

# Predicting Survival in Patients with Heart Failure

by  
Zulu Healthcare Analytics



# Client and Objectives

Who: Faisalabad Institute of Cardiology and Allied Hospital in Faisalabad

Problem: Can we predict survival in patients with heart failure? Can we identify the most important features?



# Data

## **Data source:**

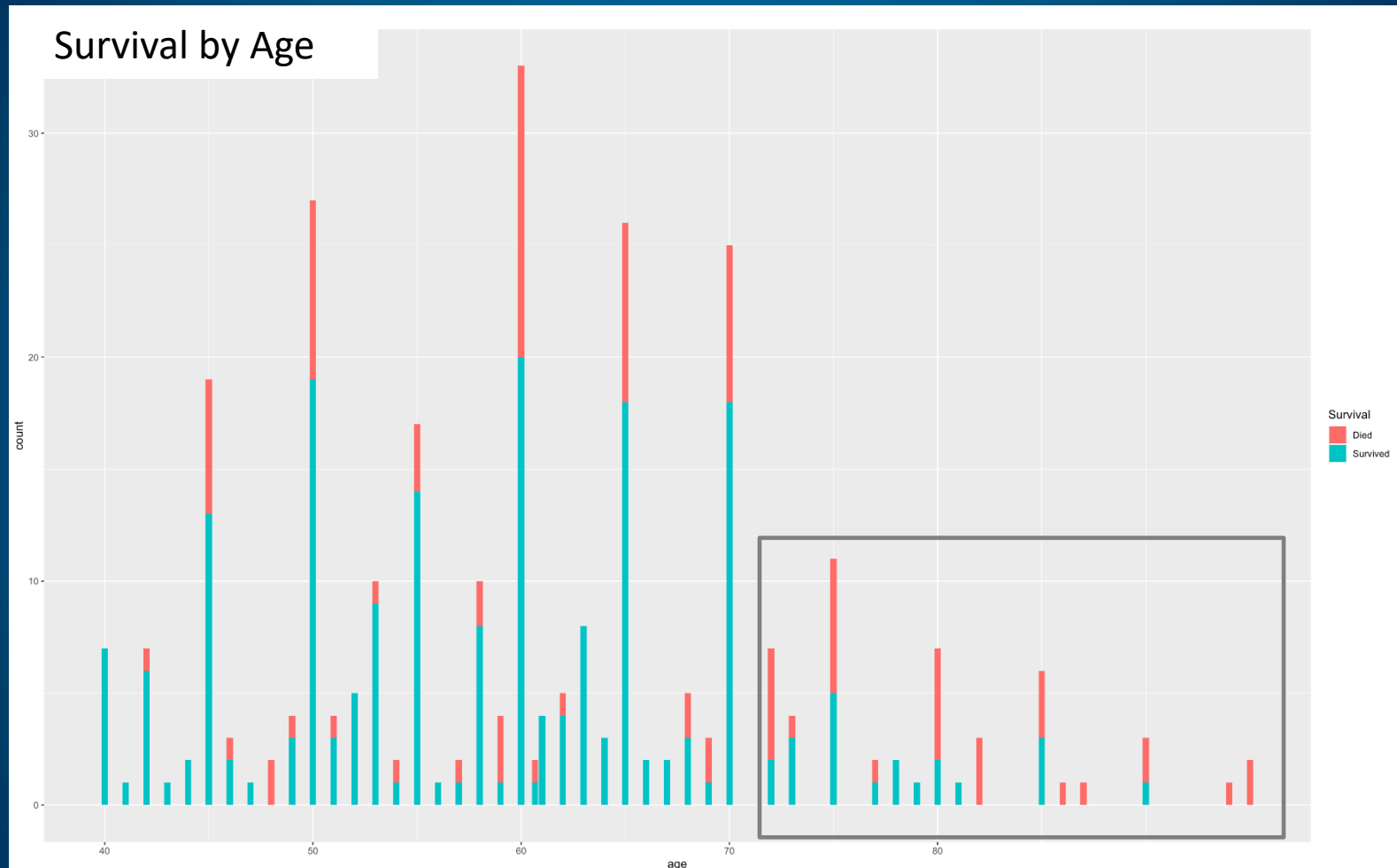
UCI Machine Learning  
Repository

**Originated:** Faisalabad  
Institute of Cardiology  
and Allied Hospital in  
Faisalabad (Pakistan) in  
2015.

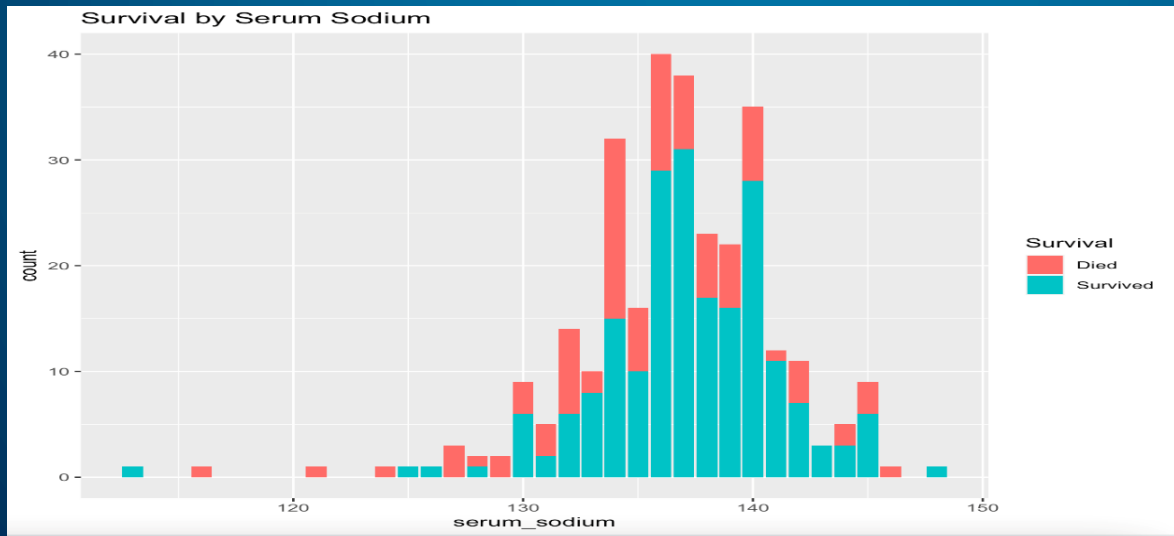
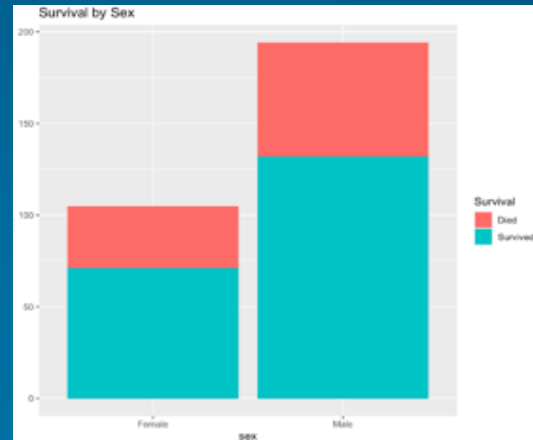
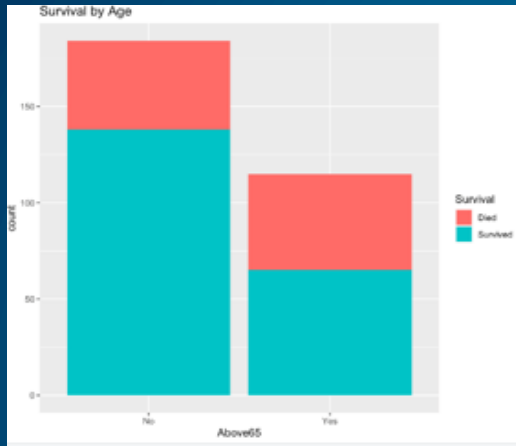
299 records  
(heart failure patients)

13 Features

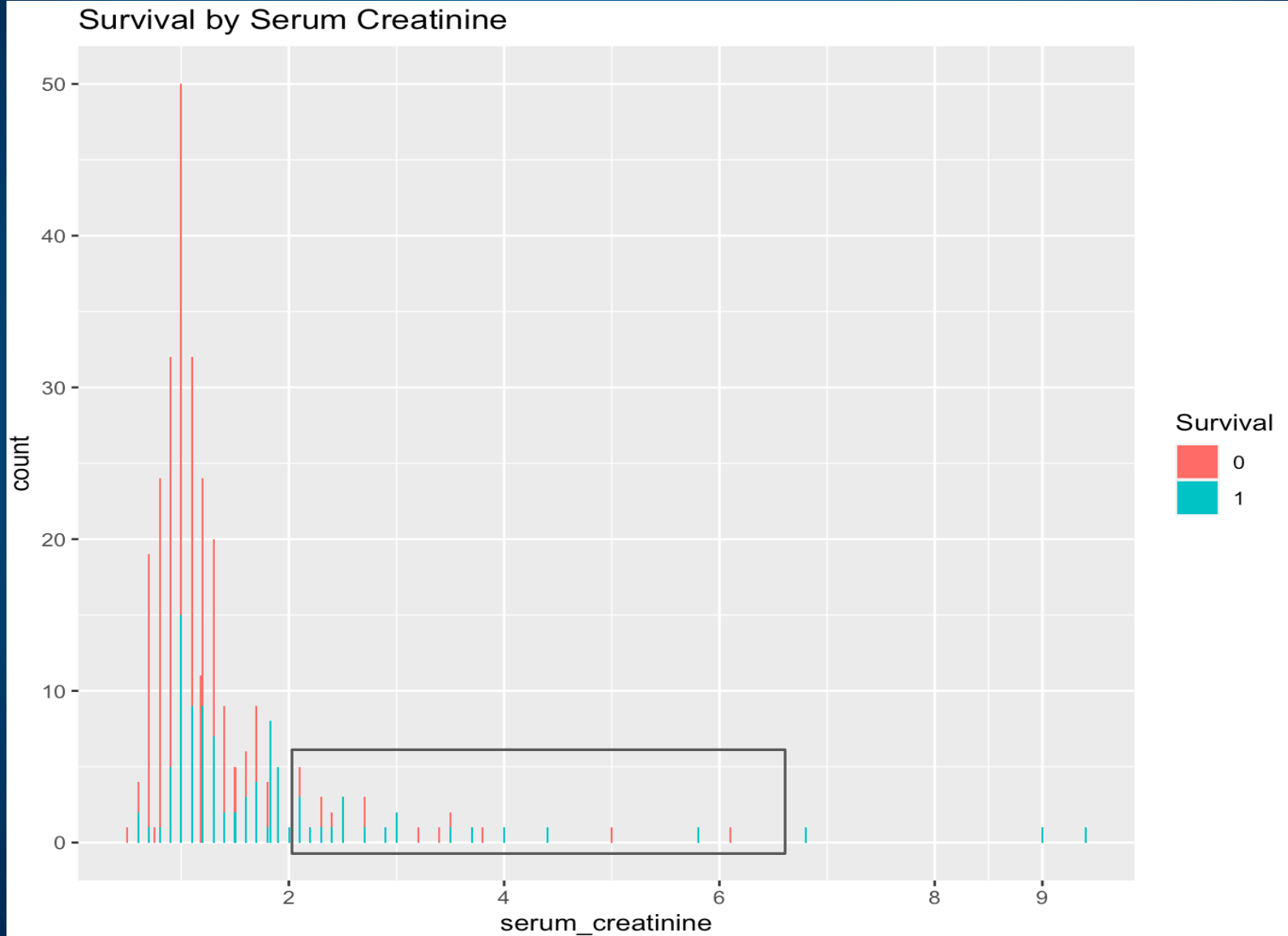
# Data Exploration



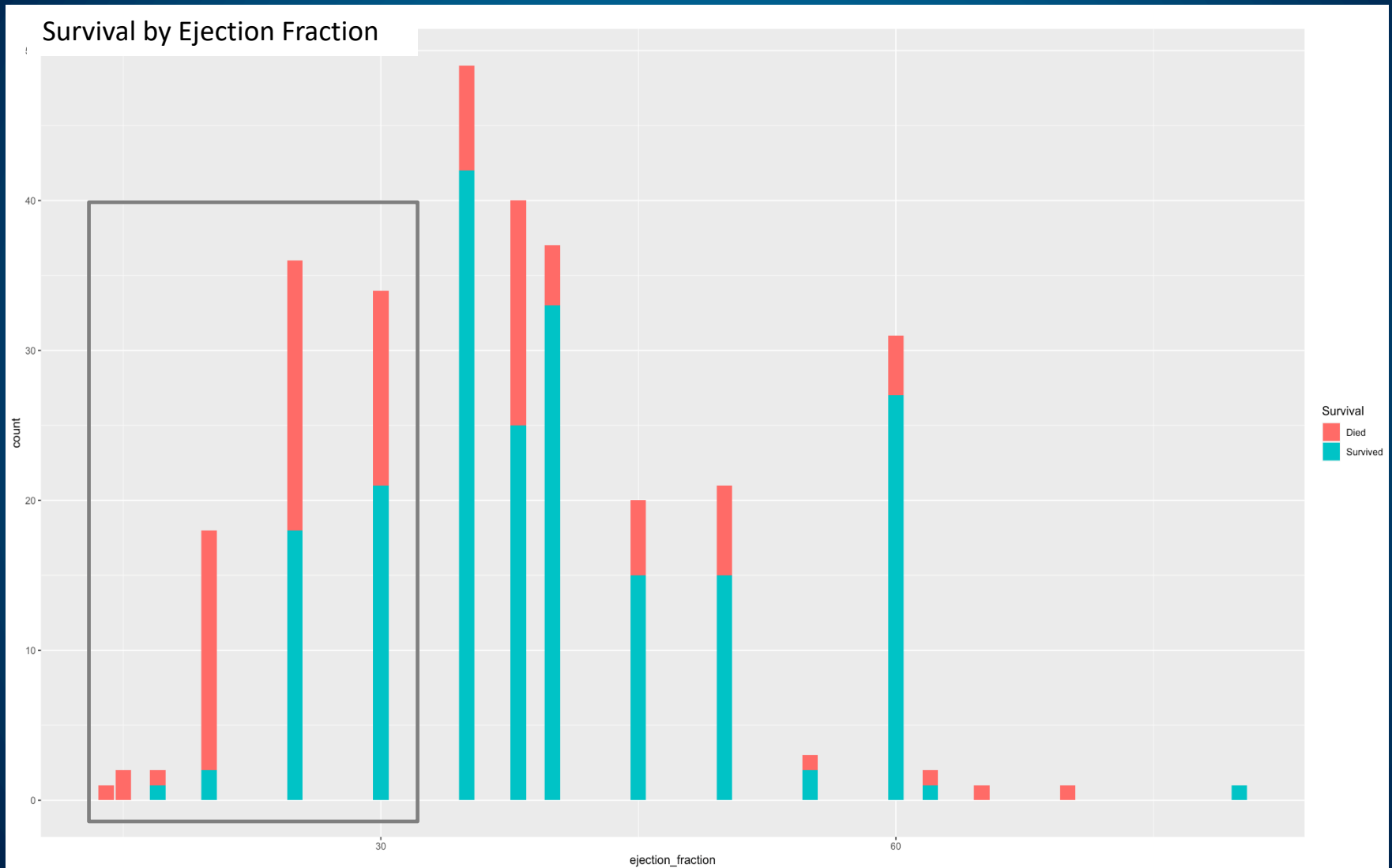
# Data Exploration cont.



# Data Exploration cont.



# Data Exploration cont.



# Preprocessing & Feature Selection

- 6 features into factors
- Age group
- Removed the feature 'time'
- No missing values
- Imbalanced data (ROSE – Training Set)
- Boruta for feature selection





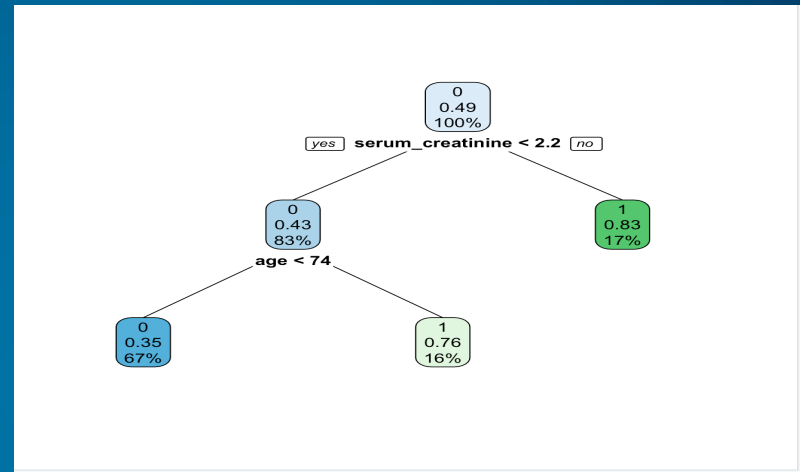
# Model Comparison

Model	Accuracy	Sensitivity	Specificity
With ROSE			
Random Forest	0.7841	0.9167	0.5000
Decision Tree	0.7273	0.8667	0.4286
GBM	0.7614	0.9500	0.3571
Logistic Regression	0.6364	0.6167	0.6786
KNN	0.6818	0.8500	0.3214
Without ROSE			
Random Forest	0.7386	0.9000	0.3929
Decision Tree	0.6818	0.8167	0.3929
GBM	0.6818	0.8000	0.4286
Logistic Regression	0.7273	0.9000	0.3571
KNN	0.6818	0.9667	0.0714

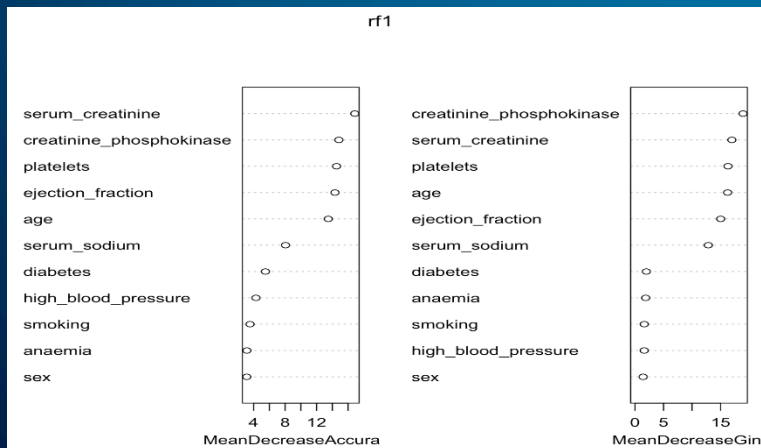
# Model Evaluation

A new data set was created with the top 6 features.

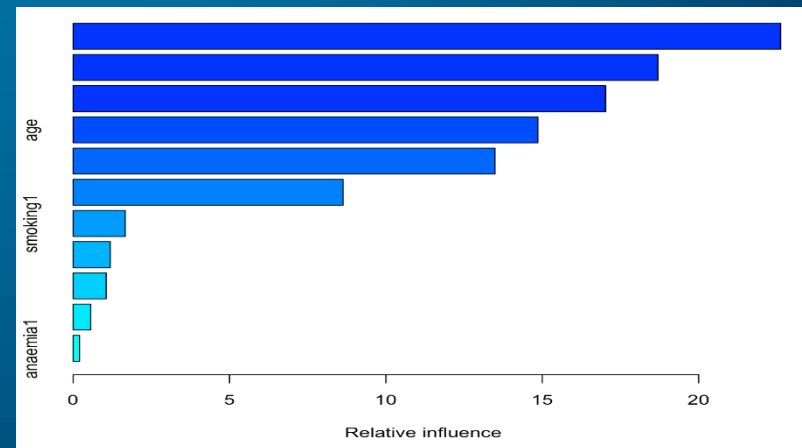
## Decision Tree














## Random Forest



## GBM

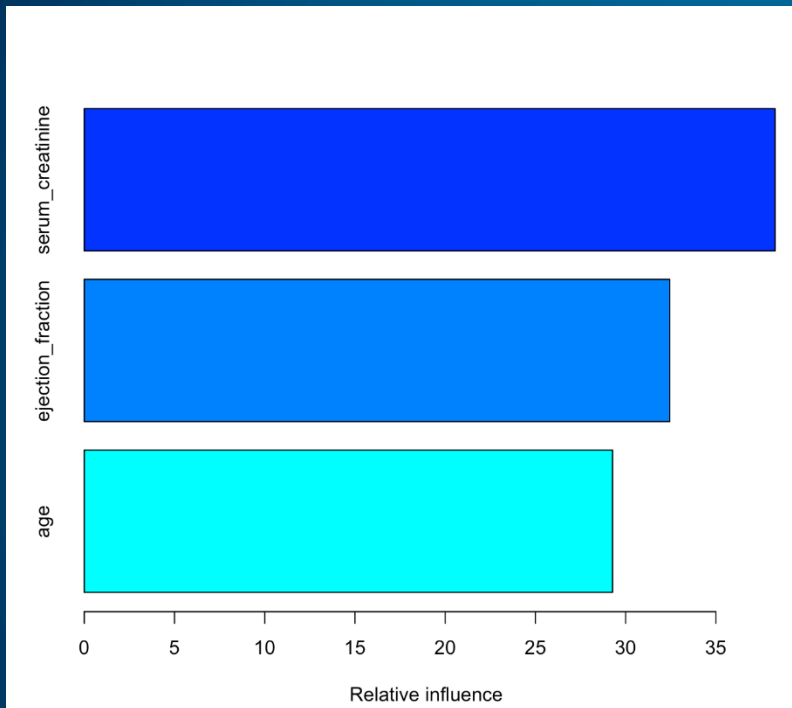


# Results

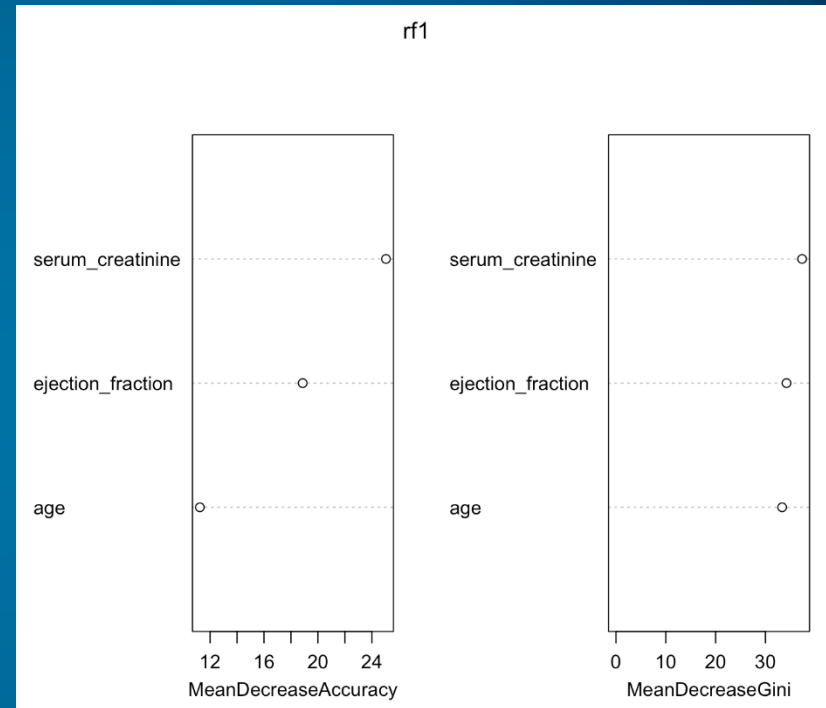
Model		Accuracy	Sensitivity	Specificity
Ejection Fraction, Serum Creatinine, Serum Phosphokinase, Age, Serum Sodium, Platelet				
Random Forest		0.7727	0.9000	0.5000
Decision Tree		0.7273	0.8667	0.4286
GBM		0.7614	0.9500	0.3571
Logistic Regression		0.7273	0.7667	0.6429
KNN		0.7273	0.9000	0.3571
Ejection Fraction, Serum Creatinine, Serum Sodium, Age (Boruta)				
Random Forest		0.7500	0.8167	0.6071
Decision Tree		0.7273	0.8667	0.4286
GBM		0.7955	0.8667	0.6429
Logistic Regression		0.7045	0.7500	0.6071
KNN		0.7045	0.8500	0.3929
Ejection Fraction, Serum Creatinine, Age				
Random Forest		0.8068	0.8833	0.6429
Decision Tree		0.7273	0.8667	0.4286
GBM		0.8182	0.9000	0.6429
Logistic Regression		0.6932	0.7333	0.6071
KNN		0.7500	0.8667	0.5000

# Results

## GBM

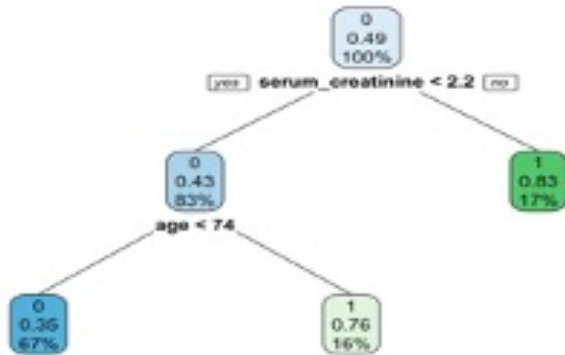


## Random Forest

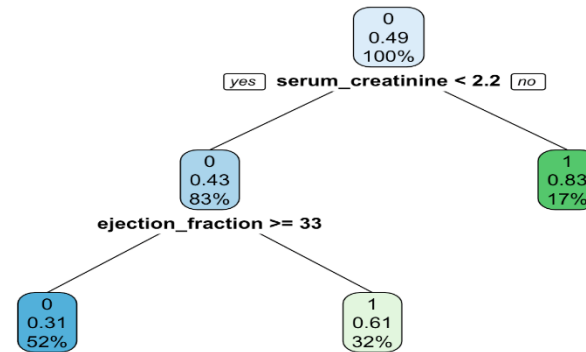


# Results

Model	Accuracy	Sensitivity	Specificity
Decision Tree	0.7273	0.8667	0.4286



Model	Accuracy	Sensitivity	Specificity
Decision Tree	0.7955	0.8833	0.6071



# Results

Model		Accuracy	Sensitivity	Specificity
Ejection Fraction, Serum Creatinine				
Random Forest	●	0.7727	0.8167	0.6786
Decision Tree	●	0.7955	0.8833	0.6071
GBM	●	0.8295	0.9500	0.5714
Logistic Regression	●	0.8295	0.8833	0.7143
KNN	●	0.7386	0.7333	0.7500

# Conclusion

ML models can predict survival of patients with heart failure.

Most important features:

- Serum Creatinine
- Ejection Fraction

# Limitations



The sample size is relatively small.



Additional patient information could be useful to the study such as body mass index, alcohol use, other medical conditions, genetic history, medications, and occupational history.



# Research & Cited Work

Mayo Clinic. 'Heart Failure'. 1998 -2022.

<https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142>

Package 'Boruta'. 2020. [https://cran.r-](https://cran.r-project.org/web/packages/Boruta/Boruta.pdf)

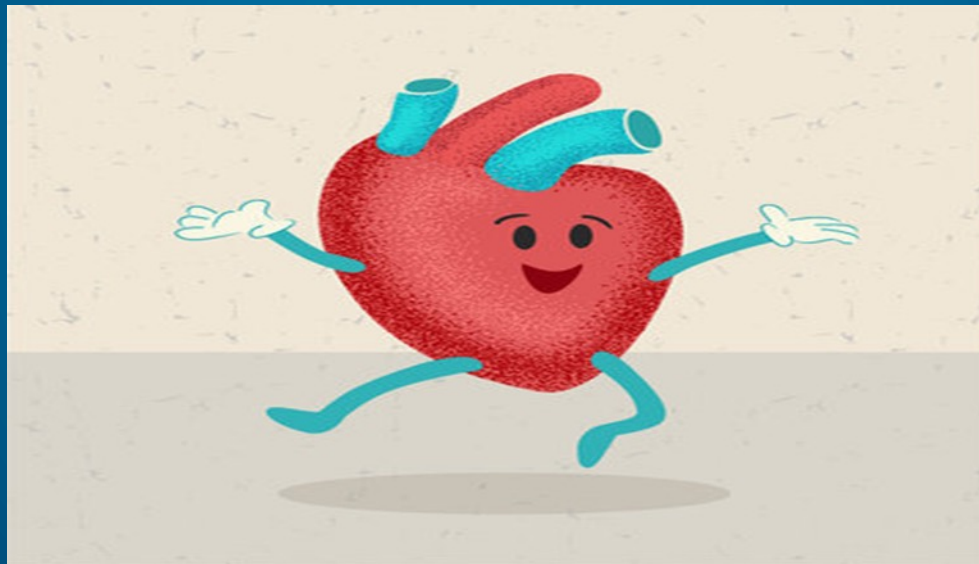
[project.org/web/packages/Boruta/Boruta.pdf](https://cran.r-project.org/web/packages/Boruta/Boruta.pdf)

Package 'ROSE'. 2021. [https://cran.r-](https://cran.r-project.org/web/packages/ROSE/ROSE.pdf)

[project.org/web/packages/ROSE/ROSE.pdf](https://cran.r-project.org/web/packages/ROSE/ROSE.pdf)

UCI Machine Learning Repository. Tanvir Ahmad, Assia Munir, Sajjad Haider Bhatti, Muhammad Aftab, and Muhammad Ali Raza. 2020. UCI Machine Learning Repository: Heart failure clinical records Data Set.

# Questions or Comments?



# Code Examples

```
#balance dataset  
data_rose <- ROSE(Survival ~., data = TrainingSet)$data  
table(data_rose$Survival)
```

```
#Feature selection  
boruta <- Boruta(Survival~., data = new_data, doTrace = 2, maxRuns = 500)  
print(boruta)
```

Without ROSE			
Model	Accuracy	Sensitivity	Specificity
Ejection Fraction, Serum Creatinine			
Random Forest	0.8977	0.9667	0.7500
Decision Tree	0.8182	1.0000	0.4286
GBM	0.8523	0.9667	0.6071
Logistic Regression	0.7841	0.9833	0.3571
KNN	0.8636	0.9833	0.6071