****

**Assessment Report**

on

**“Titanic Survival Prediction”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

GROUP-6  
 1. Arpit Choudhary - Roll No. 66   
2. Ankit Kumar Gautam - Roll No. 47  
 3. Aryan Singh - Roll No. 71  
 4. Aastha Srivastava - Roll No. 4   
5. Alok Verma - Roll No. 30

Section: A

**Under the supervision of**

“Teaching Faculty Name”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

## 1. Introduction

The sinking of the RMS Titanic in 1912 is one of the most infamous shipwrecks in history. A wealth of data was collected about the passengers on board, and this dataset is often used to explore predictive modeling. This project aims to analyze the Titanic dataset and predict whether a passenger survived based on certain features using a machine learning algorithm — specifically, **Naive Bayes**.

## 2. Problem Statement

The primary problem addressed in this project is to determine which features influenced the survival of Titanic passengers and use this information to **predict survival outcomes** using a classification model. This is framed as a **binary classification problem** (survived or not).

## 3. Objectives

* To explore and understand the Titanic dataset.
* To perform data preprocessing and cleaning.
* To implement a **Naive Bayes classifier** for survival prediction.
* To evaluate the model’s performance using appropriate metrics.
* To interpret and analyze the results for insights.

## 4. Methodology

* Load and explore the Titanic dataset.
* Clean the data to handle missing or irrelevant entries.
* Encode categorical variables numerically.
* Apply the **Gaussian Naive Bayes** classification algorithm.
* Split the dataset into training and testing sets.
* Evaluate the model using accuracy and classification metrics.

## 5. Data Preprocessing

* **Dataset Source**: The dataset is loaded from Seaborn's Titanic dataset or Kaggle’s Titanic competition dataset.
* **Cleaning Steps**:  
  + Removed columns not helpful for prediction (e.g., deck, embark\_town, alive).
  + Handled missing values:  
    - Replaced missing age with median.
    - Replaced missing embarked with the mode.
  + Converted categorical variables (sex, embarked, class) to numeric values.

## 6. Evaluation Metrics

* **Accuracy**: Measures how often the model predicts correctly.
* **Precision, Recall, F1-score**: Provide a better understanding of model performance across both survival classes.

print("Accuracy:", accuracy\_score(y\_test, y\_pred))

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

## 7. Results and Analysis

* **Accuracy**: ~78%
* **Classification Report**:
* precision recall f1-score support
* 0 0.82 0.87 0.84 67
* 1 0.72 0.63 0.67 38
* accuracy 0.78 105
* macro avg 0.77 0.75 0.75 105
* weighted avg 0.78 0.78 0.78 105

**Insights**:

* The model performs better at identifying passengers who did not survive.
* Performance for predicting survival could be improved by advanced feature engineering or more complex models.

## 8. Conclusion

The Naive Bayes classifier provided a simple yet effective approach to Titanic survival prediction. Through proper data preprocessing and feature encoding, we achieved a respectable accuracy of around 78%. While this model offers a good baseline, its performance could be enhanced with feature engineering, cross-validation, or more complex classification algorithms like Random Forest or XGBoost.

## 9. References

* Titanic Dataset - Seaborn Titanic Dataset
* Scikit-learn Documentation -<https://scikit-learn.org/>
* Kaggle Titanic ML Challenge - https://www.kaggle.com/competitions/titanic