.29  | 0.38  |   | 0.50   |
| g -             | 0.36  | 0.09  | 0.23  | 0.10  | 0.12  | 0.15   | 0.13   | -0.26   | 0.15   | 0.30  | 0.11  | 1.00  | 0.26  | 0.52  |   |        |
| thal -          | 0.13  | 0.38  | 0.27  | 0.13  | 0.01  | 0.07   | 0.02   | -0.28   | 0.33   | 0.34  | 0.29  | 0.26  | 1.00  | 0.51  |   | 0.75   |
| m -             | 0.22  | 0.22  | 0.41  | 0.16  | 0.07  | 0.06   | 0.18   | -0.42   | 0.40   | 0.50  | 0.38  | 0.52  | 0.51  | 1.00  | 3 |        |

fbs restecgthalach exang oldpeak slope ca

cp trestbps chol

- -1.00

thal num

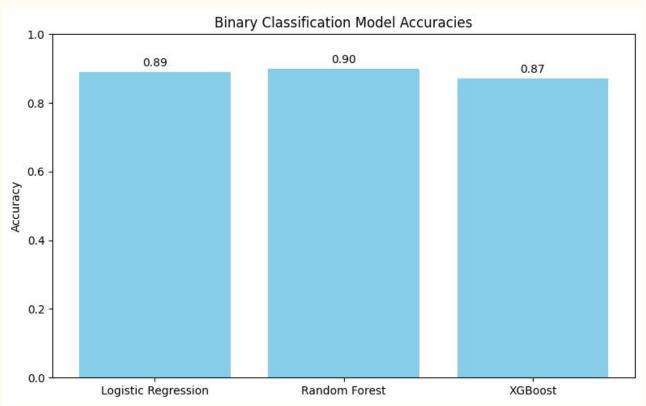
#### Model Development and Evaluation

- Used an 80/20 train-test split to train and evaluate our models
- Models Used:
  - Logistic Regression
  - Random Forest
  - XGBoost
  - Applied to both binary and multiclass tasks

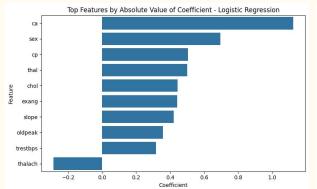
```
"Logistic Regression": LogisticRegression(max_iter=1000),
"Random Forest": RandomForestClassifier(n_estimators=100, random_state=42),
"XGBoost": XGBClassifier(eval_metric='logloss')
```

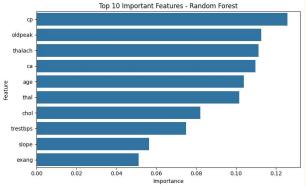
## Results and Analysis

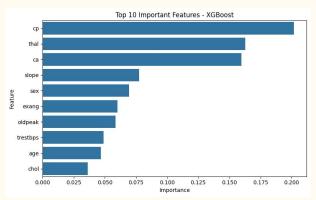
#### Binary Classification Results



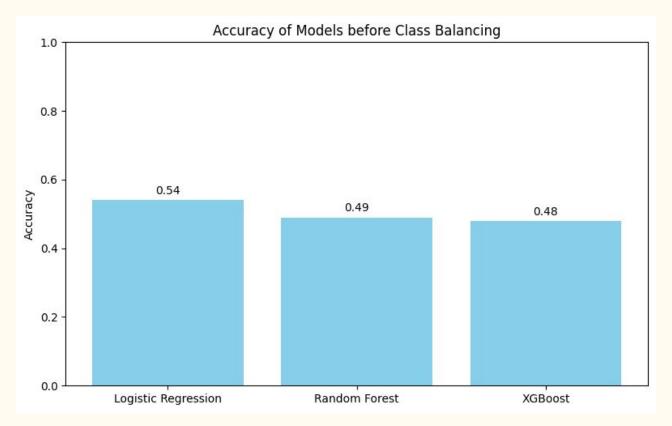
#### Most Relevant Features



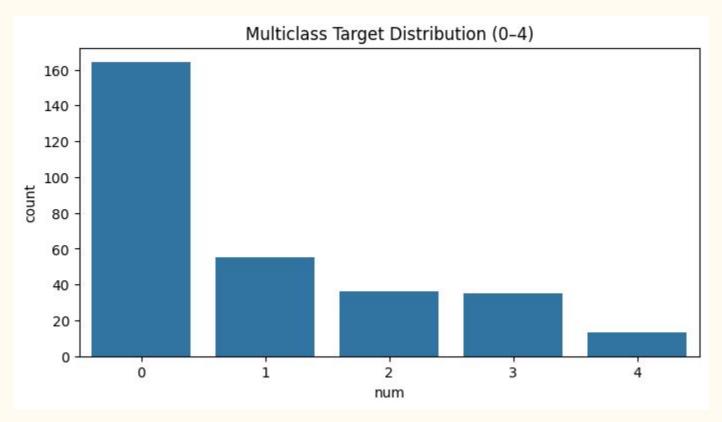




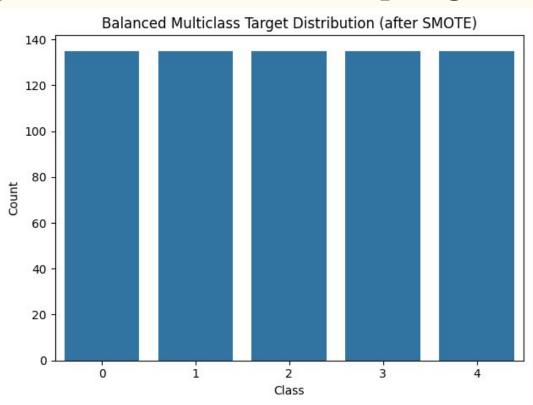
#### Results for Multiclass Prediction



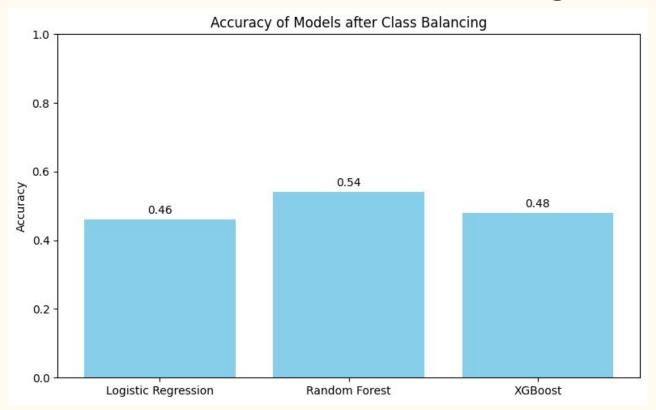
#### Unbalanced Multiclass Dataset



#### Balancing Dataset via Oversampling



#### Multiclass results after Class Balancing



#### Analysis

- All three models performed relatively similarly across binary and multiclass prediction
- Random Forest slightly improved after class balancing while Logistic Regression got slightly worse
  - Overall, class balancing did not help model performance
- All models performed very well on binary while significantly worse on multiclass

# Discussion and Reflection

#### What We Learned

- Machine learning can be very useful in detecting presence of heart disease
- Severe class imbalance likely harmed performance
- Multiclass prediction looks to be a difficult problem in general
- Relatively small sample size could have limited model generalization

#### Future Directions

- Potentially use other models
  - $\circ$  SVM
  - Neural Networks
- Attempt methods on more balanced dataset
- Apply to other diseases