**TITLE:** CodTech IT Solutions Internship - Task Documentation: “TITANIC SURVIVAL PREDICTION” Using Python programming

**INTERN INFORMATION:**

**Name:** Dibjyoti Hota

**ID:** ICOD6251

**INTRODUCTION**

The sinking of the RMS Titanic on its maiden voyage in April 1912 remains one of the most infamous maritime disasters in history. The tragic event resulted in the loss of over 1,500 lives, sparking global sorrow and raising questions about maritime safety and emergency preparedness. In the decades since, the Titanic disaster has captivated the public imagination, serving as the backdrop for countless books, films, and studies.

In this project, we embark on a journey to explore the Titanic dataset and employ machine learning techniques to predict the survival outcome of passengers. Our objective is to build a model that accurately determines whether a passenger aboard the Titanic survived or perished based on various demographic and socio-economic factors.

The dataset provides a treasure trove of information about individual passengers, including their age, gender, ticket class, fare, cabin, and ultimately, whether they survived the ordeal. By leveraging this rich dataset and employing predictive modeling techniques, we aim to uncover patterns and insights that shed light on the factors influencing survival rates aboard the ill-fated Titanic.

Through meticulous data pre-processing, exploratory data analysis (EDA), feature engineering, and model selection and training, we will meticulously craft a predictive model capable of discerning survival outcomes. Our approach involves employing popular machine learning algorithms, such as Random Forest, to harness the predictive power hidden within the data.

By the culmination of this project, we not only seek to build a robust predictive model but also to contribute to the broader understanding of the Titanic disaster. Our endeavor is not merely an exercise in data science but a tribute to the memory of those who perished aboard the Titanic and a testament to the enduring fascination with this tragic event.

Join us as we embark on this voyage through data and embark on a quest to unveil the hidden truths of the Titanic disaster, one prediction at a time.

**Implementation**

* Python Programming: Used python to build the titanic survival prediction
* Numpy/pandas: use of numpy and pandas modules to store data and to calculate mathematical problems matplot and sklearn is used.
* Responsive Design: Implement responsive design principles to ensure optimal viewing experience across desktop and mobile devices.
* User Interface Components: Utilize UI libraries for designing interactive and visually appealing components.

**CODE EXPLAINATION**

# Step 1: Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix

# Step 2: Load the dataset

titanic\_data = pd.read\_csv("train.csv")

# Step 3: Data Preprocessing

# Check for missing values

print(titanic\_data.isnull().sum())

# Handle missing values

titanic\_data["Age"].fillna(titanic\_data["Age"].median(), inplace=True)

titanic\_data["Embarked"].fillna(titanic\_data["Embarked"].mode()[0], inplace=True)

# Convert categorical variables into numerical

titanic\_data = pd.get\_dummies(titanic\_data, columns=["Sex", "Embarked"])

# Drop irrelevant features

titanic\_data.drop(["PassengerId", "Name", "Ticket", "Cabin"], axis=1, inplace=True)

# Step 4: Exploratory Data Analysis (EDA)

sns.countplot(x="Survived", data=titanic\_data)

plt.show()

sns.countplot(x="Survived", hue="Sex\_male", data=titanic\_data)

plt.show()

sns.countplot(x="Survived", hue="Pclass", data=titanic\_data)

plt.show()

# Step 5: Feature Engineering (No additional features in this example)

# Step 6: Model Selection and Training

X = titanic\_data.drop("Survived", axis=1)

y = titanic\_data["Survived"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Step 7: Model Training

scaler = StandardScaler()

X\_train\_scaled = scaler.fit\_transform(X\_train)

X\_test\_scaled = scaler.transform(X\_test)

rf\_classifier = RandomForestClassifier(random\_state=42)

rf\_classifier.fit(X\_train\_scaled, y\_train)

# Step 8: Model Evaluation

y\_pred = rf\_classifier.predict(X\_test\_scaled)

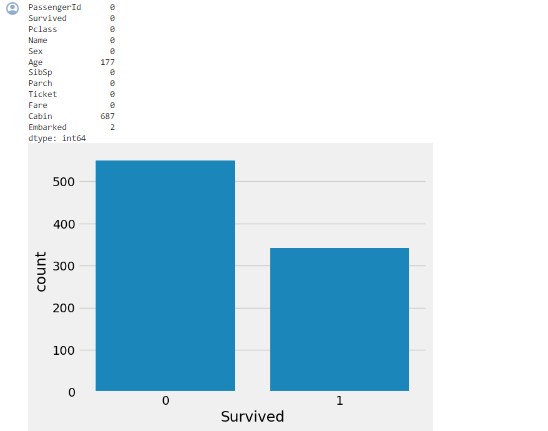
accuracy = accuracy\_score(y\_test, y\_pred)

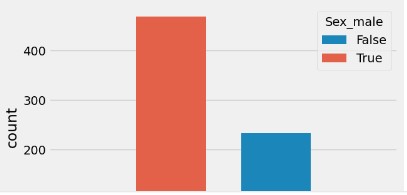
print("Accuracy:", accuracy)

print("Classification Report:\n", classification\_report(y\_test, y\_pred))

print("Confusion Matrix:\n", confusion\_matrix(y\_test, y\_pred))

* OUTPUTS





**USAGE**

**Titanic survival prediction:** Users can use this machine learning model to predict the survival rates

**Observation:** Users can know all sorts male and female population with this dataset

**Easiness:** Easiness in sorting the data and the survival rates.

**CONCLUSION**

In conclusion, the Titanic dataset to build a model that predicts whether a passenger on the Titanic survived or not. This is a classic beginner project with readily available data. . The dataset typically used for this project contains information about individual passengers, such as their age, gender, ticket class, fare, cabin, and whether or not they survived.