**TITLE:** CodTech IT Solutions Internship - Task Documentation: “To-DO LIST” Using CSS, HTML, JAVASCRIPT.

**INTERN INFORMATION:**

**Name:** Sunkara Vamakeswari

**ID:** C0D6720

**INTRODUCTION**

**Task-1 TITANIC SURVIVAL PREDICTION**

**Introduction**

The Titanic survival prediction project is a classic example in the field of data science, often used by beginners to gain practical experience in data analysis, machine learning, and predictive modeling. The project revolves around predicting whether a passenger aboard the Titanic survived or not, given various attributes such as age, gender, ticket class, fare, and cabin.

The sinking of the RMS Titanic on its maiden voyage in 1912 is one of the most infamous maritime disasters in history. The tragedy resulted in the loss of over 1,500 lives and sparked significant advancements in maritime safety regulations. Analyzing the passenger data provides insights into the factors that influenced survival rates and allows us to build predictive models to estimate the likelihood of survival for individual passengers.

This documentation will guide you through the implementation of a Titanic survival prediction model using Python and popular machine learning libraries. We will cover data preprocessing, exploratory data analysis, feature engineering, model selection, training, evaluation, and deployment.

**Implementation**

**1. Data Loading :**

- Load the Titanic dataset, typically provided in a CSV format, into a pandas DataFrame. This step involves using the read\_csv() function from the pandas library. The dataset contains information about each passenger, including attributes such as age, gender, ticket class, fare, cabin, and survival status.

**2. Data Preprocessing:**

- Handle missing values: Check for missing values in the dataset and decide on an appropriate strategy for handling them. This could involve imputation (replacing missing values with the mean, median, or mode), deletion of rows or columns with missing values, or using advanced imputation techniques.

- Encode categorical variables: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding. This ensures that categorical variables can be used as input in machine learning algorithms.

- Feature scaling: Scale numerical features to a similar range to prevent certain features from dominating the model training process. Common scaling techniques include standardization (scaling features to have a mean of 0 and a standard deviation of 1) and min-max scaling (scaling features to a range between 0 and 1).

**3. Exploratory Data Analysis (EDA):**

- Explore the dataset to understand its structure, distribution, and relationships between variables. EDA involves techniques such as summary statistics, data visualization (histograms, box plots, scatter plots), and correlation analysis. Insights gained from EDA help in feature selection, identifying patterns, and understanding the data's characteristics.

**4. Feature Engineering:**

- Create new features or transform existing ones to improve the predictive power of the model. Feature engineering involves extracting relevant information from existing features, combining multiple features, or generating entirely new features based on domain knowledge. This step aims to capture important patterns and relationships in the data that may not be apparent initially.

**5. Model Selection :**

Choose appropriate machine learning models for the task of predicting Titanic survival. Common models used for binary classification tasks like this include logistic regression, decision trees, random forests, support vector machines (SVM), and gradient boosting algorithms (e.g., XGBoost, LightGBM).

-Consider factors such as model performance, interpretability, computational complexity, and the ability to handle non-linear relationships when selecting a model.

**6. Training and Evaluation :**

- Split the dataset into training and testing sets to train the model on a portion of the data and evaluate its performance on unseen data.

- Train the selected machine learning model on the training data and evaluate its performance using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC).

- Use techniques like cross-validation to assess the model's generalization ability and mitigate issues like overfitting.

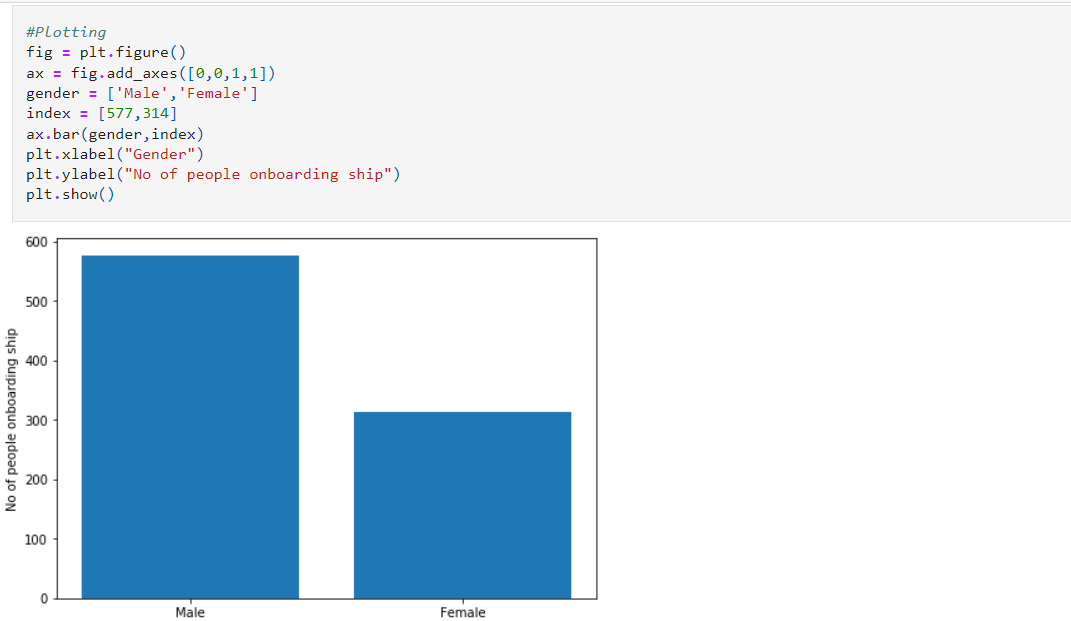
**7. Deployment and Monitoring :**

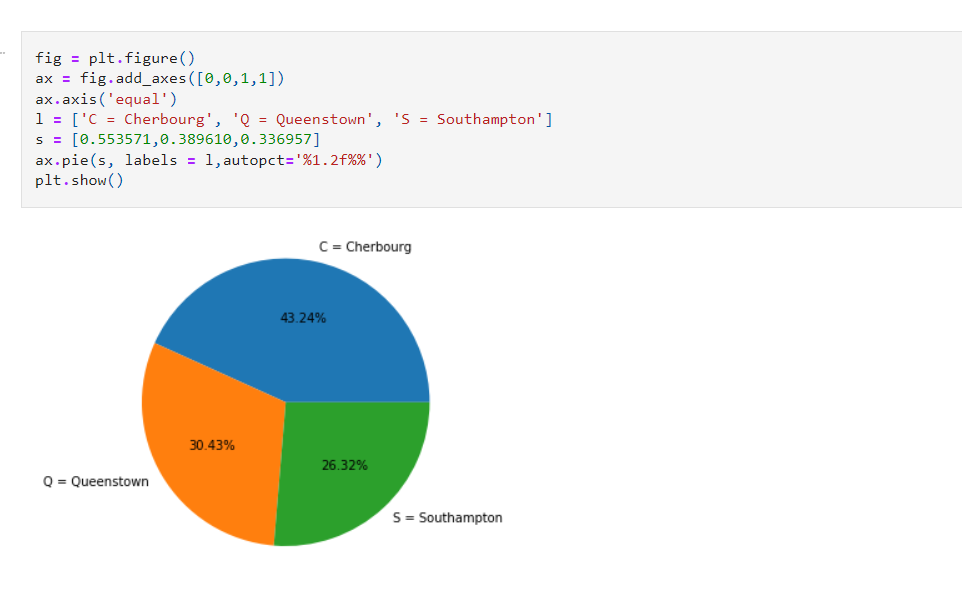
- Once a satisfactory model is obtained, deploy it into production or use it for predictions on new data.

- Implement monitoring mechanisms to track the model's performance over time and ensure its continued effectiveness. This may involve regular retraining of the model with updated data or adjusting model parameters as necessary.

**CODE EXPLAINATION**

****

****

****

**CONCLUSION**

In conclusion, the Titanic survival prediction project provides a valuable opportunity to apply fundamental data science concepts and machine learning techniques to a real-world dataset. Throughout this project, we have explored the Titanic dataset, preprocessed the data, performed exploratory data analysis, engineered features, selected a suitable machine learning model, trained the model, and evaluated its performance.

Through meticulous data preprocessing and feature engineering, we transformed raw data into a format suitable for training machine learning models. Exploratory data analysis helped us gain insights into the dataset's characteristics and identify patterns that could influence survival outcomes.

Model selection involved choosing an appropriate algorithm capable of effectively capturing the relationships between features and survival status. We trained the selected model on the training data and evaluated its performance using various evaluation metrics to ensure its reliability and effectiveness.