CodSoft: Data Science Internship

Titanic Survival Prediction

Description: The Titanic Survival Prediction project aims to analyze and predict the survival of passengers aboard the RMS Titanic, the infamous ship that sank in 1912 after hitting an iceberg. Using a dataset containing information about the passengers, including their age, gender, class, and more, the goal is to build a predictive model that can estimate the likelihood of survival for individuals on the ship.

Dataset : Titanic Dataset

Tools and Libraries:

- Python: Programming language used for the entire pipeline.
- Pandas & NumPy: For data manipulation and preprocessing.
- Matplotlib & Seaborn: For data visualization.
- Scikit-learn: For machine learning algorithms and model evaluation.

Key Features:

- **Data Exploration and Visualization:** Analyze the distribution of the dataset, focusing on features such as age, gender, passenger class, and ticket fare. Explore relationships between different features and the target variable (survival) using data visualization tools like Seaborn and Matplotlib to identify patterns and correlations.

- **Data Preprocessing:** Prepare the data for model training by addressing issues such as missing values, outliers, and scaling. Convert categorical features into numerical representations using techniques like one-hot encoding.
- **Model Selection:** Choose and train a suitable classification model to predict survival. Although models like Logistic Regression, Decision Tree, or k-Nearest Neighbors (k-NN) could be considered, the Logistic Regression algorithm is selected for this project due to its robustness and ability to handle complex relationships between features.
- **Model Evaluation:** Assessing the model's performance using metrics like accuracy, precision, recall, and F1 score. Cross-validation may be employed to ensure the model's robustness.
- **Deployment:** Implementing the trained model for predictions on new data, making it practical for real-world applications, such as in gardening or botanical studies.