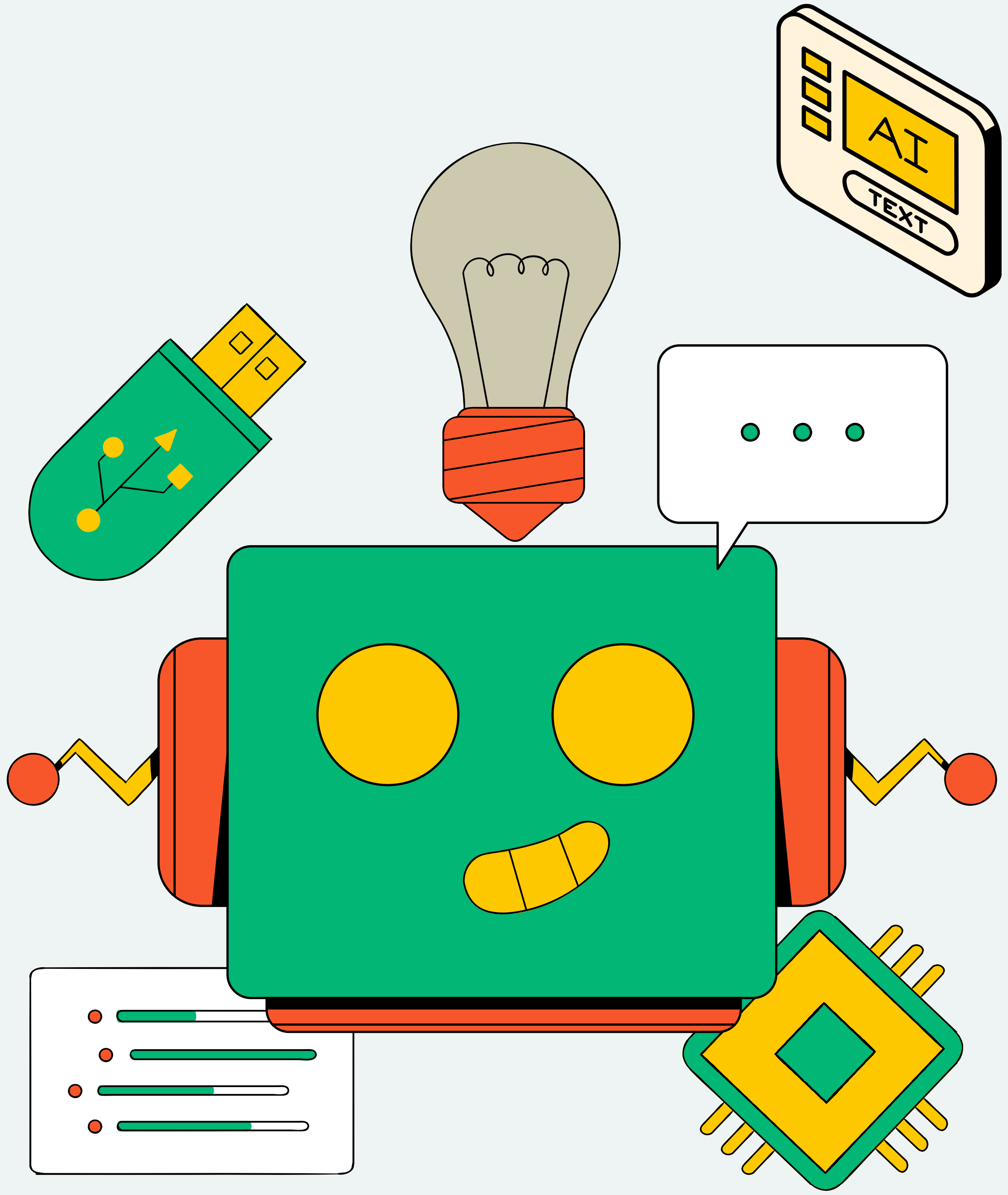
AFAME TECHNOLOGIES

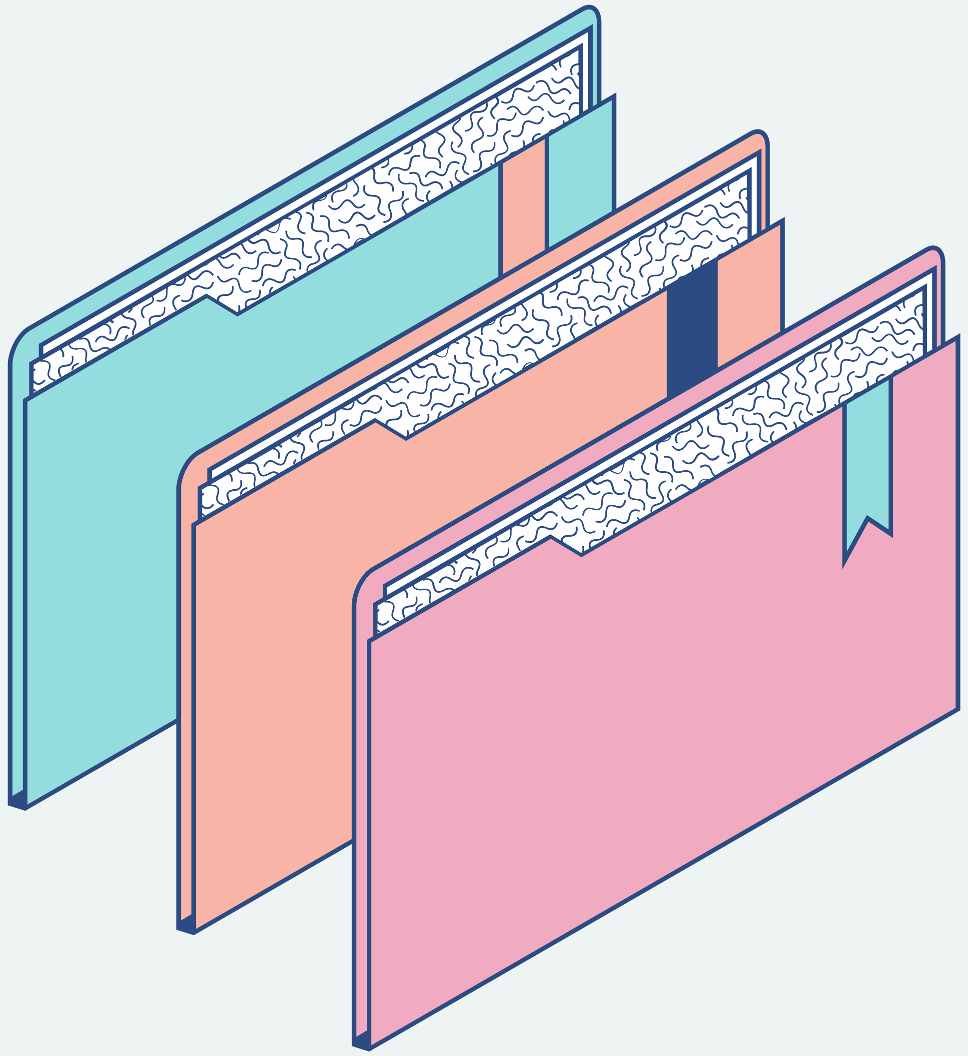


PROJECT REPORT

TITANIC SURVIVAL PREDICTION

PRESENTED BY:
AKSHIT GANDOTRA

Self Introduction



I'm Akshit, thrilled to present myself in the realm of data analysis. With a keen interest in uncovering insights and driving informed decisions through data, I eagerly embrace opportunities to contribute meaningfully to analytical endeavors. Collaboration forms the cornerstone of my approach, recognizing the collective power in diverse perspectives and teamwork.

Throughout my journey, I've remained steadfast in upholding a commitment to excellence, ensuring that every analytical endeavor reflects the utmost precision and depth. My mission is rooted in leveraging data to empower individuals and businesses, facilitating informed strategies and fostering growth. I'm genuinely excited about the prospect of connecting with fellow data enthusiasts and embarking on transformative analytical journeys together.

Introduction to Afame Technologies

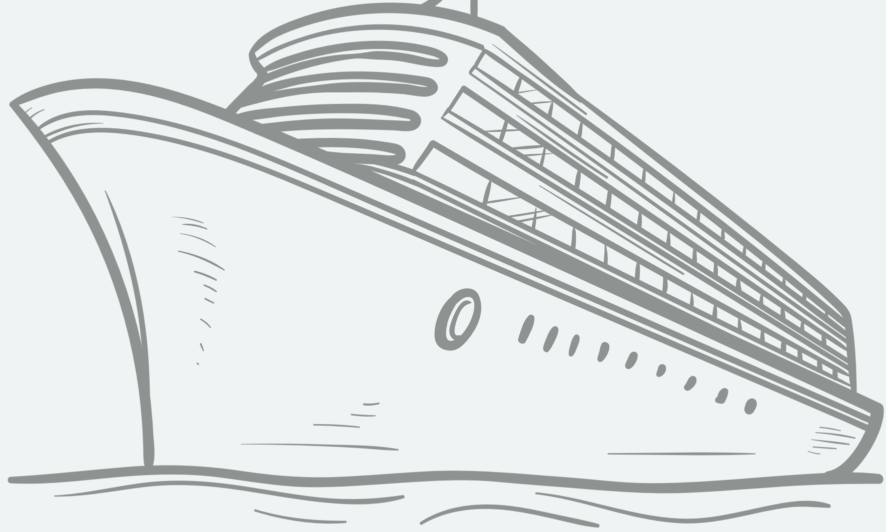
Afame Technologies stands as a beacon in the digital landscape, offering a comprehensive suite of services tailored to meet the diverse needs of businesses. Specializing in web development, digital marketing, and strategic consulting, Afame Technologies is committed to delivering cutting-edge solutions that drive success in the digital realm.

Afame Technologies envisions a future where businesses of all sizes thrive in the digital ecosystem. Their vision is to be the catalyst for this transformation, empowering our clients to reach new heights of success through technology, innovation, and strategic guidance.

Importance of Data Analysis for Business Growth:

Data analysis is vital for business growth as it provides actionable insights, allowing informed decision-making and strategic planning. By leveraging data effectively, businesses can identify opportunities, optimize operations, and gain a competitive edge in today's dynamic market landscape.

Goals and Objectives



Project Goals:

1. Predict survival likelihood for Titanic passengers using machine learning.
2. Explore Titanic dataset to understand passenger characteristics.
3. Develop a model for predicting survival based on passenger attributes.
4. Assess model accuracy and reliability using appropriate metrics.
5. Identify key factors influencing survival rates among Titanic passengers.

Project Objectives:

1. Preprocess dataset (handle missing values, encode variables, scale features).
2. Explore feature relationships with survival.
3. Engineer features for improved model performance.
4. Train machine learning algorithms and assess interpretability.

About the Data

Dataset Link:

https://drive.google.com/drive/folders/1SBScS_ixyyh4VtI3DaCbmgvCflgdyyJv

Data Summary:

Total Entries: The dataset contains information on a total of 891 passengers.

Features: There are several features available for each passenger:

- PassengerId: A unique identifier for each passenger.
- Survived: Whether the passenger survived or not (0 = No, 1 = Yes).
- Pclass: Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd).
- SibSp: Number of siblings/spouses aboard the Titanic.
- Parch: Number of parents/children aboard the Titanic.
- Passenger Name, Sex, Age, Ticket Number, Fare, Cabin Number
- Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

Target Variable: The target variable is "Survived," indicating whether each passenger survived the Titanic disaster.

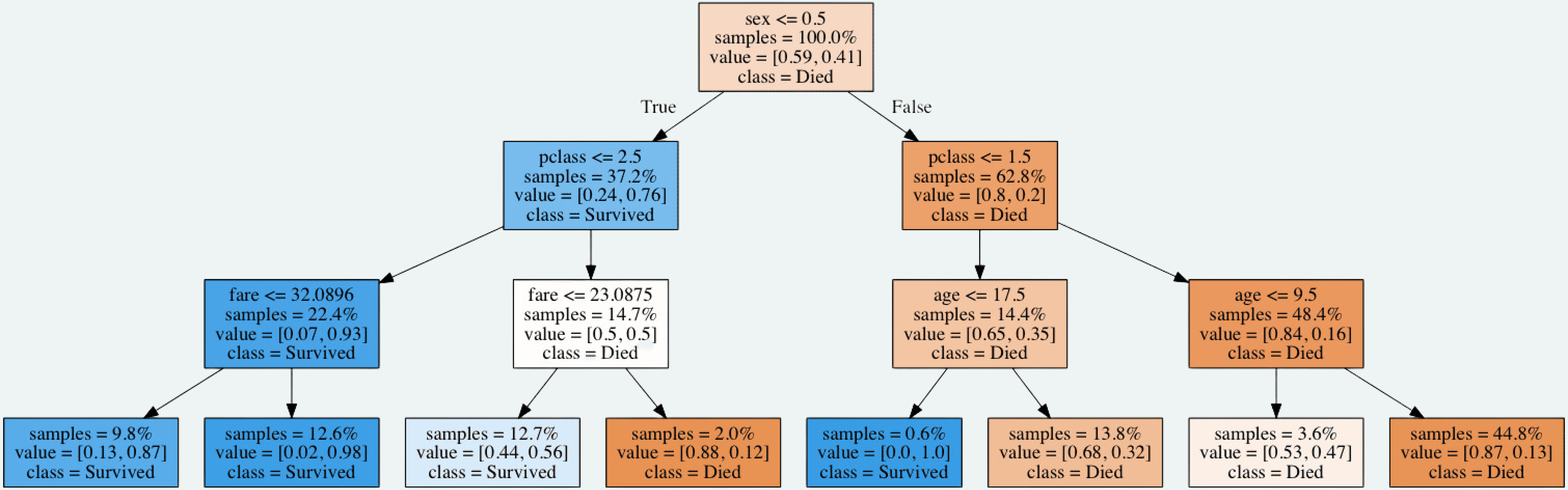
Data Description

The Titanic dataset comprises information on 891 passengers, including features like age, gender, ticket class, and whether they survived the disaster, providing a foundation for predictive modeling and analysis of survival

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Ha	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John I	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Lain	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacque	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William He	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timot	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gos	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas	female	14	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Ma	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, Miss. Elizab	female	58	0	0	113783	26.55	C103	S
13	0	3	Saundercock, Mr. Wi	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Ande	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eugen	male	2	4	1	382652	29.125		Q
18	1	2	Williams, Mr. Charles	male		0	0	244373	13		S
19	0	3	Vander Planke, Mrs.	female	31	1	0	345763	18		S
20	1	3	Masselmani, Mrs. Fa	female		0	0	2649	7.225		C
21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26		S
22	1	2	Beesley, Mr. Lawren	male	34	0	0	248698	13	D56	S
23	1	3	McGowan, Miss. Ann	female	15	0	0	330923	8.0292		Q

About the Model

The Random Forest classifier is chosen for its ability to handle complex, non-linear relationships between features and survival, its robustness to overfitting, and its capacity to provide feature importance insights, making it a suitable choice for predicting survival on the Titanic dataset.



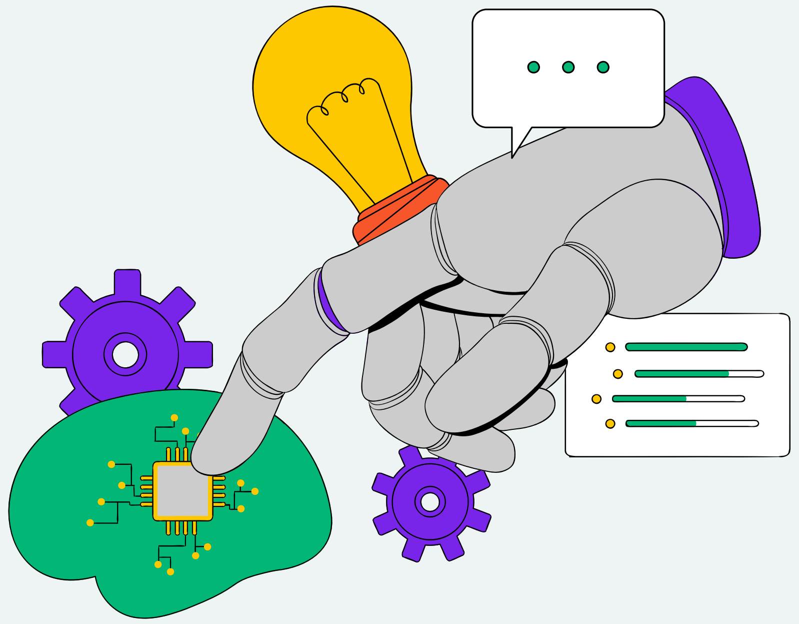
Dataset Statistics

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
Count of Null value:

PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
```

Building the Model



Data Loading: Load the Titanic dataset into the environment.

Data Preprocessing:

- Handle missing values for 'Age' and 'Embarked'.
- Drop irrelevant columns like 'Cabin', 'Ticket', 'Name', and 'PassengerId'.
- Encode categorical variables ('Sex' and 'Embarked') using one-hot encoding.

Model Selection:

- Split the dataset into features (X) and target variable (y) and further split the data into training and testing sets using train_test_split().

Model Training and Evaluation

- Train a Random Forest classifier on the training data using RandomForestClassifier().
- Evaluate the trained model's accuracy on the testing data using accuracy_score().
- Examine the confusion matrix to understand model performance.

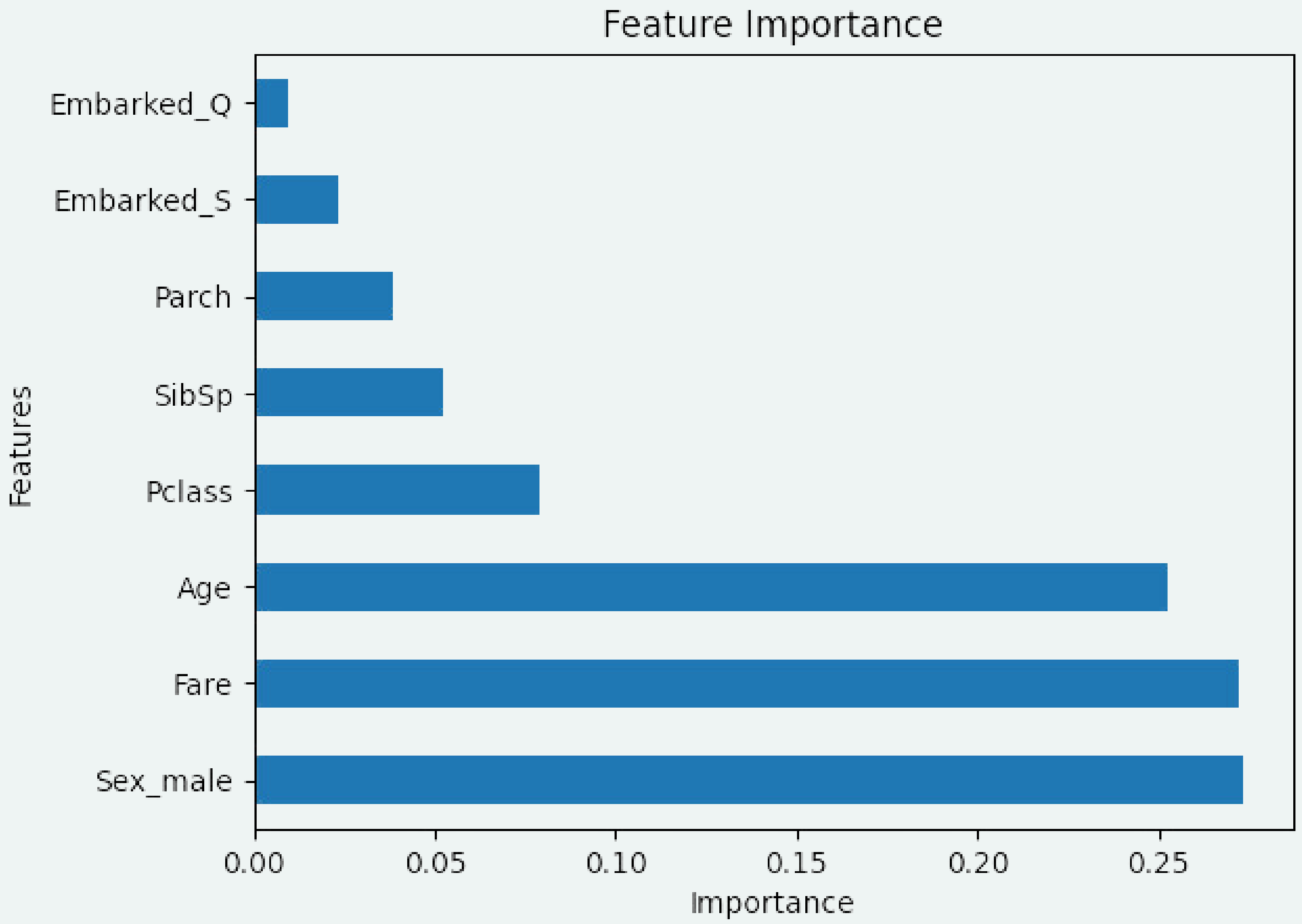
Visualizations:

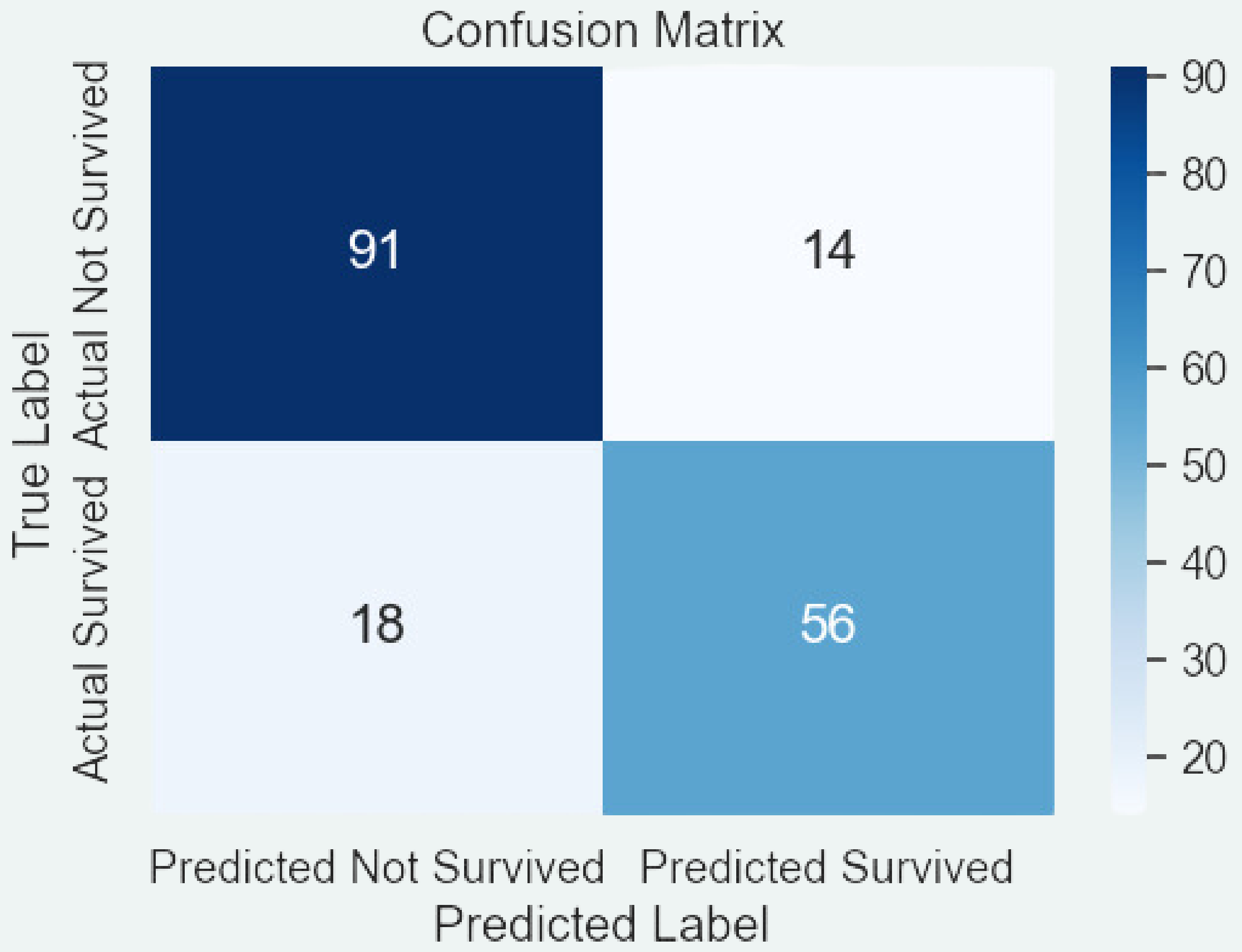
- Plot feature importance using a bar plot to understand influential factors in survival.
- Create count plots to visualize survival counts by gender and passenger class using Seaborn.

Feature Importance and Evaluation

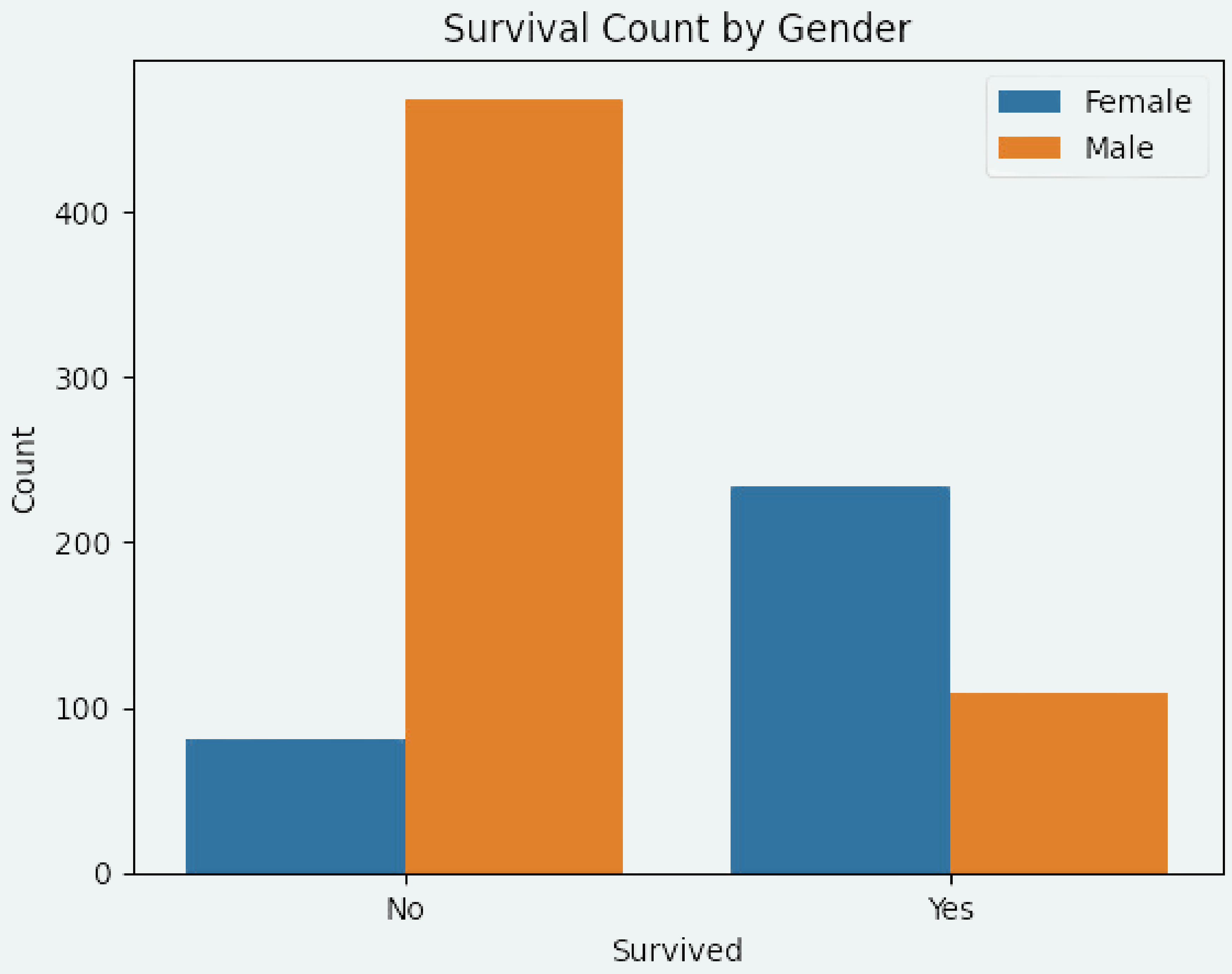
An accuracy of 82.12% suggests that the model performs reasonably well in classifying passengers' survival status.

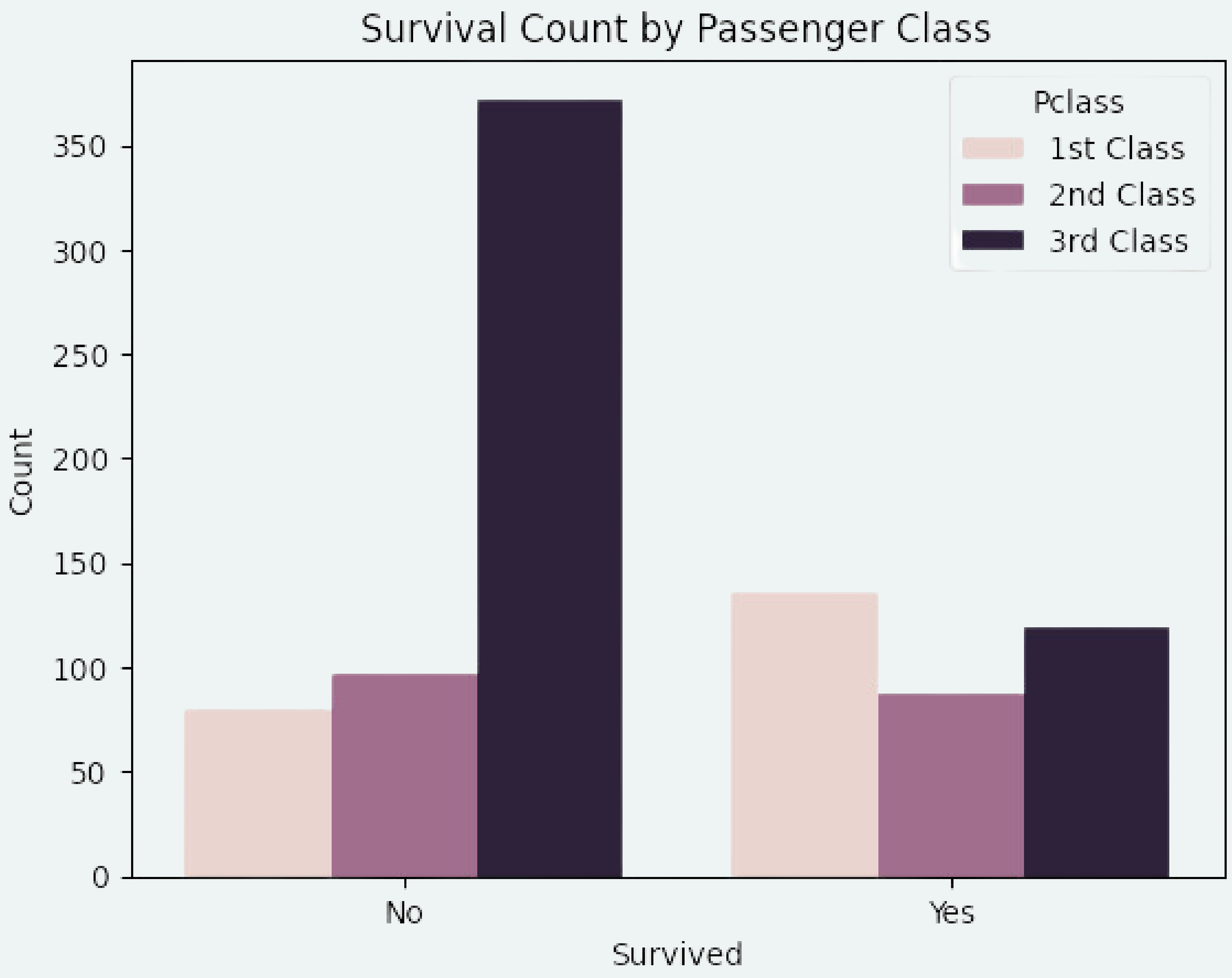
The relatively balanced distribution between true positives and true negatives indicates that the model is performing well in both predicting survival and non-survival cases. However, false positives and false negatives are still present, indicating areas where the model may misclassify passengers' survival status.





Survival Analysis





Thank you