

# Titanic Survival Prediction Project Report

## Project Overview

This project aims to predict passenger survival on the Titanic using machine learning algorithms. The analysis compares Logistic Regression and Support Vector Machine (SVM) models.

## Dataset Description

- **Source:** Titanic passenger dataset (891 records)
- **Target Variable:** Survival (0 = No, 1 = Yes)
- **Features:** 12 features including passenger class, age, gender, fare, etc.

## Data Preprocessing

### Handling Missing Values

- Age: Filled with median values based on passenger class and gender
- Embarked: Filled with mode value
- Cabin: Dropped due to excessive missing values

### Feature Engineering

- Created new features: FamilySize, IsAlone, AgeGroup
- Encoded categorical variables: Sex, Embarked, Title
- Scaled numerical features for SVM compatibility

## Exploratory Data Analysis

### Key Findings

1. **Overall Survival Rate:** 38.38% of passengers survived
2. **Gender Disparity:**
  - Female survival rate: 74.20%
  - Male survival rate: 18.89%
3. **Class Impact:**
  - 1st Class: 62.96% survival
  - 2nd Class: 47.28% survival
  - 3rd Class: 24.24% survival
4. **Age Factor:** Children had higher survival rates than adults

### Visualizations

![[EDA Visualizations]](images/eda\_summary.png)

### Key Insights from Visualizations:

- Strong correlation between survival and gender/passenger class

- Higher fare passengers had better survival chances
- Moderate family size (2-4) correlated with higher survival

## Model Performance

### Baseline Models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8156	0.7895	0.7547	0.7717
SVM	0.8212	0.8036	0.7547	0.7785

### After Hyperparameter Tuning

Model	Accuracy	Precision	Recall	F1-Score
Tuned Logistic Regression	0.8324	0.8036	0.7925	0.7980
Tuned SVM	0.8436	0.8208	0.7925	0.8065

### Performance Improvement

- **Logistic Regression:** +1.68% accuracy improvement
- **SVM:** +2.24% accuracy improvement

## Key Findings

### Most Important Features

1. **Gender:** Strongest predictor of survival
2. **Passenger Class:** Higher classes had priority
3. **Fare:** Indicator of socioeconomic status
4. **Title:** Reflects age and social status

### Model Comparison

- **SVM** performed slightly better after tuning
- **Logistic Regression** offers better interpretability
- Both models achieved >80% accuracy

## Conclusion

The analysis successfully predicted Titanic survival with over 84% accuracy using tuned SVM.

The project demonstrates:

- Effective data preprocessing and feature engineering
- Comprehensive model evaluation and optimization
- Clear insights into factors affecting survival

## Recommendations

1. Use SVM for highest accuracy requirements

2. Use Logistic Regression when interpretability is important
3. Consider ensemble methods for potential further improvement

---

**Project Code:** [titanicsurvivalprediction.ipynb](titanicsurvivalprediction.ipynb)

**Web App:** [app.py](app.py)