

Titanic Dataset Cleaning and Preprocessing

![Titanic](https://upload.wikimedia.org/wikipedia/commons/thumb/f/fd/RMS_Titanic_3.jpg/1200px-RMS_Titanic_3.jpg)

Project Overview

Complete data cleaning pipeline for the Titanic dataset, preparing it for machine learning. Handles missing values, categorical encoding, feature scaling, and outlier detection.

Technical Implementation

Data Cleaning Steps

1. ****Missing Values****:

- Age: Filled with median (28.0)
- Embarked: Filled with mode ('S')
- Cabin: Dropped (77% missing)

2. ****Feature Engineering****:

```
```python
```

```
Convert Sex to numerical
```

```
df['Sex'] = df['Sex'].map({'male': 1, 'female': 0})
```

```
One-hot encode Embarked
```

```
df = pd.get_dummies(df, columns=['Embarked'])
```