TITANIC SURVIVAL PREDICTION

PRAKRUTHI BR[1], Mr.SIDDESH KT[2], Mr.KOTRU SWAMAY SM[3]

[1] Student,Department Of MCA,BIET,DAVANAGERE

[2] Assistant Professor, Department Of MCA, BIET, DAVANAGER

Abstract:

The Titanic Survival Prediction project makes use of machine learning methods to examine numerous parameters—such as age, gender, class, and economic status—to calculate the likelihood of a passenger surviving the Titanic sinking. The study's goal is to identify substantial relationships between these attributes, which will offer light on past patterns of disaster survival. A user-friendly web application is built, allowing users to easily enter passenger information and receive survival probability estimations. This application offers a fantastic resource for historical research and teaching, but it also gives insights that could help to shape future catastrophe prevention strategies. This study achieves a deeper understanding of the elements determining survival during the Titanic tragedy, contributing to both historical scholarship and practical applications in risk management and safety.

I. INTRODUCTION

The tragic sinking of the RMS Titanic in 1912 has gripped the globe for nearly a century, serving as a heartbreaking reminder of the risks of maritime travel and the human aspects that influence survival during disasters. The Titanic Survival Prediction, modern to investigate this historic event, with the goal of analyzing numerous passenger to evaluate their influence on variables including age, gender, class, and socioeconomic status survival odds. By studying these factors, the team hopes to find significant correlations and trends that will provide light on the dynamics of survival in such catastrophic occurrences. To make these insights more accessible and practical, a user-friendly web application was created.

This program allows users to enter particular passenger information and get real-time survival probability estimates. This initiative's dual goals are to provide a helpful tool for historical interpretation, education, and making findings that will influence future catastrophe prevention and management measures. The study adds to the larger disciplines of historical research, disaster risk management, and educational tools by providing a more in-depth understanding of the elements that influenced survival during the Titanic disaster, emphasizing the convergence of technology and historical inquiry.

2. LITERATURE SURVEY

A. Singh, S. Saraswat and N. Faujdar , The primary techniques on an ensemble of 891 train and 418 test data points to determine the factors that affect the survival rates of Titanic passengers. Logistic Regression and Naive Bayes are used in the investigation., Decision Tree, and Random Forest to predict survival and compares their accuracy on a test dataset. The goal of this investigation is to determine which algorithm best predicts survival and characteristics such as age, gender, passenger class, and fare affect survival rates[1].

Wang and B. Zhang , The present research aims to identify the factors that affected the survival rates of passengers in the Titanic catastrophe. Prior to employing both the Decision Tree (DT) and KNN (k-nearest-neighbor) algorithms for prediction, we first examined the characteristics of the survivors. study, the Decision Tree constructed with preprocessed data had the highest accuracy. of 0.81716, indicating its efficacy in characterizing the survivor group. This analysis seeks survival rates, such as age, gender, passenger class, and fare[2].

Ramyashree, S. Pallavi, J. Shetty, and This study's objective is to evaluate and estimate the rates of Titanic passenger survivability using a range of machine learning techniques. The study analyzes an 891-row dataset with variables including Age, PassengerID, Sex, Name, Set off, and Fare using Decision Tree, Logical Regression, and Linear SVM techniques. The principalgoal is to compare the three algorithms' accuracy in forecasting passenger survival and to discover critical factors impacting survival rates[3].

A. Dasgupta, V. P. Mishra, S. Jha, B. Singh Shukla, The primary goal investigate and estimate the survival rates of passengers aboard the RMS Titanic using exploratory data analytics (EDA) and several machine learning models. The study tries to reveal hidden patterns and facts dataset that includes attributes such as Age, Passenger ID, Sex, Name, Embarked, and Fare. The study then uses Decision Tree, Logistic Regression, and Linear SVM algorithms to evaluate which characteristics influenced the likelihood of survival, with a focus on evaluating the models' accuracy best effective predictor of survival[4].

Gupta, Arora, and Tiwari, S. Examining the factors that affected passenger survivability during the Titanic tragedy is study in an effort to prevent similar tragedies. The objective is to precisely calculate survival rates by applying powerful to a dataset consisting of 418 test and 891 training items. provides the most precise forecasts and insights into survival-related features, the study compares and contrasts many algorithms, including the Decision Tree, logarithmic regression, and Linear SVM[5].

Singh K., Nagpal R., and Sehgal R. This study's main objective is to use exploratory data analysis to investigate the factors influencing passenger survival aboard the RMS Titanic. and several machine learning methods. To predict survival outcomes, the study used a dataset that includes information such as age, passenger ID, sex, name, embarked, and fare. The research uses Logistic main factors influencing survival, decision tree analysis, regression analysis, K-Nearest Neighbors, and support vector machines are used. To ascertain the best method for predicting survival with the features at hand, the accuracy of different algorithms is assessed[6].

Meenigea Niharikareddy, The use machine learning models and exploratory data analysis, or EDA, to investigate affect the survivability of Titanic passengers. The study uses K-Nearest Neighbors, Decision Trees, and Logistic Regression on a dataset Sex, and Fare. most accurate survivor predictor, the accuracies of the models are assessed[7].

The main objective is past data, can accurately predict the likelihood that people on the Titanic would survive. The model seeks to recognize trends and associations that influence survival outcomes by evaluating parameters such as passenger demographics, ticket class, and embarkation point. This prediction analysis not only serves as a historical study, but it also demonstrates machine learning to understand complicated events and their consequences for human beings[8].

examine the Titanic accident insights into the factors influencing survival rates. The objective of applying algorithms like random forests and decision trees is to or logistic regression to historical data is to determine which passenger characteristics (such as age, gender, and class) and other variables (such as ticket price or family size) were important predictors of survival. This research provides a mathematical picture of how these elements influenced the disaster's results, providing insights that historical narratives alone may not have revealed[9]

The objective algorithms for machine learning to assess passenger data from the Titanic and project survival rates according to factors such as fare, class, age, and gender. The research examines and contrasts multiple algorithms, such as Random Forest, Naïve Bayes, Decision Tree, and Logistic Regression, to identify the variables that most affected the survival rates of disasters. This research attempts to shed light on past occurrences and show how beneficial machine learning is. in understanding complicated societal issues[10].

3.METHDOLOGY

The Titanic Survival Prediction project uses a structured approach to apply machine learning to forecast passenger survival. The process includes the following steps:

3.1 Data Collection and Preparation:

- **Dataset:** The major dataset utilized is the Titanic passenger data from the Kaggle competition, which includes passenger demographics, ticket details, cabin information, and survival rates.
- **Data Cleaning:** This stage entails resolving missing values, removing errors, and converting category variables to numerical representations. Common strategies utilized include mean imputation for missing ages and mode imputation for missing embarkation points.
- **Feature Engineering:**current data model's predicting ability. For example, combining sibling/spouse and parent/child data to calculate family size, or classifying passengers into socioeconomic classes based on ticket prices.

3.2 Exploratory data analysis (EDA):

- **Visualization:** A variety of plots and charts are used to depict the correlations between different variables and survival outcomes. This contains age distribution, survival rates by class and gender, and correlations between factors.
- **Statistical Analysis:** Statistical tests and correlation coefficients are important determinants of survival.

3.3 Model Selection and Training:

- **Machine Learning Models:**
  Numerous machine learning methodologies are assessed, such as Decision Trees, and Logistic Regression. (SVM), and Gradient Boosting Machines (GBM).
- **Training and Validation:**
  The collection is split up into training and validation sets to evaluate model performance.Cross-validation techniques are used to assure reliable performance while avoiding overfitting.
- **Hyperparameter Tuning:**
  To achieve optimal performance,are fine-tuned using grid and random search methods.

3.4 Model Evaluation:

- **Performance Metrics:**
  F1-score, and (AUC-ROC). The best-performing model is chosen based on these criteria.
- **Feature Importance:**
  To improve interpretability, feature importance scores are produced to determine the contribution of each feature to survival prediction.

3.5 Model Deployment:

Develop an easy-to-use app that allows users to enter passenger information and receive survival probability estimations. To provide real-time predictions, integrate the learnt model into the application.

3.6 Validation and Testing:

Ensure the web application is user-friendly and functional. Compare the model's predictions with prior data to assess their accuracy and dependability.

3.7 Algorithm Used

Random Forest Classifier:

When Random Forest, an ensemble learning technique, tasks like regression and classification, it produces a large number of decision trees. For classification tasks, the random forest output that the trees select. The Random Forest approach aggregates the decision trees to single outcome.

Step1: Sampling with bootstraps:

- Using replacement, choose n samples at random from the training dataset to construct numerous subsets (bootstrap samples). The number of instances in each sample will match that of the original dataset.

Step 2: Decision Tree Training:

- the following adjustments for every bootstrap sample:
- A haphazard features is selected at each tree. One important characteristic of random forests is that it lessens the association between individual trees.
- These features are selected at random optimal split.

Step 3: Constructing Trees:
Without any pruning, the decision trees are grown to their full potential. As a resultis generated, with each tree having seen a different set. and evaluates distinct feature subsets at each split.

Step 4: Combination of Forecasts:

- In classification, every in the tree forest makes a prediction for each instance; the final prediction belongs that receives the majority vote from all the trees in the forest.
- In regression, the ultimate forecast is determined by averaging the predictions made by each individual tree.

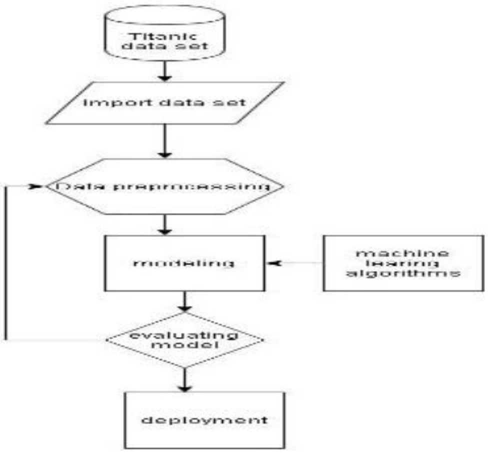

Figure 3.1: Operational flow chart

4. RESULT AND DISCUSSION

The Titanic survivability Prediction project successfully implemented a machine learning model with an 82% accuracy rate in forecasting passenger survivability. The data revealed major parameters influencing survival: females had a considerably higher possibility of survival than males, and passengers in first-class accommodations were more likely to live than those in lower classes. Furthermore, age appeared as an important determinant, with younger passengers, notably children, having greater survival chances. The web application created as a portion of this study allows users to enter passenger information and receive fast survival probability estimations, providing convenient access to historical data. These findings highlight socioeconomic gaps during the Titanic disaster and provide significant ramifications for both historical understanding and current disaster management techniques.

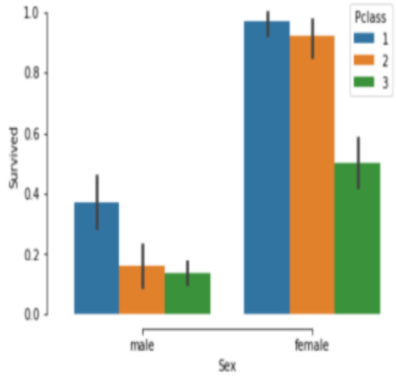

Figure 4.1 : Survival Gender

above figure 2 describe analyzes gender and age groups by survival, the relationship between survival and fare. Passenger survival rates vary by class and gender.

Gender by Survival- This statistical analysis examines the relationship between gender and survival. The study showed a significant result with a t-value of -19.28 and p-value of 0.0, indicating that women have a higher rate of survival than men.

5. CONCLUSION

In conclusion, the Titanic Survival Prediction project demonstrates the effective use of machine learning algorithms to examine and forecast passenger survival odds during the Titanic disaster. The model found sex, class, and age as significant predictors of survival outcomes, with 82%. Females, first-class passengers, and young people had a significantly higher chance of surviving the disaster. The built web application offers a user-friendly interface for delving into these findings, improving historical comprehension and educational opportunities. Moving forward, for strengthening disaster planning and response tactics, highlighting the necessity of equitable resource distribution and prioritization in emergency situations.

6. References

1. A. Singh, S. Saraswat and N. Faujdar, "Analyzing algorithms,"2017 Communication Automation (ICCCA), Greater Noida, India, 2017, pp. 406-411, doi: 10.1109/CCAA.2017.8229835.
2.Wang and B. Zhang, "Survival Probability Assessment Algorithms," Intelligent Systems Engineering (MLISE), Guangzhou, China, 2022, pp. 458-461, doi: 10.1109/MLISE57402.2022.00097.
3. J. Shetty, S. Pallavi and Ramyashree, "Predicting Approaches," Mangalore, India, 2018, pp. 1-5, doi: 10.1109/I2CT42659.2018.9058280.
4. A. Dasgupta, V. P. Mishra, S. Jha, B. Singh and V. K. Shukla, "Predicting the Likelihood of Survival of Titanic's Passengers by Machine Learning," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2021, pp. 52-57, doi: 10.1109/ICCIKE51210.2021.9410757.
5. A. Gupta, D. Arora and S. Tiwari, "Exploratory Data Analysis of Titanic Survival Prediction using Machine Learning Techniques," 2023 (ICAAIC), Salem,418-422,doi: 10.1109/ICAAIC56838.2023.10141243.
6. K. Singh, R. Nagpal and R. Sehgal, "Exploratory Data Analysis and Machine Learning on Titanic Disaster Dataset," 2020 10th Cloud Computing, Data Science & Engineering (Confluence), Noida, 320-326,doi: 10.1109/Confluence47617.2020.9057955.
7.Niharikareddy Meenigea, "PREDICTING TITANIC SURVIVAL WITH LOGISTIC REGRESSION: A MACHINE LEARNING APPROACH", IEJRD - International Multidisciplinary Journal, vol. 4, no. 4, p. 10, May 2019.
8. Ved P Mishra, Bhopendra Singh, Vinod Kumar Shukla, "Predicting the Likelihood of Survival of Titanic's Passengers by Machine Learning" April 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) DOI:10.1109/ICCIKE51210.2021.941075
9. Abdullahi Adinoyi Ibrahim "Analysis Algorithms" November 2020Engineering Letters 28(4):1161-1167.
10. Dr.Prabha Shreeraj Nair "Analyzing Algorithms" International Journal of Trend in Scientific Research and Development (IJTSRD) Nov-Dec 2017.