Titanic Survival Prediction

February 05, 2026

--- title: "Titanic Survival Prediction" subtitle: "Predict passenger survival on the Titanic using demographic and ticket information." author: "Jotty SwarmMLComprehensive" date: "February 05, 2026" geometry: "margin=0.9in" fontsize: 11pt documentclass: article classoption: twoside colorlinks: true linkcolor: NavyBlue urlcolor: NavyBlue toccolor: NavyBlue toc-depth: 3 numbersections: true header-includes: Typography - \usepackage{fontspec} - \setmainfont{DejaVu Serif}Scale=1.0 - \setsa

Tables - \usepackage{booktabs} - \usepackage{longtable} - \usepackage{array} - \usepackage{multirow} - \usepackage{float} - \usepackage{tabularx} - \usepackage{colortbl} - \renewcommand{\arraystretch}{1.3}

Graphics - \usepackage{graphicx} - \usepackage{adjustbox}

Colors - \usepackage{xcolor} - \definecolor{NavyBlue}{RGB}{26,54,93} - \definecolor{TableHeader}{RGB}{44,82,130} - \definecolor{TableAlt}{RGB}{240,245,250} - \definecolor{AccentBlue}{RGB}{49,130,206} - \definecolor{SuccessGreen}{RGB}{56,161,105} - \definecolor{WarningGold}{RGB}{214,158,46} - \definecolor{DangerRed}{RGB}{229,62,62}

Header/Footer - \usepackage{fancyhdr} - \pagestyle{fancy} - \fancyhf{} - \fancyheadLE,RO{\small\thepage} - \fancyheadLO{\small\textit{Titanic Survival Prediction}} - \fancyheadRE{\small\textit{Jotty SwarmMLComprehensive}} - \fancyfootC{\small\textcolor{gray}{Jotty ML Comprehensive Report}} - \renewcommand{\headrulewidth}{0.4pt} - \renewcommand{\footrulewidth}{0.2pt}

Title page styling - \usepackage{titling} - \pretitle{\begin{center}\LARGE\bfseries\color{NavyBlue}} - \posttitle{\par\end{center}\vskip 0.5em} - \preauthor{\begin{center}\large} - \postauthor{\par\end{center}} - \predate{\begin{center}\large} - \postdate{\par\end{center}}

Section styling - \usepackage{titlesec} - \titleformat{\section}{\Large\bfseries\color{NavyBlue}}{\thesection}{1em}{} - \titleformat{\subsection}{\large\bfseries\color{TableHeader}}{\thesubsection}{1em}{} - \titleformat{\subsubsection}{\normalsize\bfseries}{\thesubsubsection}{1em}{}

Captions - \usepackagefont=small,labelfont=bf,textfont=it{caption}

Hyperref (load last) - \usepackage{hyperref} - \hypersetup{pdfauthor={Jotty SwarmMLComprehensive}, pdftitle={Titanic Survival Prediction}, pdfsubject={Machine Learning Analysis}}

\thispagestyle{empty} \begin{center} \vspace{2cm} {\Huge\bfseries\color{NavyBlue} Titanic Survival Prediction} \vspace{0.5cm}

{\Large\textit{Predict passenger survival on the Titanic using demographic and ticket information.}} \vspace{2cm}

{\large Jotty SwarmMLComprehensive} \vspace{0.3cm}

{\large February 05, 2026} \vspace{3cm}

\includegraphicswidth=0.3\textwidth{professionalreports/figures/featureimportance.png} \vfill

{\small\textit{Generated by Jotty SwarmMLComprehensive}} \vspace{0.5cm}

{\small\textcolor{gray}{Comprehensive ML Analysis Report}} \end{center} \newpage

\tableofcontents \newpage

Executive Summary

Predict passenger survival on the Titanic using demographic and ticket information.

Key Results

Best Model: Logistic Regression

Performance Metrics:

| Metric | Value | |--------|-------| | Accuracy | 0.8244 | | Precision | 0.7872 | | Recall | 0.7400 | | F1 | 0.7629 | | Auc Roc | 0.8708 |

Dataset: 11 features analyzed

Data Quality Analysis

A comprehensive analysis of data quality, identifying potential issues before modeling.

Dataset Overview

| Metric | Value | |--------|-------| | Total Samples | 262 | | Total Features | 11 | | Numeric Features | 11 | | Categorical Features | 0 | | Features with Missing | 0 | | Total Missing Values | 0 (0.00%) |

Distribution Analysis

| Feature | Skewness | Kurtosis | Assessment | |---------|----------|----------|------------| | pclass | -0.43 | -1.50 | Symmetric | | sex | -0.59 | -1.65 | Left-skewed | | age | 0.44 | 0.68 | Symmetric | | sibsp | 4.39 | 24.11 | Right-skewed, Heavy-tailed | | parch | 3.51 | 15.81 | Right-skewed, Heavy-tailed | | fare | 4.57 | 27.27 | Right-skewed, Heavy-tailed | | embarked | -1.14 | -0.56 | Left-skewed | | familysize | 3.21 | 12.68 | Right-skewed, Heavy-tailed | | isalone | -0.25 | -1.94 | Symm

Feature Distributions

!Feature Distributions(professionalreports/figures/distributions.png)

Outlier Analysis

Method: Interquartile Range (IQR) with 1.5x multiplier

Total Outliers Detected: 174 across 7 features

| Feature | Outliers | % of Data | Min | Max | |---------|----------|-----------|-----|-----| | parch | 62 | 23.7% | -0.45 | 6.51 | | age | 29 | 11.1% | -2.26 | 3.24 | | fare | 29 | 11.1% | -0.66 | 9.83 | | familysize | 22 | 8.4% | -0.56 | 5.85 | | fareperperson | 22 | 8.4% | -0.60 | 15.20 | | sibsp | 9 | 3.4% | -0.48 | 7.44 | | ageclass | 1 | 0.4% | -2.01 | 2.83 |

Outlier Distribution

!Outlier Boxplot(professionalreports/figures/outlierboxplot.png)

Correlation & Multicollinearity Analysis

Understanding feature relationships is critical for model interpretation and feature selection.

Correlation Matrix

!Correlation Matrix(professionalreports/figures/correlationmatrix.png)

Highly Correlated Feature Pairs (|r| >= 0.7)

| Feature 1 | Feature 2 | Correlation | |-----------|-----------|-------------| | fare | fareperperson | 0.876 | | sibsp | familysize | 0.870 | | parch | familysize | 0.759 |

Variance Inflation Factor (VIF)

VIF measures multicollinearity. VIF > 5 indicates moderate, VIF > 10 indicates severe multicollinearity.

| Feature | VIF | Assessment | |---------|-----|------------| | ageclass | 8.69 | High | | pclass | 8.50 | High | | fare | 7.59 | High | | age | 7.36 | High | | fareperperson | 6.77 | High | | isalone | 1.92 | OK | | sex | 1.14 | OK | | embarked | 1.09 | OK |