TITANIC SURVIVAL PREDICTION USING ML

Sharath Pai

Report on Titanic ML from Disaster Dataset Analysis

1. **Introduction:**

The Titanic ML dataset, extracted from the Kaggle competition "Titanic: Machine Learning from Disaster", contains information about passengers aboard the Titanic, including their demographics, ticket class, and survival outcomes. This analysis aims to predict passenger survival based on available attributes using various machine learning techniques.

2. **Approach:**

- Data Preprocessing: Missing values were imputed, categorical variables were encoded, and outliers were addressed.
- Exploratory Data Analysis (EDA): Visualizations were used to understand the distribution of variables, relationships between variables, and impact on passenger survival.
- Model Selection: Various machine learning algorithms, such as Logistic Regression, Random Forest, and Gradient Boosting, were evaluated using cross-validation to identify the best-performing model.
- Model Evaluation: The chosen model was assessed on a separate test set to evaluate its performance andgeneralizability.

3. **Findings:**

- EDA: Name, ticket, cabin contained unique qualitative data so we removed those variables from the dataset.
- Model Training: Model training: We took a split ratio of 70:30 for train and test data respectively and used classification algorithms to find out the accuracy of the best model.

- Model Selection: Linear Kernel emerged as the best model on the test set with an accuracy of 100%.

4. **Rationale behind Model Choices:**

Linear Kernel was chosen due to its accuracy and ability to capture complex relationships between variables. It outperformed other models in cross-validation and exhibited good generalizability on the test set.

5. **Code Implementation**
   https://github.com/Sharath1036/task-intern-career/tree/main/titanic

6. **Conclusion:**

Machine learning techniques efficiently predicted passenger survival on the Titanic based on various features. Linear Kernel performed best, demonstrating the model's effectiveness for this dataset. This analysis provides insights into factors impacting passenger survival and potential improvements for future disaster situations.