Concept Notes: Titanic Survival Prediction Project

Project Objective:

To build a machine learning model that predicts whether a passenger on the Titanic survived or not based on various features such as age, sex, ticket class, fare, and more.

1. Data Loading:

- Loaded the dataset using pandas from 'Titanic-Dataset.csv'.

- Displayed the first few rows to understand the structure of the data.

2. Initial Data Exploration:

- Checked data types and null values using .info() and .isnull().sum().

- Observed summary statistics using .describe() to understand numeric distributions.

3. Data Cleaning:

- Filled missing values in 'Age' with the median.

- Filled missing values in 'Embarked' with the mode.

- Dropped irrelevant or high-missing-value columns like 'Cabin', 'Name', and 'Ticket'.

4. Data Visualization:

- Plotted survival counts using seaborn.

- Visualized survival distribution across gender and passenger class.

- Plotted the age distribution of passengers.

5. Label Encoding:

Concept Notes: Titanic Survival Prediction Project

- Converted categorical columns ('Sex', 'Embarked') to numerical using LabelEncoder.

6. Feature Selection and Splitting:

- Selected input features (X) and target variable (y).

- Split the data into training and testing sets (80-20 split).

7. Model Building:

- Used Logistic Regression from sklearn to build the predictive model.

- Trained the model on training data using .fit().

8. Prediction and Evaluation:

- Made predictions on test data using .predict().

- Evaluated the model using accuracy score, classification report, and confusion matrix.

9. Model Saving:

- Saved the trained model using joblib.dump() to use it later without retraining.

Conclusion:

The project followed a structured data science pipeline from data loading to model evaluation. The logistic regression model gave a reasonable performance and could predict the survival status of Titanic passengers based on historical data.

This end-to-end workflow reflects a typical machine learning classification project and helps in understanding the real-world application of predictive modeling.