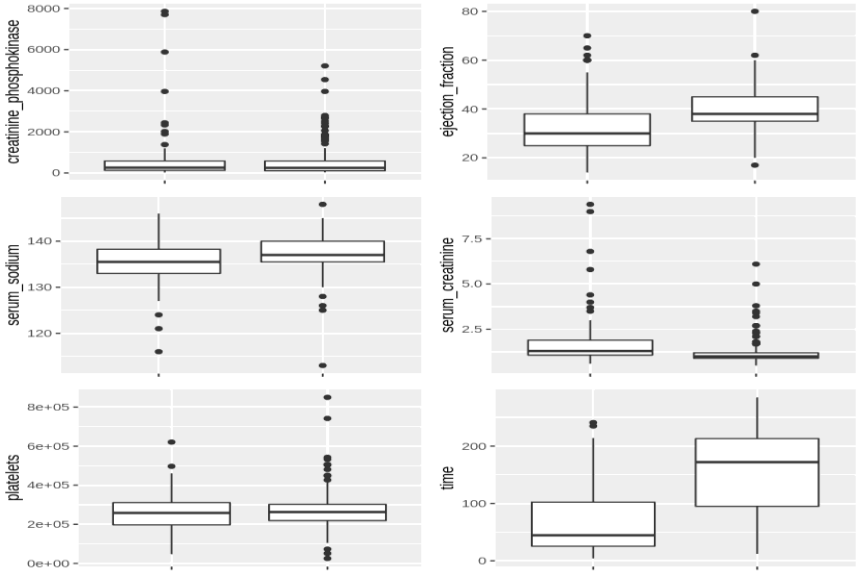
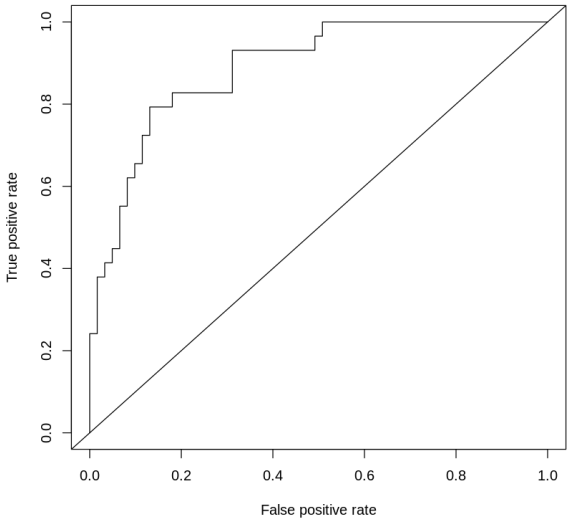
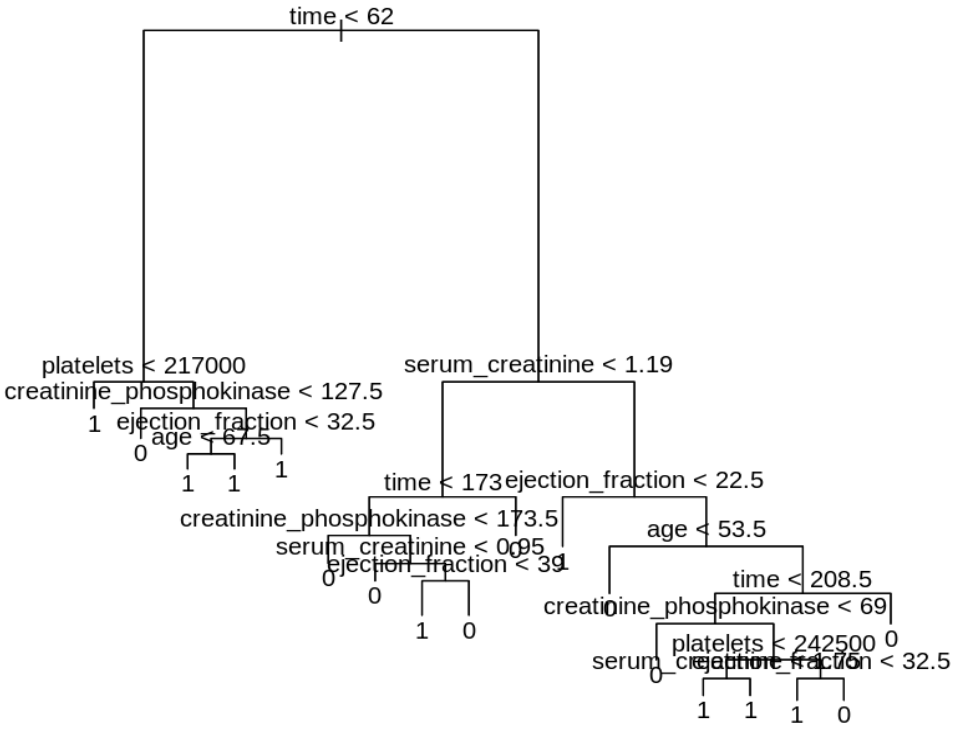
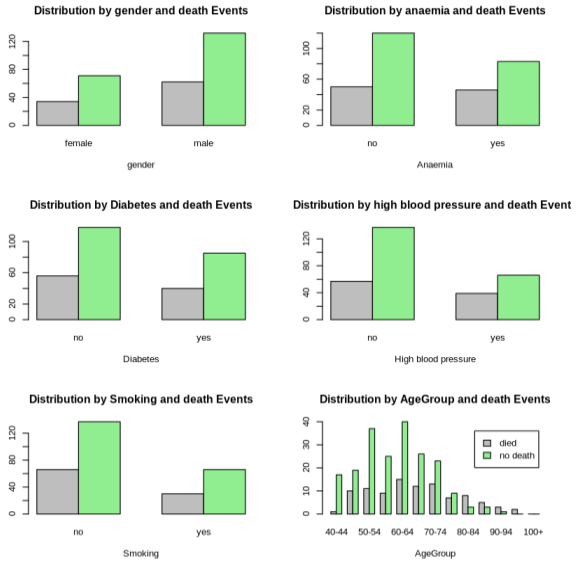
Heart Failure Clinical Record Study and Patients Survival Predictions using classification models.

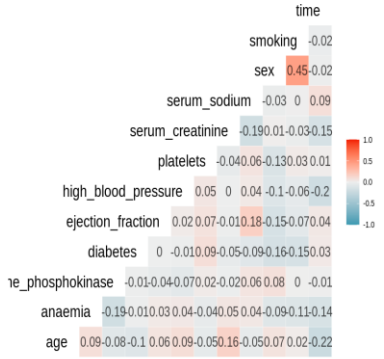
In this study, the heart failure clinical record dataset has been collected from Machine Learning Repository. (https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records#) The dataset is a cleaned one, there were no null values. An exploratory data analysis through correlation plots, and univariate and multivariate analysis of data, gives a clear picture that, no two variables are positively or negatively correlated. From the box plots, we can have a clear idea that time and ejection fraction are significant factors for predicting the survival of patients having heart failure.

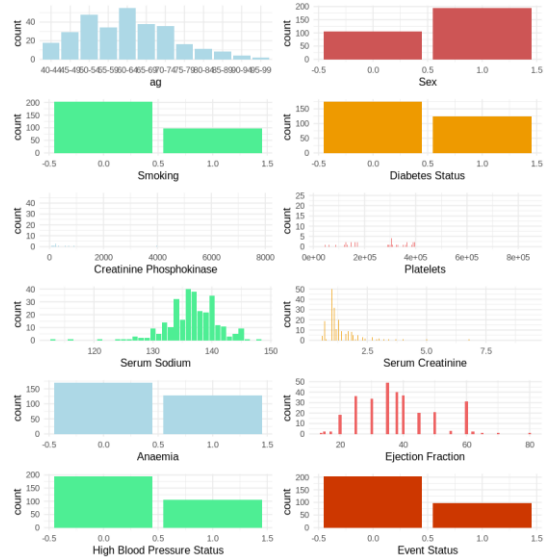












From the Monova and the increase in accuracy of the classification models, we can confirm that along with that, age, creatinine phosphokinase, platelets, and serum creatinine features play a significant role too. The latest logistic regression model predicts the survival of patients with an accuracy of 0.82 for unseen data, having a sensitivity of 0.62 and a specificity of 0.91, and a baseline accuracy of 0.67. However, The decision tree model predicts with an accuracy of 0.75