Titanic Report

Feature engineering and Preprocessing

Non-informative features like Name, Ticket, PassengerID were removed.

The missing values in Age were replaced by the median.The missing values in Embarked was replaced by mode. The cabin was dropped due to too many missing values.

Sex and Embarked were converted to numeric values using Label Encoding.

Continuous variables Fare and Age were normalized.

Influence

Females were far more likely to survive than males.
Younger passengers had marginally better survival.
Higher Fare increases probability of survival.
Passengers from Cherbourg (C) had slightly better survival chances.
A small number of family members increased survival odds, but large numbers hurt it.

Limitations

Dropping the Cabin column is a loss of some important information.
Logistic regression doesn't capture the interactions between people and their family members.

Idea to improvement

To use a better algorithm where it can capture the interactions between family members and also to enhance feature engineering by extracting the information of their socio economic status through their names.