

Analysis Report

AutoPrep

December 18, 2024

Abstract

This report has been generated with AutoPrep.

Contents

1	Overview	2
1.1	System	2
1.2	Dataset	2
1.3	Target Distribution	2
1.4	Missing Values	3
2	Exploratory Data Analysis	4
2.1	Numeric Features	5
2.2	Categorical Features	8
2.3	Correlation Analysis	11
3	Model Performance	12
3.1	Cross-validation Results	12
3.2	Test Set Performance	12
3.3	Confusion Matrix	13
4	Feature Importance Analysis	13

1 Overview

1.1 System

System	Darwin
Machine	arm64
Processor	arm
Architecture	64bit
Python Version	3.11.10
Physical Cores	8
Logical Cores	8
CPU Frequency (MHz)	4056
Total RAM (GB)	16.0000
Available RAM (GB)	4.7938
Total Disk Space (GB)	460.4317
Free Disk Space (GB)	263.7565

1.2 Dataset

Number of samples	1309
Number of features	13
Numeric features	6
Categorical features	7

1.3 Target Distribution

Category	Value
0	809
1	500

Table 1: Target Distribution

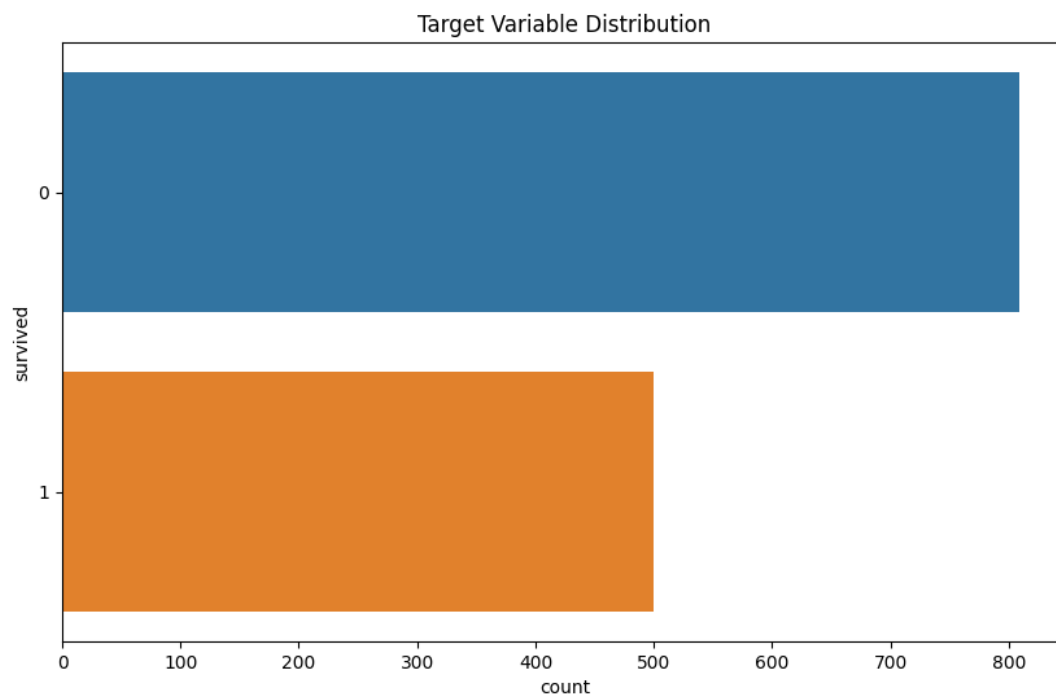


Figure 1: Target Variable Distribution

1.4 Missing Values

Category	Value
pclass	0
survived	0
name	0
sex	0
age	263
sibsp	0
parch	0
ticket	0
fare	1
cabin	1014
embarked	2
boat	823
body	1188
home_dest	564

Table 2: Missing Values

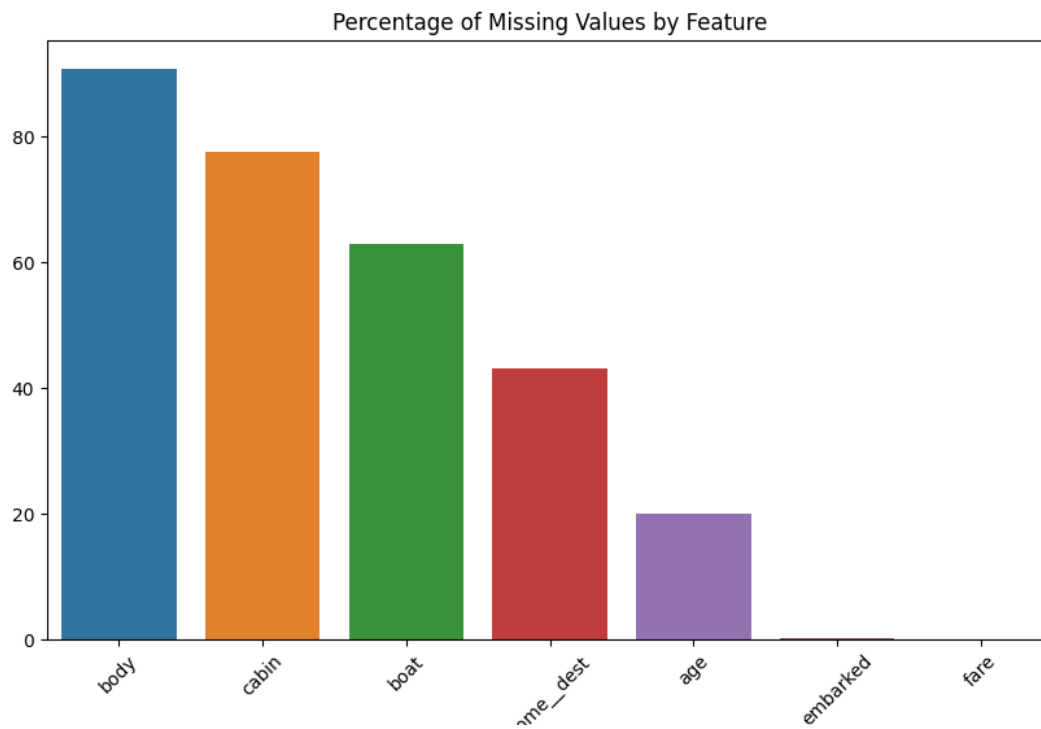


Figure 2: Missing Values Analysis

2 Exploratory Data Analysis

Visual analysis of the dataset features.

2.1 Numeric Features

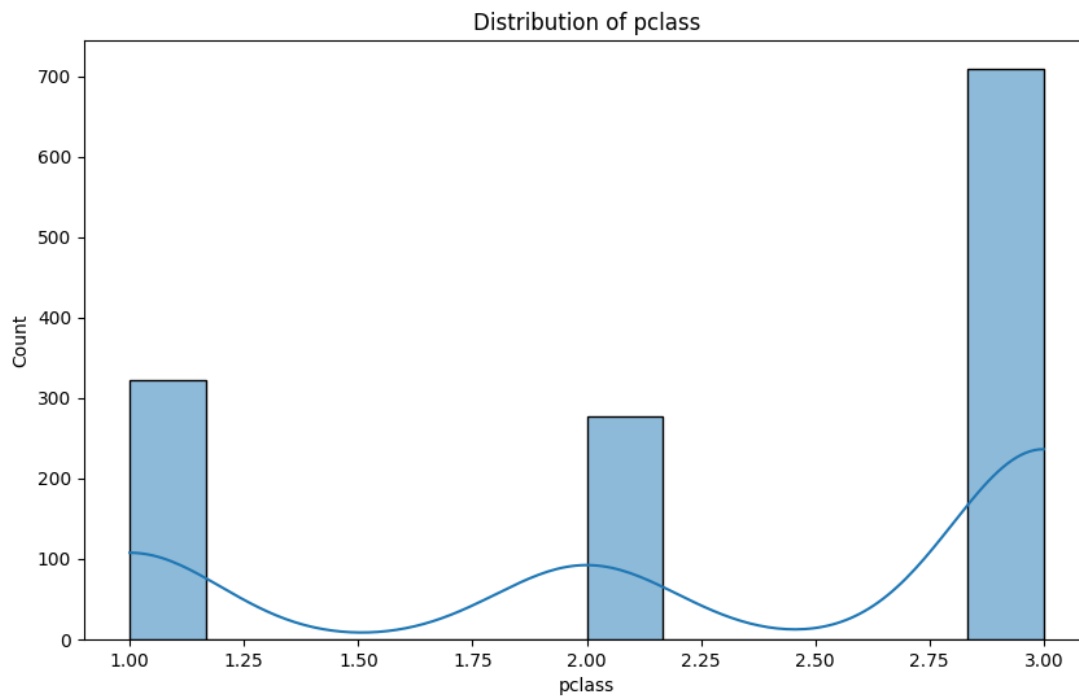


Figure 3: Distribution of pclass

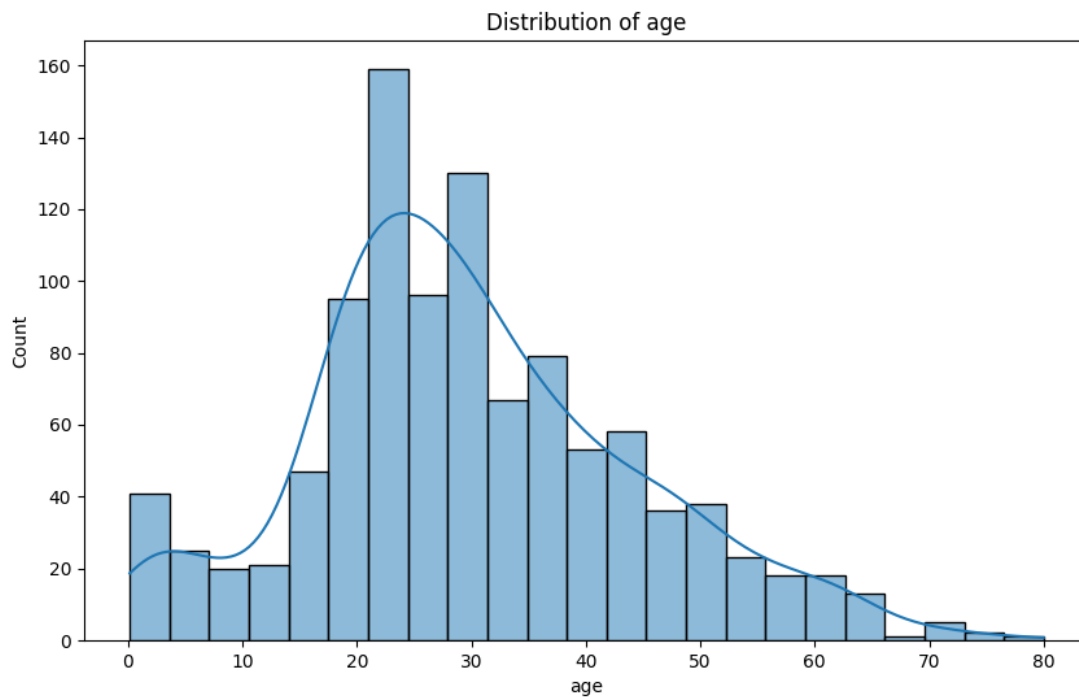


Figure 4: Distribution of age

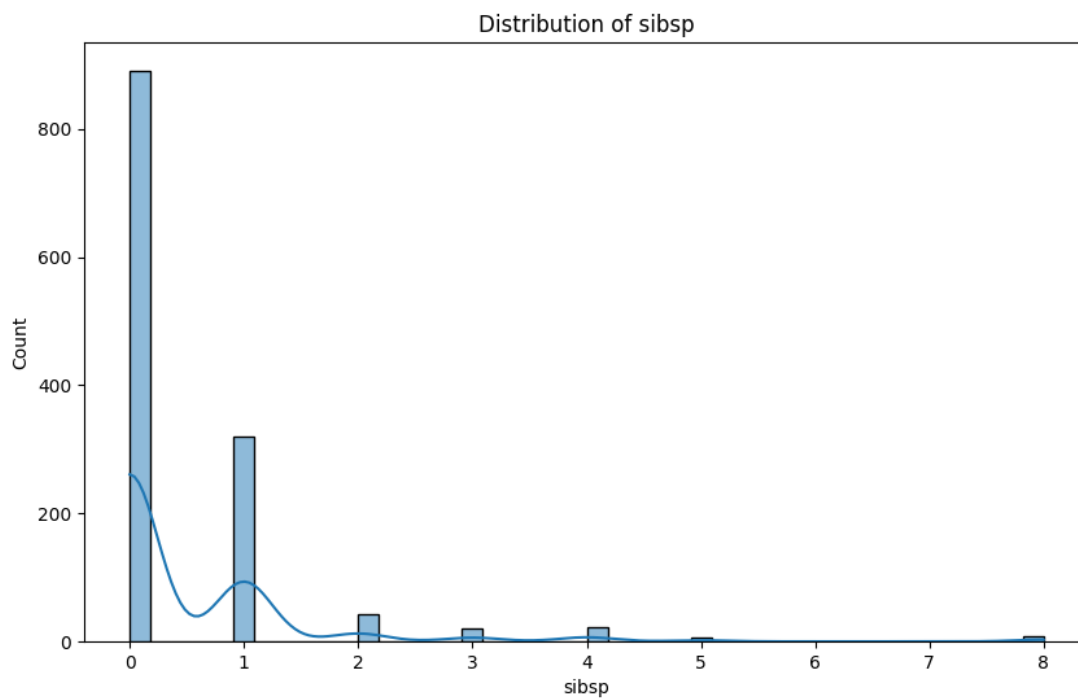


Figure 5: Distribution of sibsp

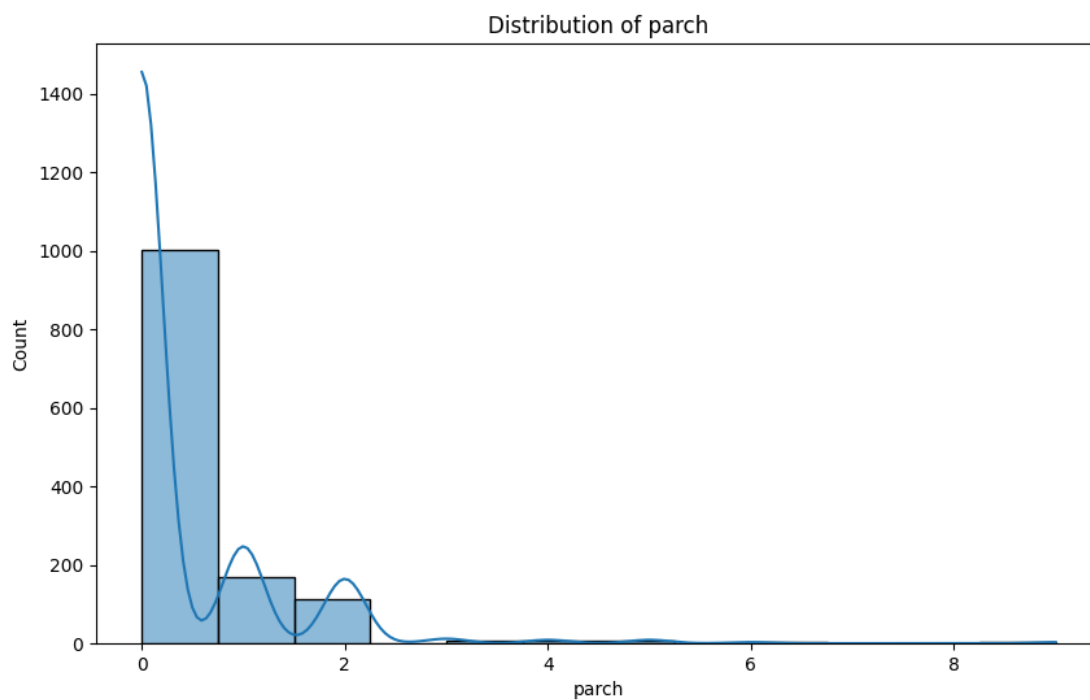


Figure 6: Distribution of parch

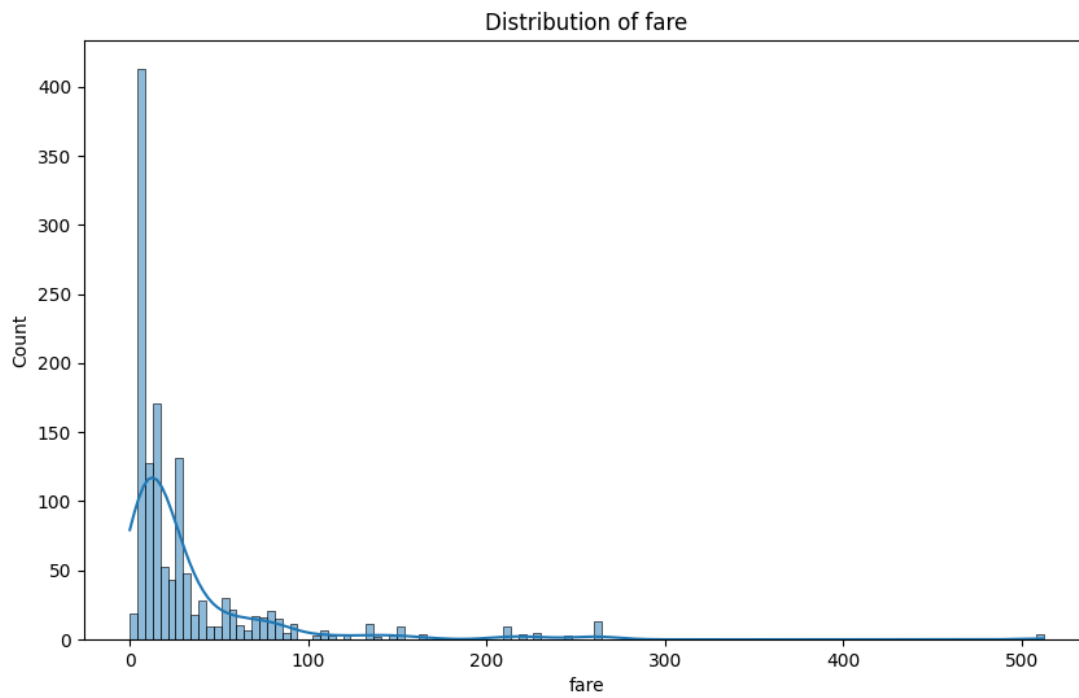


Figure 7: Distribution of fare

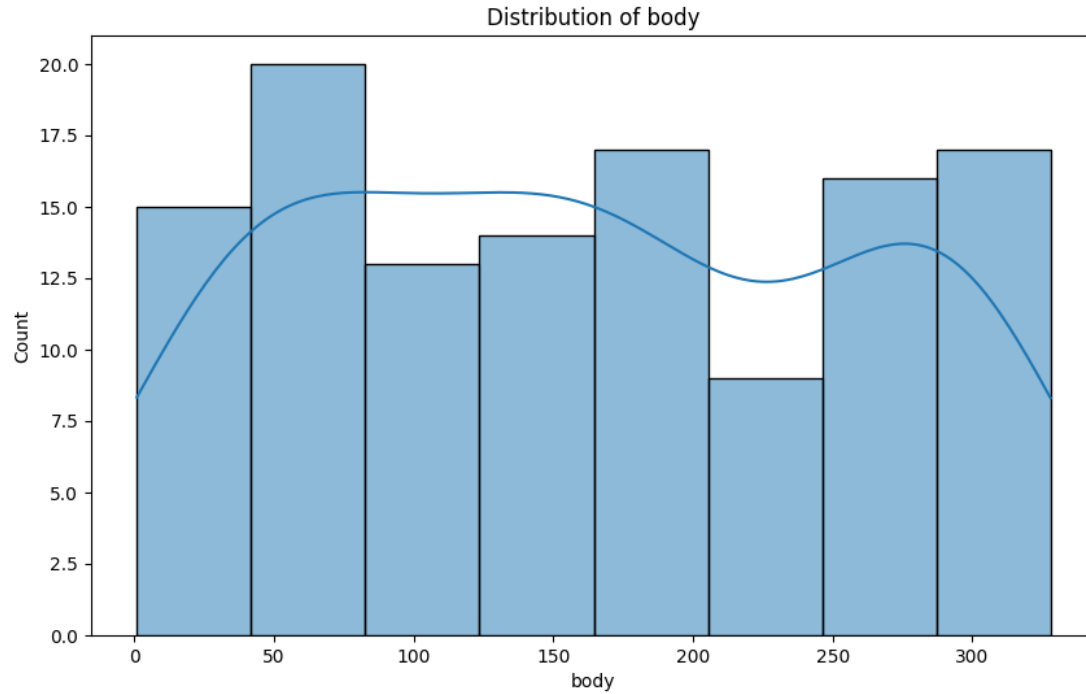


Figure 8: Distribution of body

2.2 Categorical Features

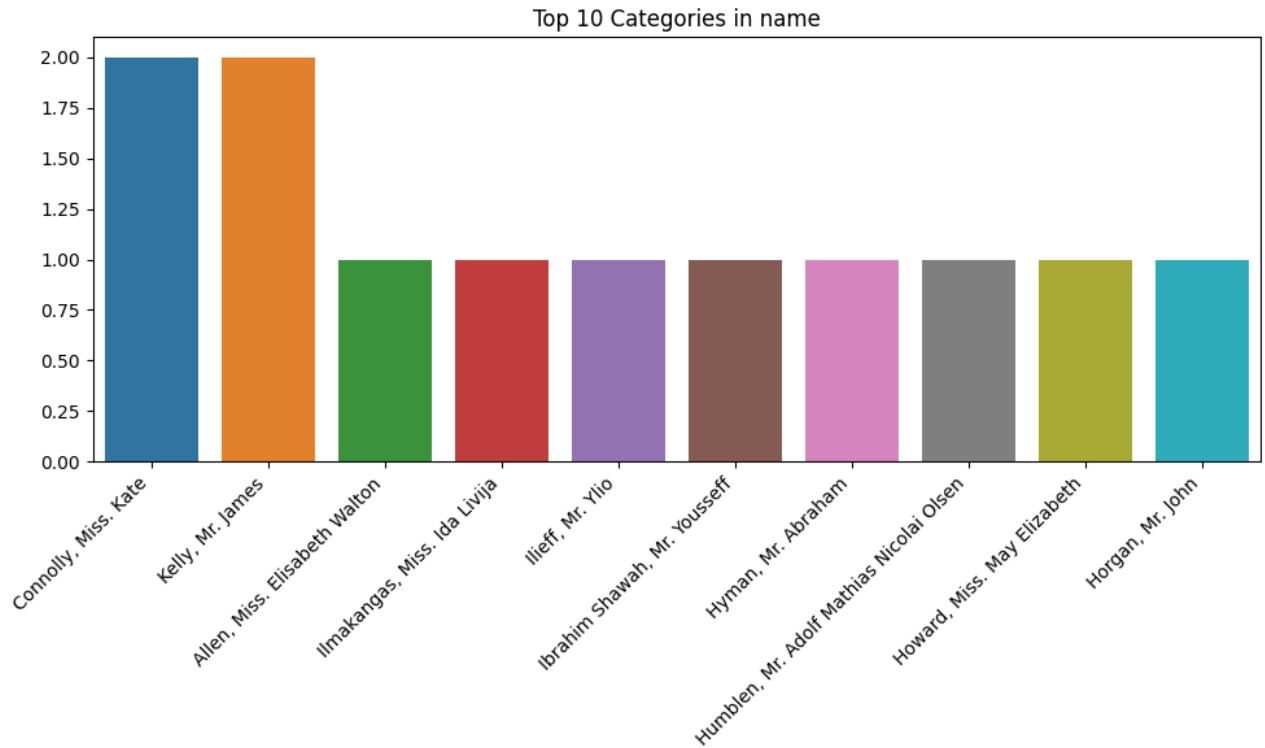


Figure 9: Distribution of name

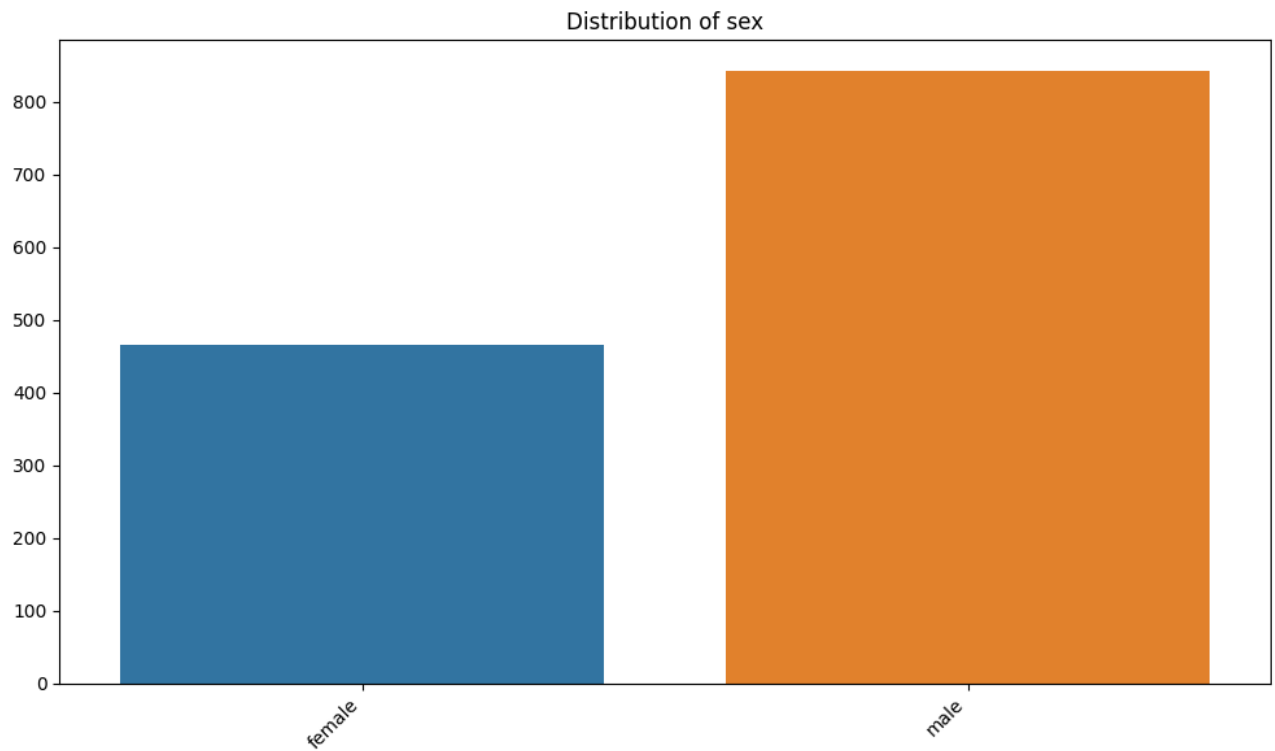


Figure 10: Distribution of sex

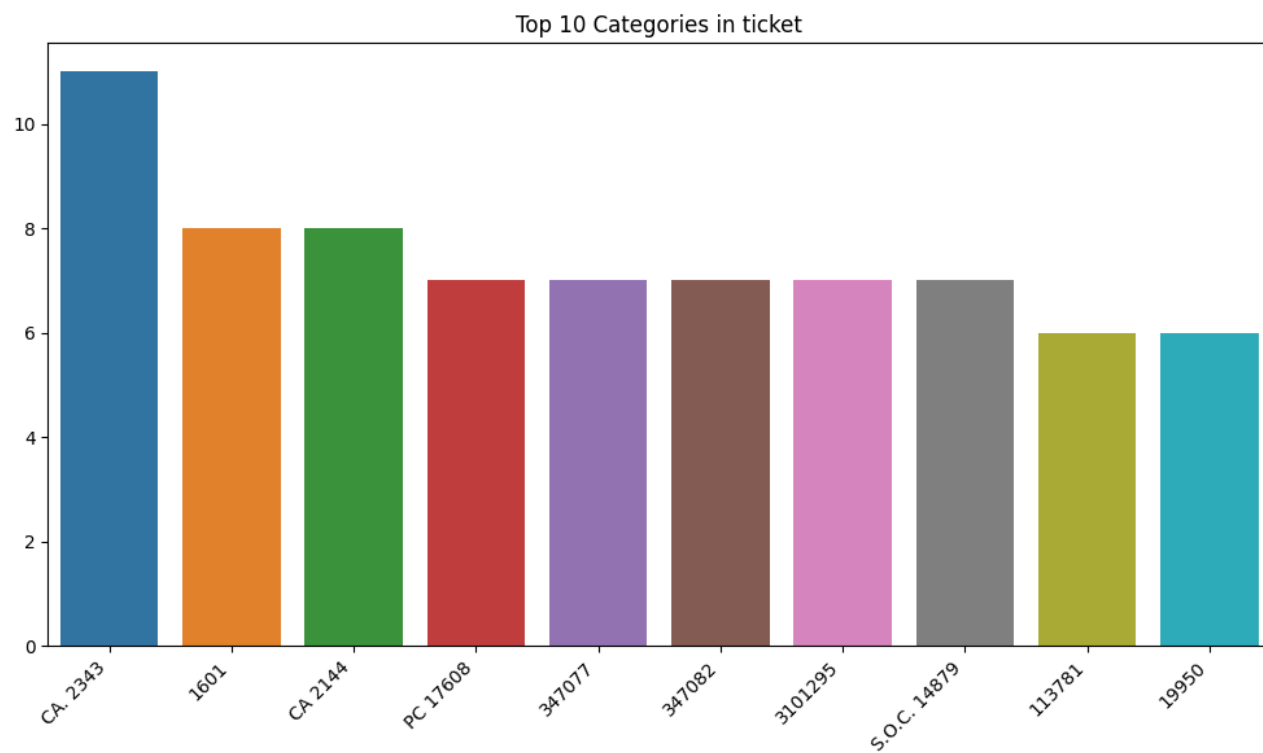


Figure 11: Distribution of ticket

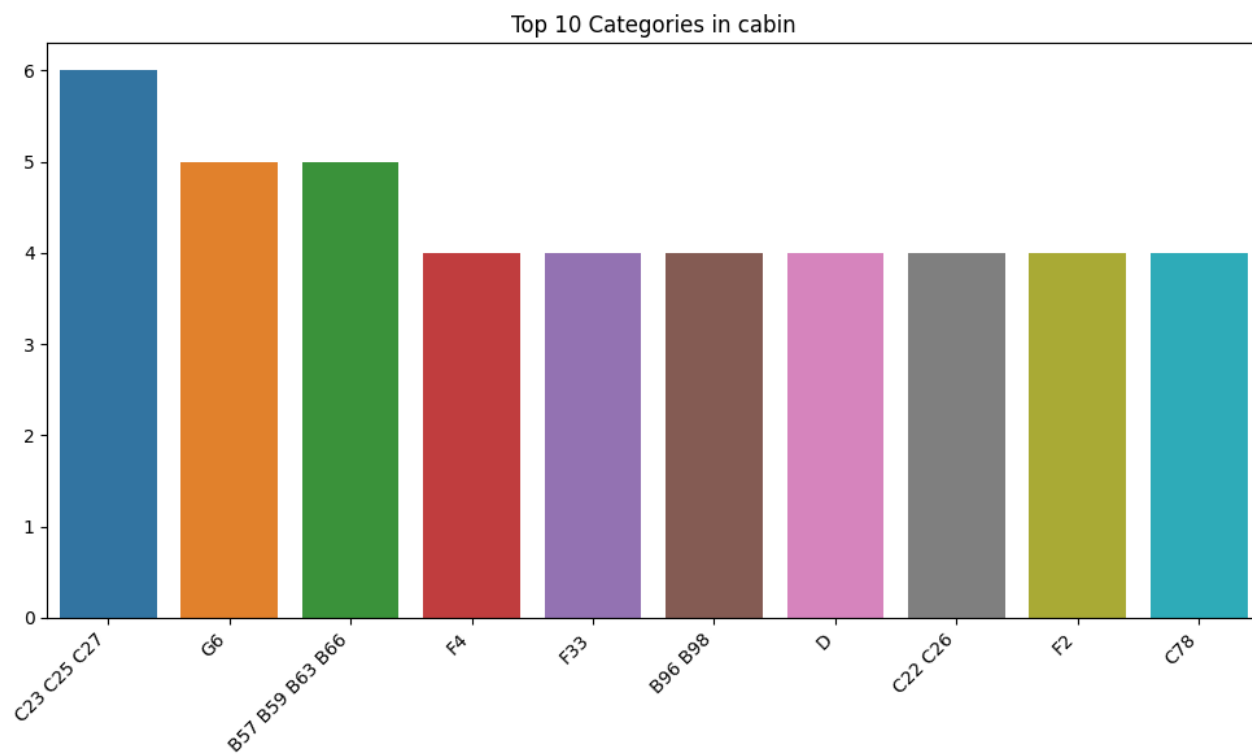


Figure 12: Distribution of cabin

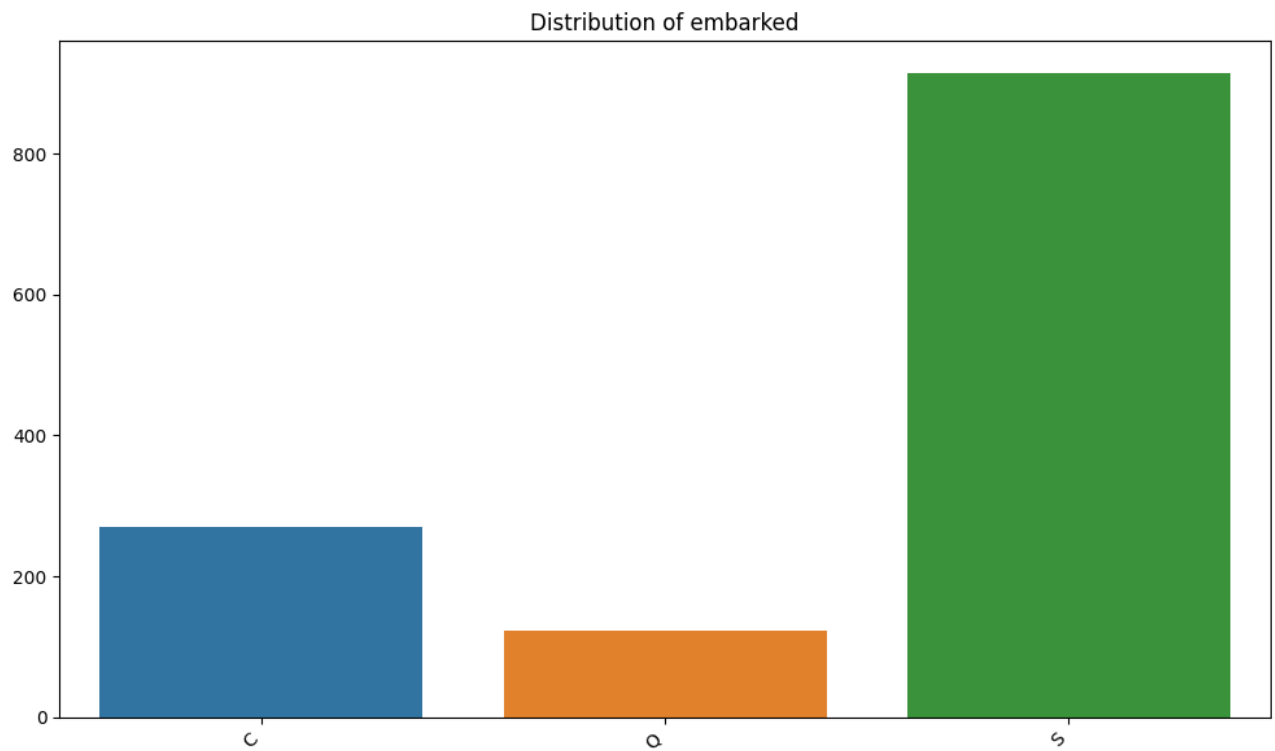


Figure 13: Distribution of embarked

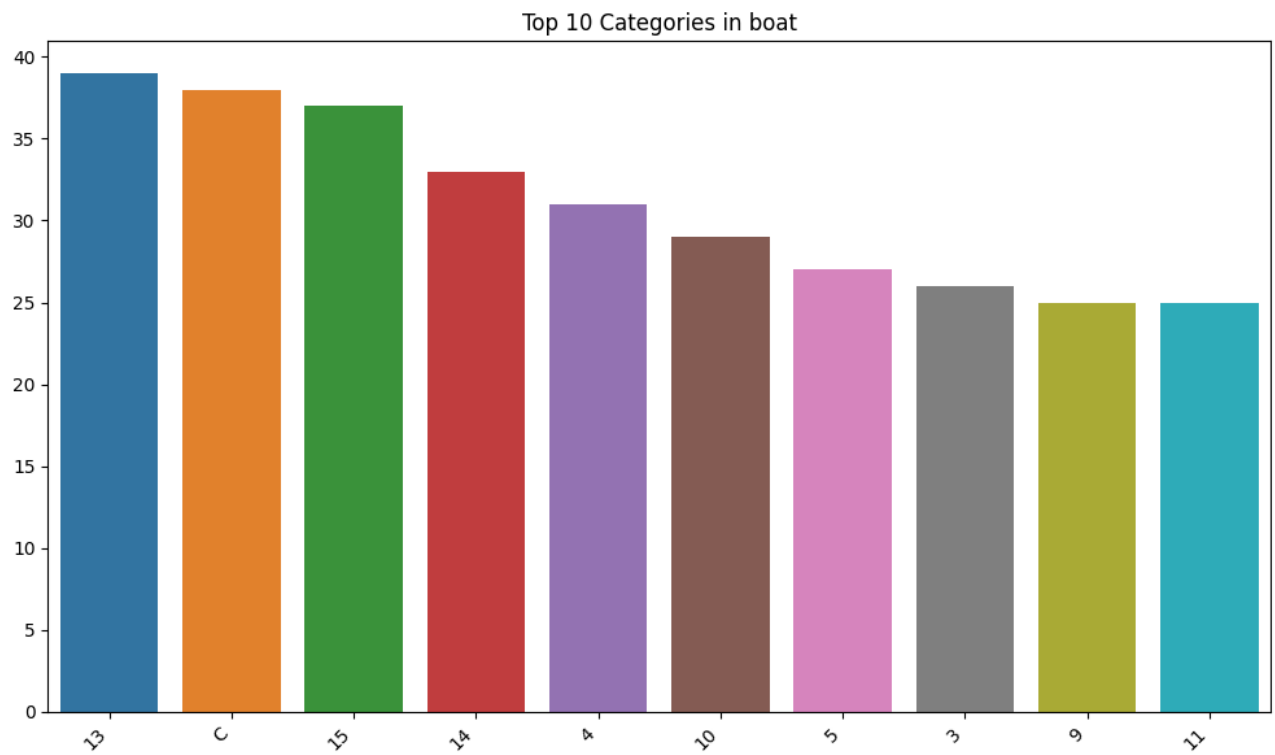


Figure 14: Distribution of boat

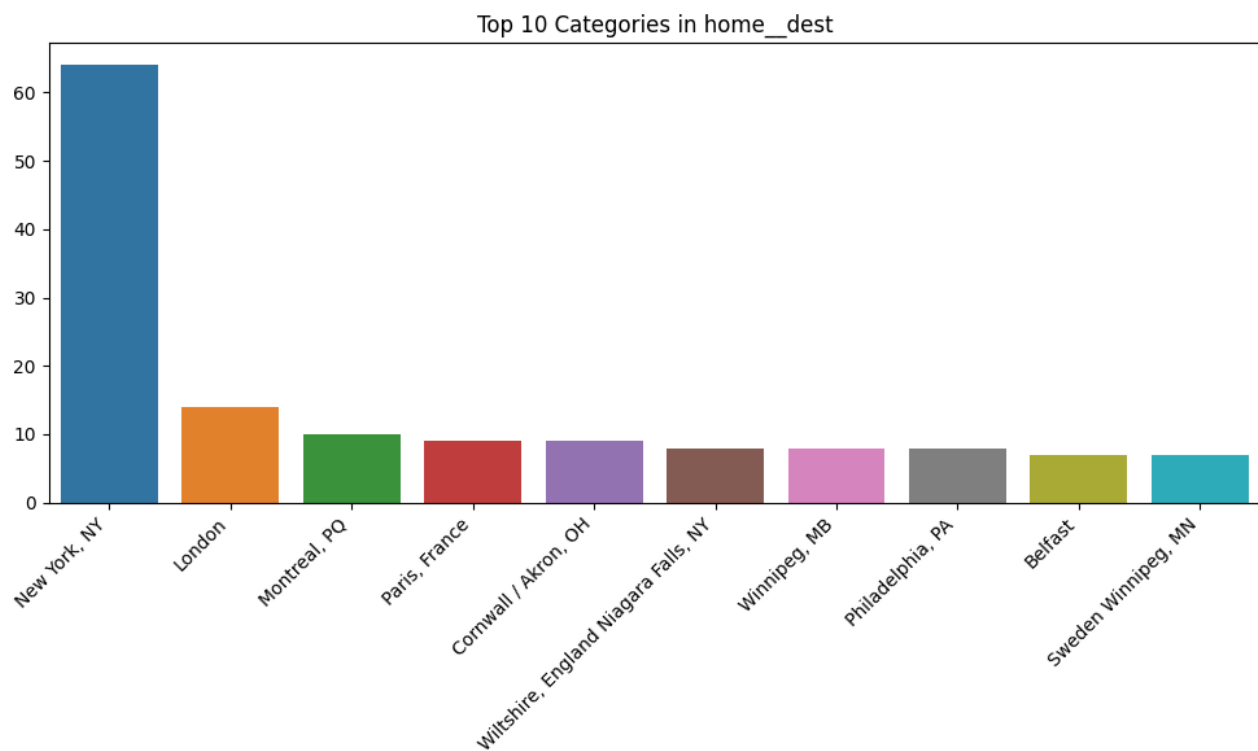


Figure 15: Distribution of home__dest

2.3 Correlation Analysis

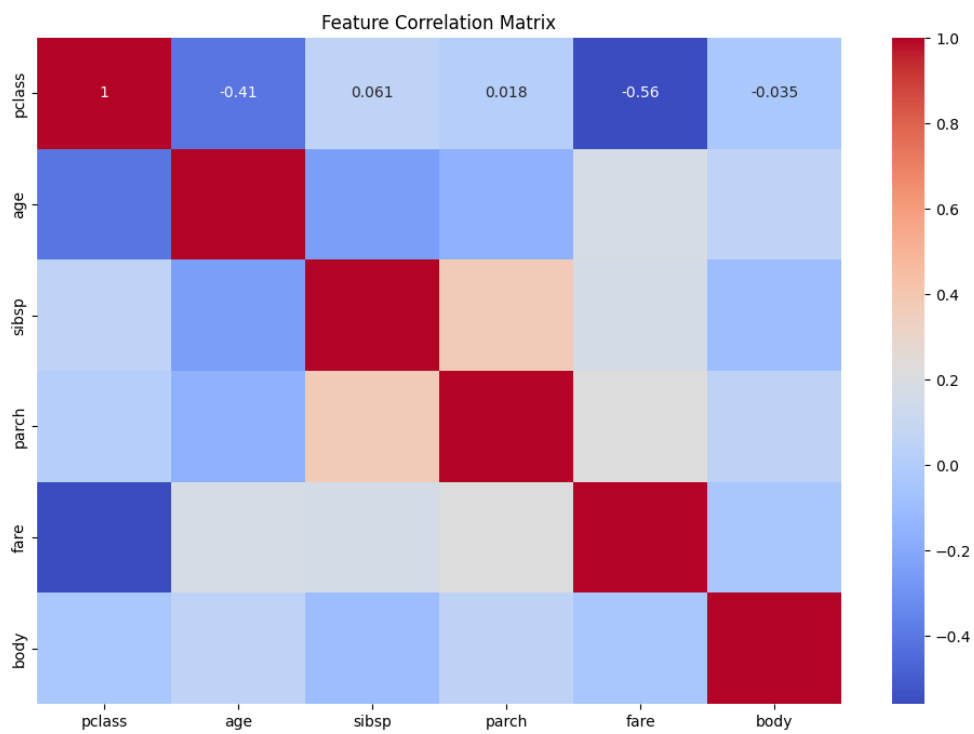


Figure 16: Feature Correlation Matrix



Figure 17: Feature Correlations with Target

3 Model Performance

3.1 Cross-validation Results

5-fold CV Score: 0.9761 (+/- 0.0121)

3.2 Test Set Performance

	precision	recall	f1-score	support
0	0.95	1.00	0.98	144
1	1.00	0.94	0.97	118
accuracy			0.97	262
macro avg	0.98	0.97	0.97	262
weighted avg	0.97	0.97	0.97	262

3.3 Confusion Matrix

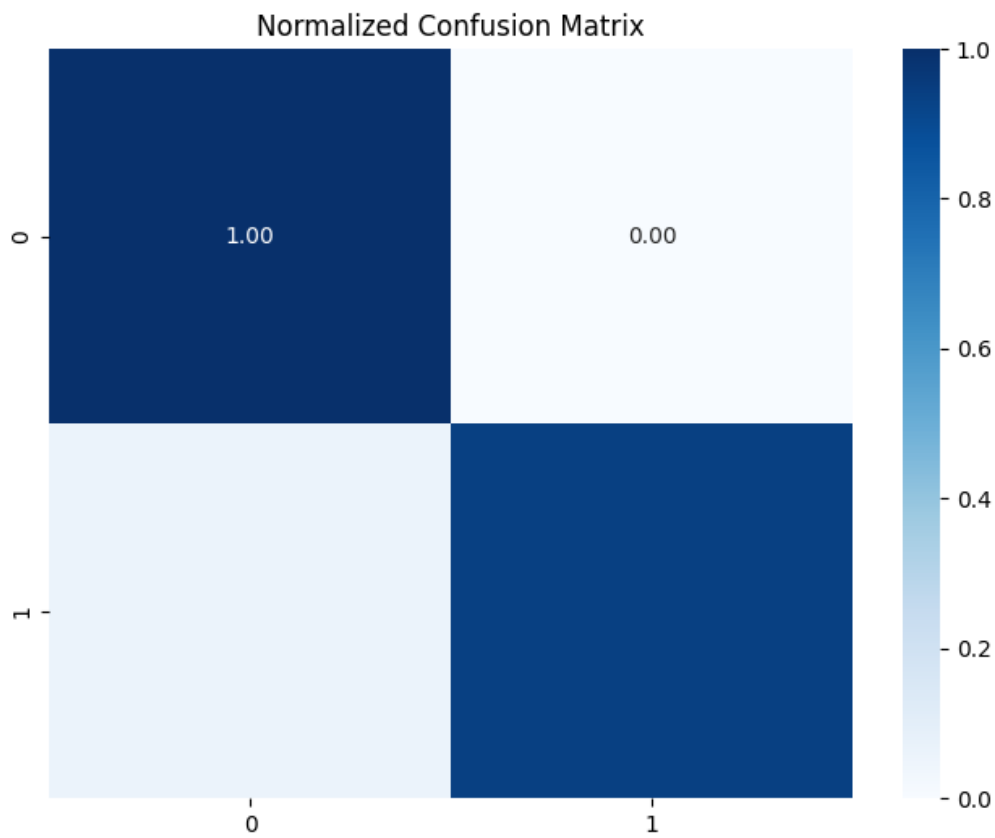


Figure 18: Normalized Confusion Matrix

4 Feature Importance Analysis

Analysis of feature importance in the model.

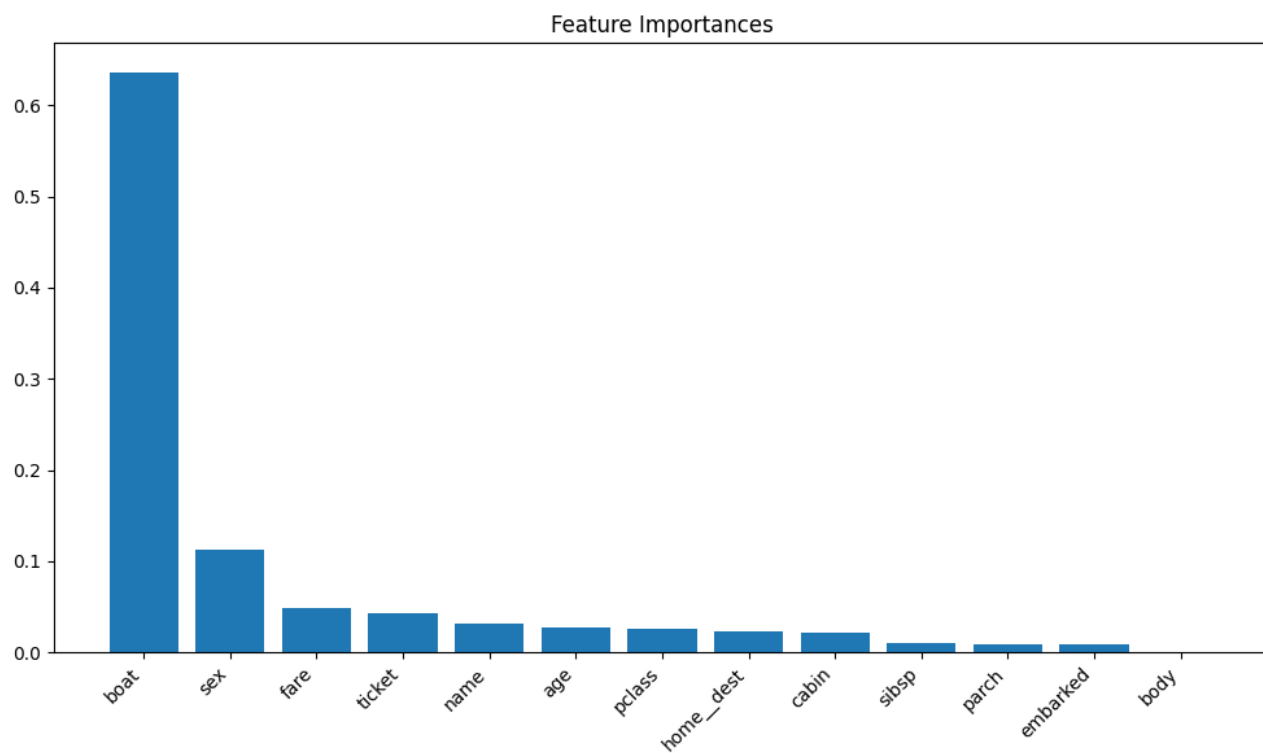


Figure 19: Feature Importance Rankings