# TITANIC SURVIVAL PREDICTION

***Submitted by***

**CB.EN.U4CCE21016 – DHIVYADHARSHAN G**

**CB.EN.U4CCE21036 – HARI KRISHNA N**

**CB.EN.U4CCE21056 – S RAJESH**

**CB.EN.U4CCE21076 – UTKARSH BHARDWAJ**



# B.TECH COMPUTER AND COMMUNICATION ENGINEERING

**2021 BATCH**

**PAPER SUMMARY :**

The RMS Titanic was a British luxury passenger liner that sank in the North Atlantic Ocean on April 15, 1912. It was a large and opulent ship known for its supposed unsinkable design. However, it collided with an iceberg, leading to its tragic demise and the loss of over 1,500 lives.

The fateful incident still compel the researchers and analysts to understand what can have led to the survival of some passengers and demise of the others. With the use of machine learning methods and a dataset consisting of 891 rows in the train set and 418 rows in the test set, the we have done an analysis determining the correlation between factors such as age, sex, passenger class, fare etc. to the chance of survival of the passengers.

The Titanic Survival Project is a classic machine learning problem that aims to predict which passengers aboard the RMS Titanic survived the ship's tragic sinking. The dataset used for this project contains information about each passenger, such as their age, gender, ticket class, and other relevant features.

The goal of the project is to build a machine learning model that can accurately predict whether a passenger survived or not based on the provided information. This is a binary classification problem, where the model needs to classify passengers into two categories: survived or not survived.

During the model building process, several steps are typically involved, including data preprocessing, feature engineering, model selection, and evaluation. Data preprocessing involves handling missing values, scaling numerical features, and encoding categorical variables. Feature engineering may involve creating new features or transforming existing ones to improve the model's performance.

Once the model is trained and tuned, it can be evaluated using performance metrics such as accuracy, precision, recall, and F1 score. Cross-validation techniques can also be applied to assess the model's generalization capabilities.

The story of the Titanic continues to fascinate and serves as a reminder of human ambition and the consequences of overconfidence.

**BLOCK DIAGRAM :**

+----------------------------------+

| |

| Titanic Dataset |

| |

+---------------------+------------+

|

|

+----------v-----------+

| |

| Data Preprocessing |

| |

+----------+-----------+

|

|

+----------v-----------+

| |

| Feature Selection |

| |

+----------+-----------+

|

|

+----------v-----------+

| |

| Model Training |

| |

+----------+-----------+

|

|

+----------v-----------+

| |

| Model Evaluation |

| |

+----------+-----------+

|

|

+----------v-----------+

| |

| Model Testing |

| |

+----------+-----------+

|

|

+----------v-----------+

| |

| Model Deployment |

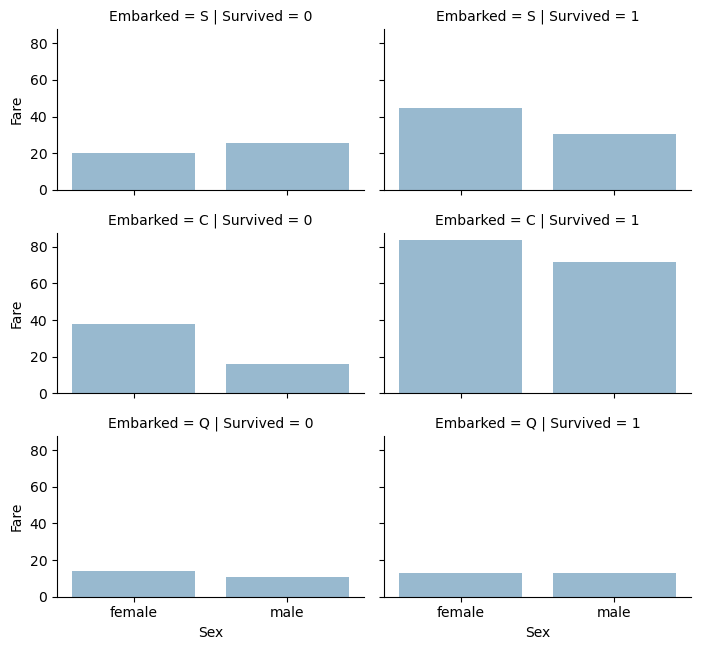
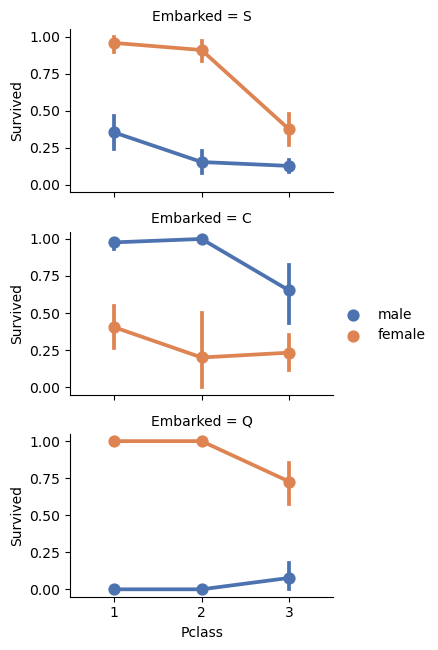
| |

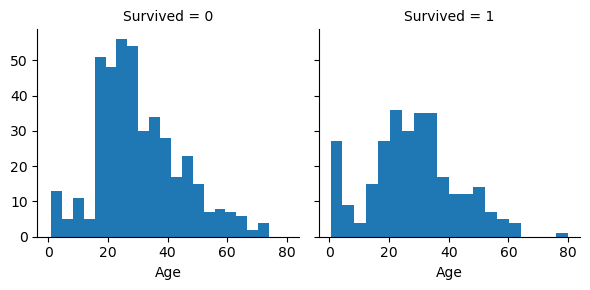
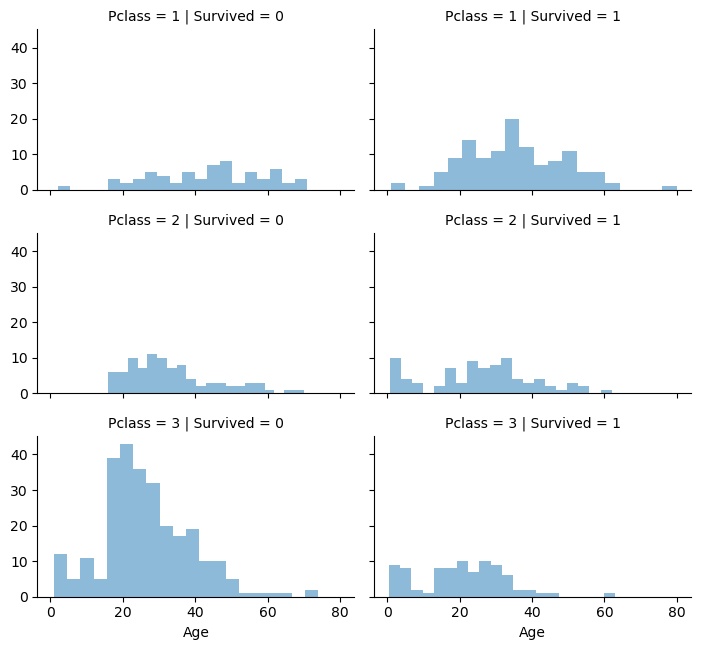
+----------------------+

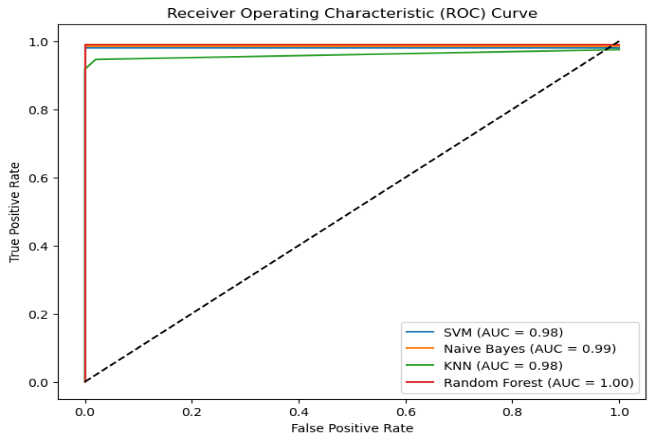
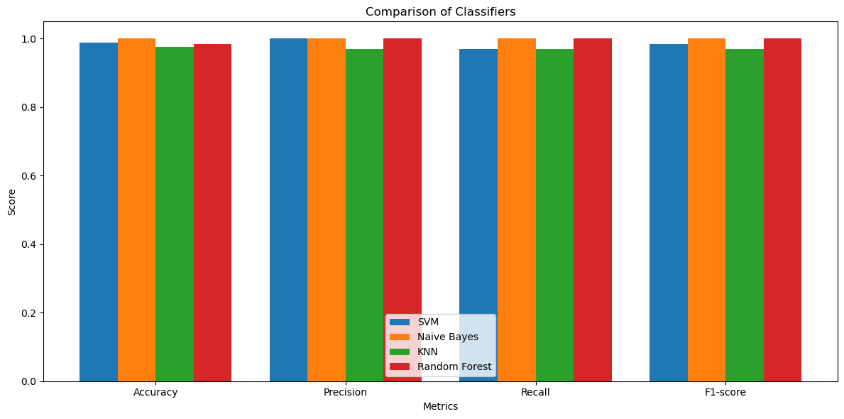
* **Titanic Dataset**: The initial source of data containing information about Titanic passengers, such as their age, gender, ticket class, fare, and survival status.
* **Data Preprocessing**: The process of cleaning and transforming the raw dataset to ensure it is suitable for analysis. This includes handling missing values, encoding categorical variables, and scaling numerical features.
* **Feature Selection**: Identifying and selecting the most relevant features from the dataset that contribute significantly to predicting the survival outcome. This step helps reduce complexity and improve model performance.
* **Model Training**: Training a machine learning model using the processed dataset to learn patterns and relationships between the input features and the target variable (survival status). Various algorithms like decision trees, random forests, or neural networks can be used for this purpose.
* **Model Evaluation**: Assessing the performance of the trained model using evaluation metrics like accuracy, precision, recall, or F1-score. This step helps determine how well the model generalizes to unseen data.
* **Model Testing**: Applying the trained model to new, unseen data to make predictions about the survival outcome of Titanic passengers. This step verifies the model's ability to make accurate predictions on real-world examples.
* **Model Deployment**: Integrating the trained model into a production environment where it can be used to make predictions on new data in real-time.

Each of these steps plays a crucial role in developing a machine learning model for the Titanic dataset, enabling the prediction of passenger survival based on the available features.

**PLOTS OBTAINED** :

**RESULT** :

We get highest score of accuracy while training with Random Forest as they have low tendency to get overfit on training set. The early observations of women aged persons infants and passengers of first class surviving more were identified using feature extraction and suitable plots.

The models have revealed patterns and relationships that highlight the social dynamics and biases prevalent during the time of the Titanic disaster. Overall, the use of machine learning in Titanic survival prediction has offered a powerful tool for uncovering hidden patterns, making accurate predictions, and enhancing our understanding of the factors that influenced the fate of the passengers.

It serves as a reminder of the potential of machine learning to contribute to historical analysis and provide valuable insights into significant events from the past.

**REFERENCE:**

M. A. Whitley, *Using statistical learning to predict survival of passengers on the RMS Titanic*, 2015.

E. Lam and C. Tang, *CS229 Titanic–Machine Learning From Disaster*, 2012.

