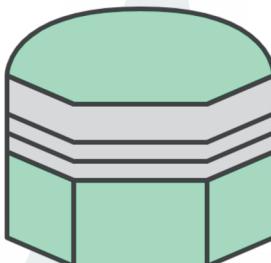




Titanic Dataset Overview

Training Set

Contains labeled data for model training



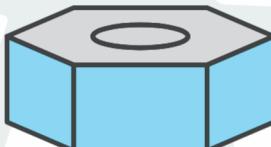
Gender Submission File

Sample submission file with gender-based assumptions



Test Set

Similar attributes, no survival outcomes





Titanic Machine Learning Challenge: Overview and Steps

Problem-Solving Steps ⚠

- Data Exploration
- Feature Engineering
- Model Selection
- Model Evaluation
- Submission

Titanic Machine Learning Challenge

Dataset

- Training Set
- Test Set
- Gender Submission File



Expected Challenges

- Handling Missing Data
- Feature Selection
- Avoiding Overfitting



Data Cleaning Process for Titanic Dataset



Load Dataset

Begin by loading the Titanic dataset for analysis



Inspect for Missing Values

Check the dataset for any missing or inconsistent data



Handle Missing Age Data

Address missing age data using median or predictive methods



Address Cabin Data

Manage cabin data by extracting the first letter or marking as unknown

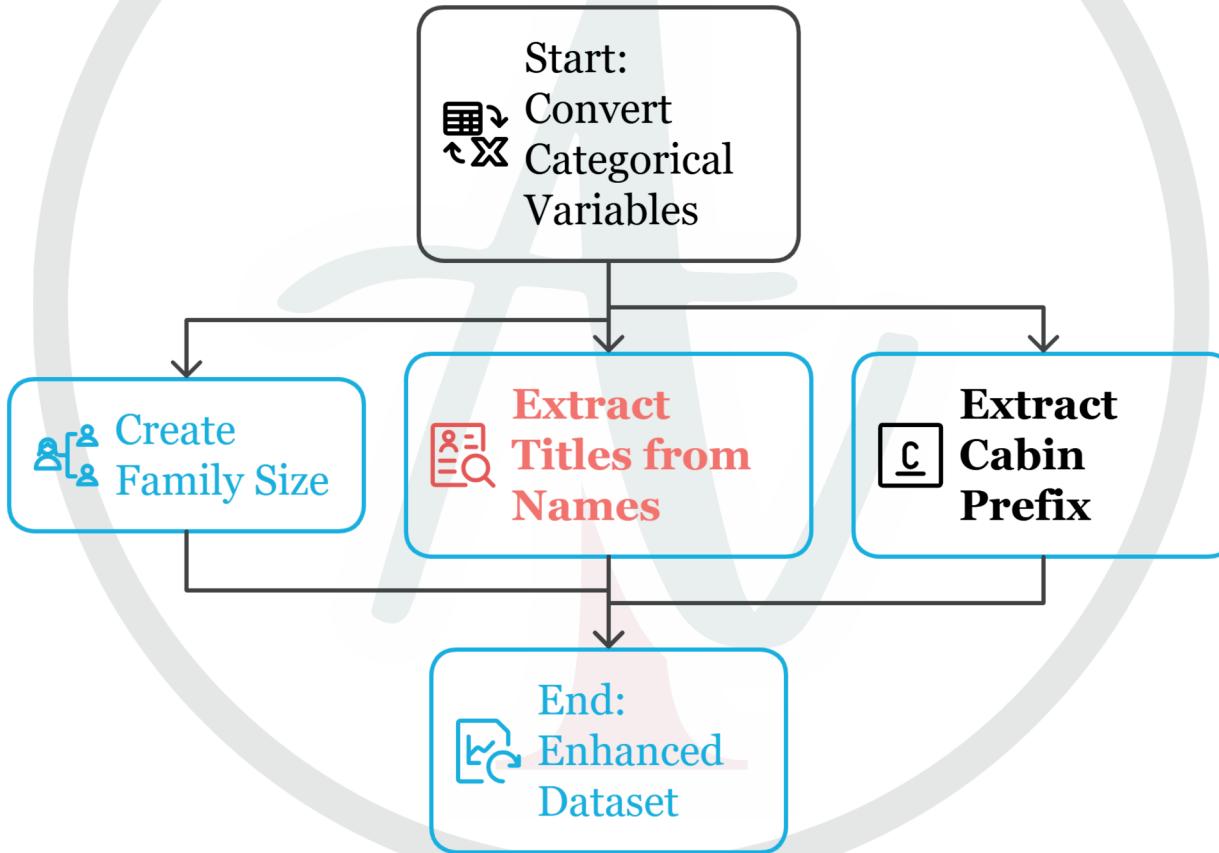


Fill Embarked Values

Complete missing embarked values with the most frequent category

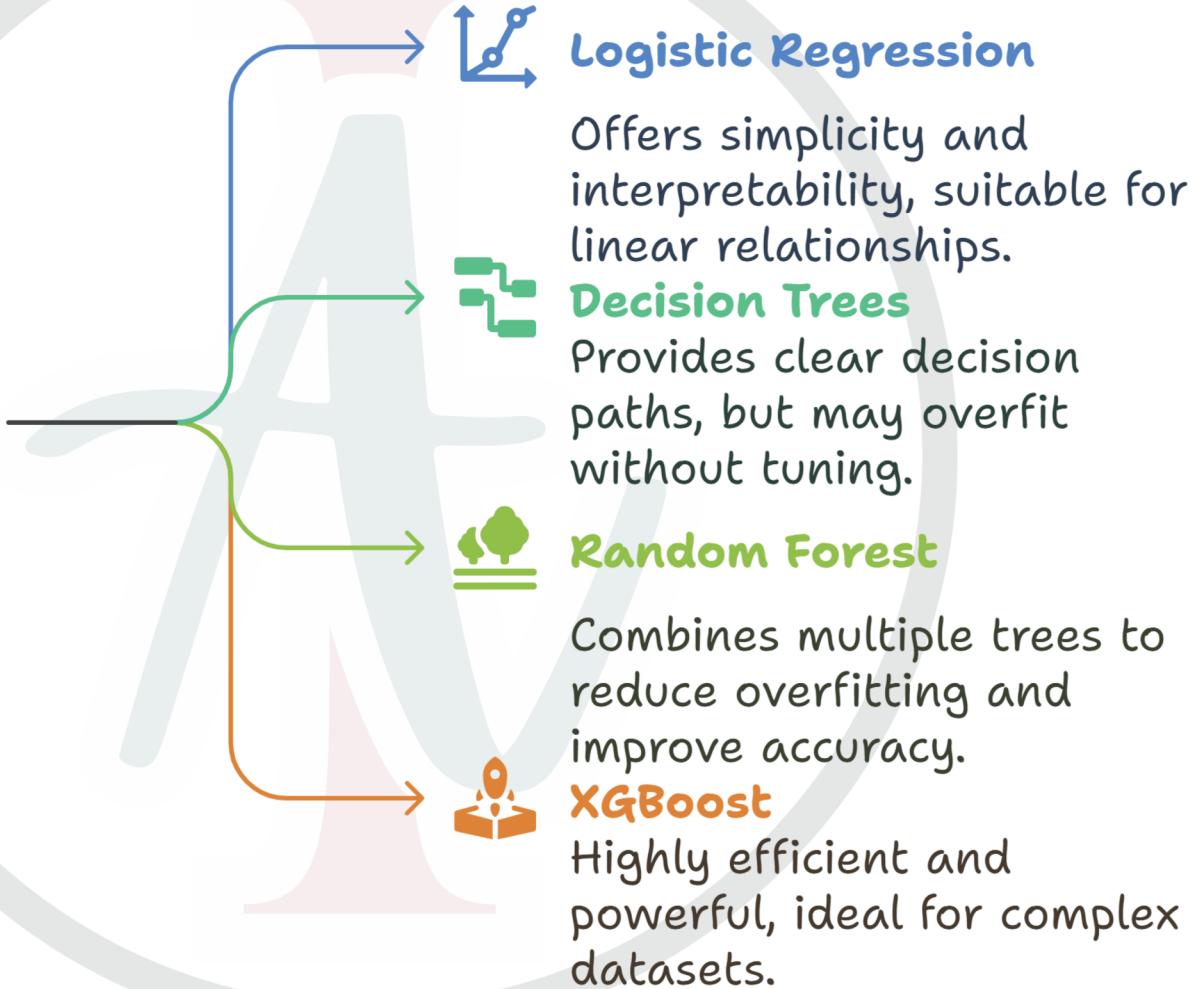


Feature Engineering Process for Titanic Dataset





Which model and parameters should be used for optimal performance?





How to evaluate and select the best model?

Use Accuracy

Accuracy provides a general measure of how often the model is correct.



Use Precision

Precision measures the correctness of positive predictions, reducing false positives.

Use Recall

Recall captures the model's ability to identify all relevant instances, minimizing false negatives.



Use F1-Score

F1-Score balances precision and recall, providing a single metric for model performance.