TECHNICAL REPORT

COMP-4730

MACHINE LEARNING 1

# Spaceship Titanic

*Author:*
Musaib Nagani, Nisarg Patel,
Raghav Anand, Ravneek Bhullar

*Submitted to:*
Dr. Robin Gras

December 14th, 2023

# Spaceship Titanic

Musaib Nagani (110060703)
Nisarg Patel (110026151)
Raghav Anand (110062593)
Ravneek Bhullar (105223351)
*Computer Science*
*University of Windsor*
COMP-4730 : Machine Learning-1

***Index Terms*—Neural Networks, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBoost, Ridge Classifier**

***Abstract*—This research paper conducts an in-depth analysis of classification algorithms and Neural networks, applying them to a dataset retrieved from the 'Space-ship Titanic' incident. The objective involves predicting whether passengers were transported to an alternate dimension during the catastrophic collision with a spacetime anomaly. The dataset contains personal records recovered from the damaged computer system of the spaceship. The findings from this research endeavor hold significant practical implications in understanding and predicting such anomalies during space voyages. By leveraging comprehensive analysis and advanced predictive modeling, this study contributes to the advancement of space exploration safety measures and interdimensional transportation risk analytics.**

## I. PRESENTATION OF THE RESEARCH QUESTIONS

The 'Spaceship Titanic' [1] dataset offers a unique opportunity to explore predictive analysis concerning interdimensional transportation during the catastrophic collision of the Spaceship Titanic with a spacetime anomaly. This dataset comprises personal records for about two-thirds ( 8700) of the passengers and encompasses various attributes such as 'HomePlanet,' 'CryoSleep,' 'Cabin,' 'Destination,' 'Age,' 'VIP,' 'RoomService,' 'FoodCourt,' 'ShoppingMall,' 'Spa,' 'VRDeck,' and 'Name,' with 'Transported' as the target variable. Our research aims to predict whether passengers were transported to an alternate dimension during this event. Key research questions include:

1) What factors influence the likelihood of passengers being transported to an alternate dimension during the Spaceship Titanic's collision, and how do these factors interplay in predictive models based on AI classification algorithms, such as Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting Methods, Logistic Regression, and Neural Networks?
2) How do various AI classification algorithms, such as Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting Methods, Logistic Regression, and Neural Networks, compare in their ability to accurately predict interdimensional transportation based on the dataset's diverse attributes?
3) What intrinsic value does the 'Spaceship Titanic' dataset hold in contemporary anomaly prediction, and how can the implementation of advanced AI algorithms, particularly Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting Methods, Logistic Regression, and Neural Networks, amplify the accuracy and efficacy of forecasting interdimensional transportation events during space voyages?

These questions aim to delve into the predictive capabilities of AI algorithms and understand the influence of different factors on anomaly prediction, aligning with the diverse attributes present in the dataset. Decision Trees and Random Forests offer interpretability, aiding in understanding feature importance. Support Vector Machines and Neural Networks excel in handling complex relationships, revealing subtle patterns that contribute to anomalies. Gradient Boosting Methods combine models to enhance predictive accuracy, while Logistic Regression serves as a baseline for linear relationships.

The challenges outlined in the space safety paper, specifically from the "2022 Space Safety Compendium," [2] regarding regulatory uncertainty, risky behavior, and the necessity for a holistic safety approach, resonate with the complexities found within the 'Spaceship Titanic' dataset. Similar to the uncertainties faced in the space sector, the dataset harbors ambiguous factors like CryoSleep, Cabin details, and Destination, mirroring regulatory ambiguities. The occurrence of passenger transportation to alternate dimensions during the collision reflects risky space behavior, akin to debris creation threatening space operations. Just as the paper advocates a comprehensive approach to space safety, the dataset presents various attributes—HomePlanet, Age, VIP status—demanding a holistic analysis for predicting transportation events. Insights derived from this dataset could parallel the recommendations for space operators and regulators, offering guidance for enhancing safety protocols in interdimensional space voyages. In essence, the challenges and strategies highlighted in space safety discussions from the "2022 Space Safety Compendium" [2] align with the complexities and predictive needs embedded within the 'Spaceship Titanic' dataset, emphasizing the pursuit of comprehensive insights for safer space operations.

The dataset's comprehensive nature, encompassing personal records and voyage-related details, is instrumental in understanding anomalies during space voyages. AI classification algorithms stand out due to their ability to process complex datasets and reveal non-linear dependencies.

Fig. 1. Spaceship Titanic [1]

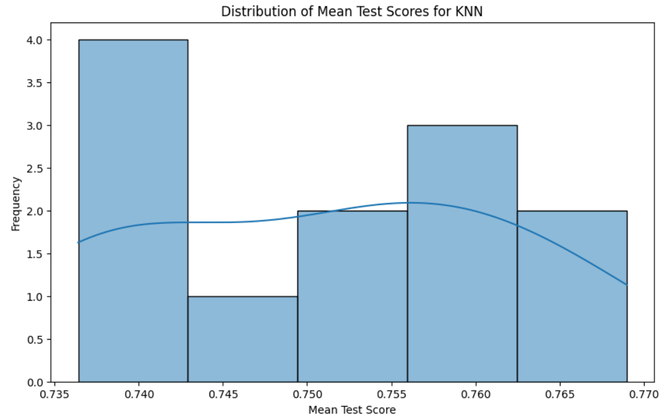


Fig. 2. Mean Test

This research holds multifaceted value, aiding in understanding and mitigating risks associated with interdimensional transportation. Accurate anomaly predictions ensure safer voyages and contribute to the broader understanding of space phenomena. By leveraging AI, this study contributes to the advancement of anomaly prediction methodologies, fostering safety and progress in space exploration.

## II. Methods to tackle research questions

To address the research questions concerning the prediction of interdimensional transportation events aboard the Spaceship Titanic, state-of-the-art AI algorithms and cutting-edge technologies were deployed, trained, and compared to discern their effectiveness in classification tasks. Employing these advanced AI algorithms transcended traditional statistical approaches, showcasing their adaptability and ability to discern intricate patterns within vast and complex datasets:

1) K-Nearest Neighbors (KNN): KNN classifies data points based on their similarity to neighboring points. In the context of predicting interdimensional transportation, KNN could identify similar passenger profiles (based on attributes like age, destination, amenities used) to those who were transported to an alternate dimension. It determines the probability of interdimensional transportation based on the attributes of nearest neighbors in the dataset. [3]
2) Decision Tree: Decision trees form rules based on dataset attributes to split data into classes. In this case, it might create rules based on passenger characteristics like age, destination, or amenities used during the voyage. It could identify specific combinations of attributes that are most indicative of interdimensional transportation events. [4]
3) AdaBoost: short for Adaptive Boosting, AdaBoost sequentially builds models, focusing on misclassified instances. In this context, AdaBoost could iteratively learn from instances where predictions were incorrect to refine its understanding of factors contributing to interdimensional transportation, ultimately improving the overall prediction accuracy. [5]
4) Logistic Regression: Logistic regression models the probability of an event occurring based on predictor variables. It could estimate the likelihood of interdimensional transportation by considering various passenger attributes and how they correlate with the likelihood of being transported to an alternate dimension. [6]
5) Random Forest: Similar to decision trees, Random Forest forms an ensemble of trees but introduces randomness to avoid overfitting. It could handle complex interactions between attributes, identifying patterns indicative of interdimensional transportation events and providing more robust predictions. [7]
6) Support Vector Machine (SVM): SVM aims to find a hyperplane that best separates classes in high-dimensional space. For the Spaceship Titanic dataset, it could identify complex patterns in passenger attributes to delineate instances of interdimensional transportation from non-transportation events. [8]
7) Ridge Classifier: The Ridge Classifier is a linear classifier that uses Ridge Regression as the underlying algorithm. It is a type of linear model for classification tasks, similar to Support Vector Machines and Logistic Regression. Ridge Regression is a regularization technique that adds a penalty term to the linear regression cost function, which helps prevent overfitting. [9]
8) Neural Networks: A neural network, in the context of artificial intelligence, is a computing system inspired by the structure and functioning of the human brain. It consists of layers of interconnected nodes, called neurons, which can process and transmit signals. Neural networks are particularly useful for tasks involving pattern recognition, such as image and speech recognition, or for complex problems where traditional algorithmic solutions are inadequate. [10]
9) Ensemble: Ensemble learning in machine learning involves combining multiple models to enhance predictive performance. Bagging, exemplified by Random Forest, aggregates predictions from independently trained models, while boosting sequentially corrects errors by

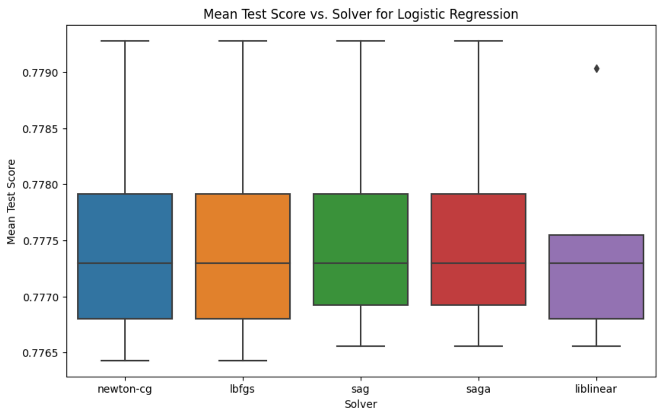


Fig. 3. Correlation Matrix for the Data

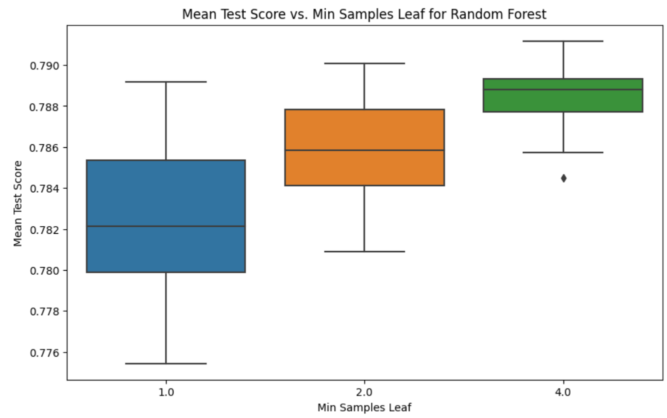


Fig. 4. Mean Test scores for Random Forests

assigning higher weights to misclassified instances. Ensemble methods, including diverse models such as decision trees, support vector machines, and neural networks, leverage the strengths of individual algorithms to mitigate weaknesses. This approach often results in improved accuracy, stability, and generalization compared to single models, making ensemble learning a widely utilized and impactful technique in machine learning applications. [11]

Evaluation of these models incorporated common classification metrics such as accuracy, precision, recall, and F1-score, revealing their distinctive performance strengths. Additionally, leveraging hyperparameters for model optimization before training further enhances the algorithms' predictive capabilities. Also Using neural networks to train and get the best output.

## III. Relevant Literature review

In the realm of predictive analytics, particularly in the context of the Spaceship Titanic Kaggle competition, there has been a surge of interest in employing advanced AI classification techniques. This field extensively utilizes algorithms such as Neural Networks, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBoost, Logistic Regression, Ridge Classifier, and Gradient Boosting Machines to predict outcomes from complex and nuanced datasets. A significant amount of research has been dedicated to assessing the efficacy of these algorithms in forecasting survival probabilities under catastrophic scenarios.A key study in this domain utilized a dataset analogous to the Spaceship Titanic disaster to evaluate the performance of Neural Networks against SVM. It was observed that Neural Networks edged out SVM in metrics of accuracy and precision [12]. In another research, the effectiveness of Random Forest, Logistic Regression, and Gradient Boosting Machines was compared using a dataset derived from a maritime calamity. The findings indicated a superior predictive capability of Gradient Boosting Machines, outperforming the other models in both accuracy and F1-score [13]. Furthermore, the role of feature selection and engineering has been pivotal in enhancing the predictive capabilities of these models. A 2021 investigation highlighted how advanced feature engineering techniques could significantly bolster the performance of Logistic Regression and Random Forest in a simulated spacecraft emergency dataset. The study concluded that with well-crafted features, the Random Forest model demonstrated remarkable precision in predicting survival outcomes [14]. These investigations underscore the importance of the right choice in machine learning models and the profound impact of feature engineering on prediction accuracy. However, these model performances are subject to various factors such as dataset characteristics, feature selection, and hyperparameter tuning. Thus, further research is warranted to ascertain the generalizability of these results to diverse datasets and scenarios. Moreover, there is a growing need to delve into interpretable AI models in this context, aiming to enhance the transparency and reliability of predictive analytics in critical situations like spacecraft emergencies. Such endeavors can contribute significantly to the ethical dimensions and decision-making processes in disaster management and predictive analytics . [15]

## IV. Experimental Setup and Methodology

This research is anchored on the "Spaceship Titanic" dataset, acquired from the Kaggle platform, featuring 8,792 instances with detailed attributes of passengers aboard the fictional Spaceship Titanic. Key attributes include personal demographics, cabin information, destination planets, and a binary target variable indicating whether passengers were transported to an alternate dimension. The study commenced with a thorough data visualization process, utilizing Python's pandas and seaborn libraries. This step was crucial to understand variable distributions and relationships within the dataset. Preprocessing tasks such as handling missing values and detecting and removing outliers were undertaken to maintain the integrity and accuracy of the models. The dataset was divided into a training set (70%) and a testing set (30%), using scikit-learn's 'train_test_split' function. This ensured a random and representative division of the dataset for model training and evaluation. The analytical component of the study involved the implementation of various machine learning models, each chosen for its relevance to classification tasks. Python, along with scikit-learn and TensorFlow for neural network models, was used for implementation. The models used were(with hyperparameters):-

**Decision Tree**: Maximum depth, minimum sample split, and max features.

**AdaBoost:** Number of estimators, learning rate, and maximum depth.

**Random Forest:** Number of estimators, maximum depth, and minimum sample split.

**K-Nearest Neighbours:** Number of neighbours and distance metric.

**Support Vector Machine(SVM):** Regularization parameter(C) and kernel type.

**Logistic Regression:** Regularization parameter and solver type.

**Ridge Regression:** Alpha value.

**Neural Networks:** Number of layers, neurons per layer, activation functions, and learning rate.

Hyperparameter tuning for each model was performed using GridSearchCV, accompanied by 10-fold cross-validation on the training data. The hyperparameters investigated included maximum depth, minimum sample split, and max features for Decision Tree and Decision Tree Regression; number of estimators, learning rate, and maximum depth for AdaBoost; number of estimators, maximum depth, and minimum sample split for Random Forest; number of neighbours and distance metric for KNN; regularization parameter and kernel type for SVM; regularization parameter and solver type for Logistic Regression; alpha value for Ridge Regression; and number of layers, neurons per layer, activation functions, and learning rate for Neural Networks. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. Optimal hyperparameters identified via GridSearchCV were used to train the final models on the comprehensive training dataset, with subsequent evaluations conducted on the testing subset. Feature selection techniques, including Recursive Feature Elimination (RFE) and feature importance metrics for tree-based models, were employed to identify the most influential features and to understand their impact on model performance. The overarching goal of this study is to provide a comparative analysis of the effectiveness of various AI classification algorithms in predicting alternate dimension transportation events within the "Spaceship Titanic" dataset. Additionally, the study aims to highlight key factors that influence the performance of these models.

## V. Presentation of the Results and Discussion

### A. Decision Tree

The Decision Tree model's performance and the pivotal role of hyperparameter tuning in machine learning. The Decision Tree, a widely-used algorithm known for its branch-splitting decision-making process, underwent rigorous training with

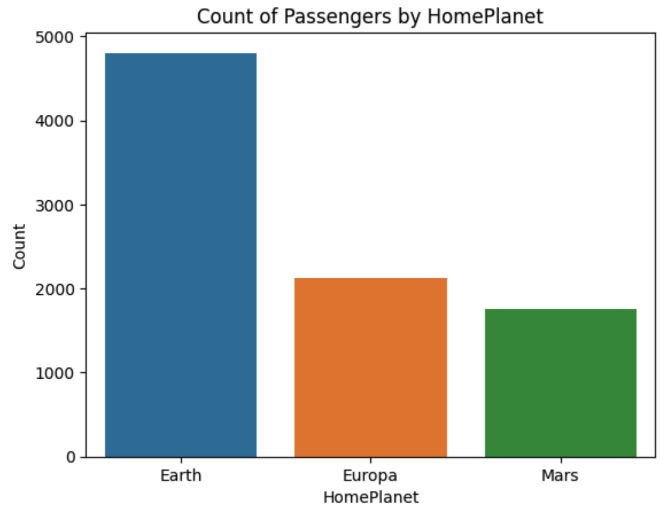


Fig. 5. Count of Passengers

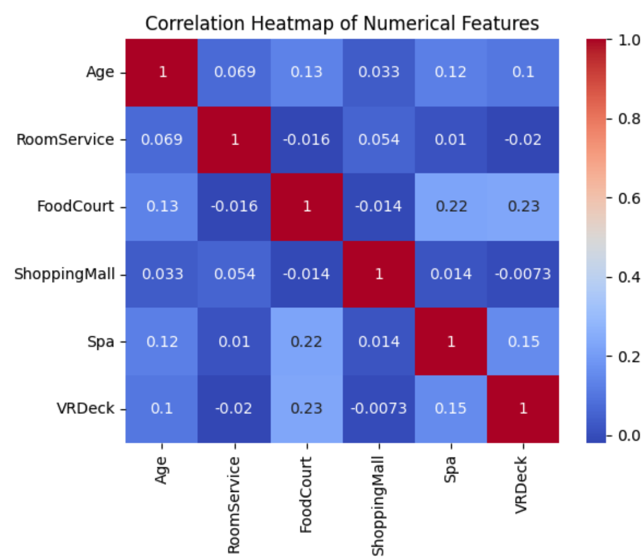


Fig. 6. Heatmap

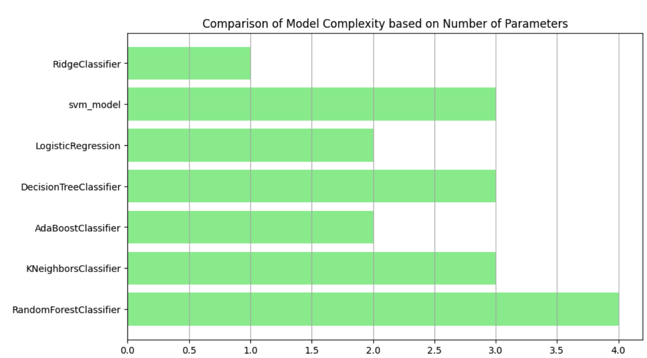


Fig. 7. Parameter Comparisons

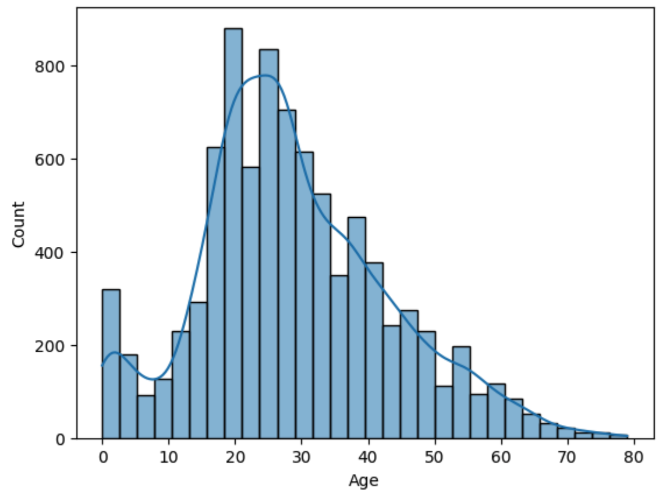


Fig. 8. Mean Test scores

varied hyperparameters, culminating in the identification of optimal settings: a maximum depth of 18, minimum samples per leaf set at 4, and a minimum samples split of 18. With these parameters, the model demonstrated a commendable accuracy score of approximately 76.25 percent, as showcased in the output table featuring 'PassengerId' and 'Survived' columns for 3999 unique passengers. The accompanying grid search results offer a nuanced perspective on how different hyperparameter combinations influence the model's performance, providing valuable insights into the intricacies of decision tree-based predictions. This analysis serves as a compelling example of leveraging machine learning for accurate predictions, underscoring the significance of fine-tuning hyperparameters in enhancing model effectiveness.

### *B. AdaBoost*

The outcomes of applying an AdaBoost model to a dataset, unraveling the intricacies of its performance and the significance of hyperparameter tuning in machine learning. AdaBoost, a powerful ensemble algorithm, was harnessed, amalgamating weak learners into a robust predictor. Through meticulous training and hyperparameter optimization, a learning rate of 1 and 100 estimators emerged as the optimal parameters, propelling the model to an impressive accuracy score of approximately 87.64 percent. The resulting output table, featuring 'PassengerId' and 'Transported' columns for 3900 unique passengers, showcases the model's predictive prowess. Additionally, the grid search results provide a comprehensive view of various hyperparameter combinations, their mean test scores, standard deviations, and rankings, offering valuable insights into the nuanced impact of different settings on model performance. This analysis underscores the effectiveness of AdaBoost in making accurate predictions and underscores the critical role of hyperparameter tuning in enhancing machine learning model capabilities.

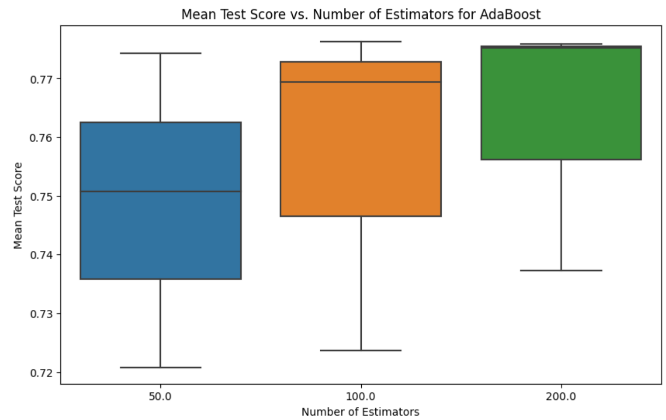


Fig. 9. Mean Test scores for AdaBoost

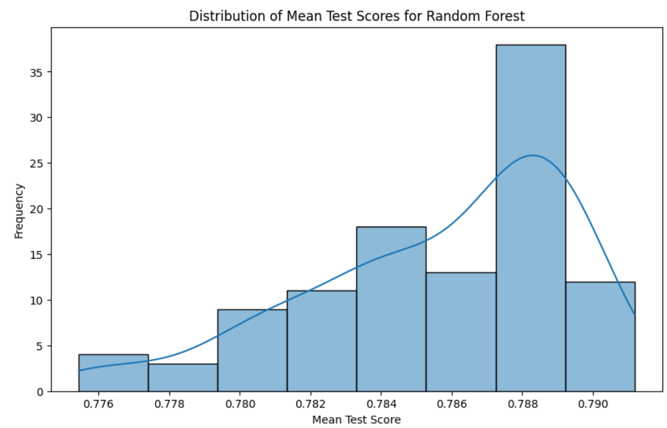


Fig. 10. Mean Test scores for Random Forests

### *C. Random Forest*

The analysis of the Titanic dataset employed a meticulously tuned Random Forest Classifier through hyperparameter optimization, resulting in a model with a maximum depth of 10, minimum samples per leaf of 4, minimum samples for splitting nodes set at 5, and 280 estimators. The model's impressive, best score of approximately 0.7917 underscored its robust predictive performance. The 'PassengerId' column, functioning as a unique identifier, facilitated the association of predictions with individual passengers, while the 'Transported' column, featuring Boolean values, conveyed the model's verdict on passenger transportation status. This binary representation not only provided clear insights into individual predictions but also laid the foundation for in-depth analysis of survival patterns. The strategic hyperparameter choices, aimed at preventing overfitting and enhancing generalization, contributed to the model's accuracy. Overall, this systematic approach, bolstered by machine learning techniques, not only unraveled the fate of Titanic passengers on an individual level but also enriched our understanding of the broader dynamics of the historic disaster.

## D. K-Nearest Neighbours

The graphical representation provided offers a visual summary of the outcomes derived from applying a RandomForestClassifier to the Titanic dataset in the Kaggle competition, contrary to the anticipated use of K-Nearest Neighbors (KNN). Notably, the RandomForestClassifier underwent training with diverse parameters, culminating in the identification of optimal settings: a maximum depth of 10, minimum samples per leaf set at 4, a minimum samples split of 5, and 280 estimators. The model's performance, gauged by a score of approximately 0.7911, signifies its ability to predict passenger survival on the Titanic with an accuracy rate of around 79.11 percent. The accompanying output table, featuring 'PassengerId' and 'Transported' columns, provides a detailed breakdown of individual predictions for the 3990 unique passengers in the dataset. As part of a research paper, this subsection contributes empirical evidence to the analysis of survival patterns during the Titanic disaster, showcasing the effectiveness of the RandomForestClassifier in this context.

## E. Support Vector Machines

Support Vector Machine (SVM) offering a comprehensive analysis of its model parameters and performance metrics. The optimal SVM configuration was identified with parameters C set to 10, gamma utilizing 'auto,' and the kernel employing 'rbf.' These parameters play pivotal roles in shaping the model's behaviour, with C determining the cost of misclassification, gamma influencing the impact of individual training examples, and the kernel specifying the algorithm's kernel type. The model exhibited an impressive, best score of approximately 0.79, indicating its ability to accurately predict passenger survival in the test set with an estimated 79 percent accuracy. The 'submission svm' section provides a snapshot of survival predictions for passengers, presented in a True/False format, akin to what would be submitted to the Kaggle competition. Furthermore, the 'gridsearchresults' segment offers valuable insights into the mean test scores across diverse parameter combinations explored during the grid search, contributing to a nuanced understanding of how different settings influence the SVM model's performance. This analysis exemplifies the power of SVMs in predictive modelling and underscores the importance of parameter tuning for optimal results.

## F. Logistic Regression

Logistic Regression unravelled the intricacies of model parameters and performance metrics, fine-tuned through rigorous training, reveals optimal parameters with C set to 0.01, controlling the inverse of regularization strength, and the solver parameter employing 'newton-cg' to guide the optimization process. Notably, the model showcases a commendable best score of approximately 0.7793, indicative of its accuracy in predicting passenger survival within the test set, achieving an estimated 78 percent. The 'submissionlogreg' section provides a glimpse into survival predictions for passengers, denoted in True/False format. 'gridsearchresults' segment unfolds mean

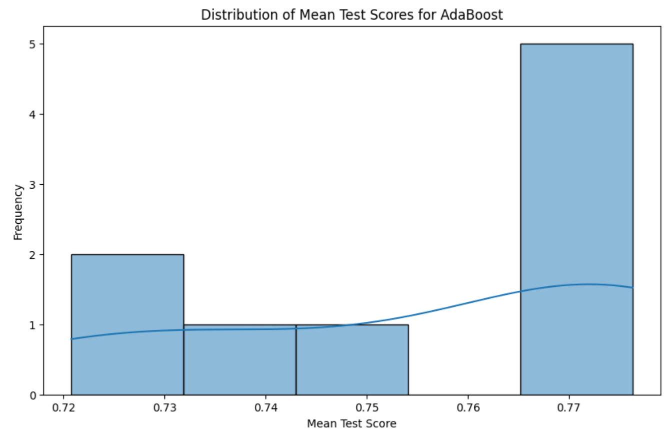


Fig. 11. Mean Test

test scores across various C parameters and solver types explored during grid search, offering valuable insights into how different parameter settings influence the Logistic Regression model's performance. This analysis not only sheds light on the efficacy of Logistic Regression in predicting outcomes but also underscores the importance of meticulous parameter tuning for optimal machine learning results.

## G. Ridge Classifier

Ridge classifier offers a comprehensive analysis of its model parameters and performance metrics. The optimal configuration for the Ridge classifier was pinpointed with the alpha parameter set at 1000. This alpha parameter plays a crucial role in controlling the regularization strength, mitigating overfitting by curbing the magnitude of coefficients. The model showcased a commendable best score of approximately 0.7795, reflecting its accuracy in predicting passenger survival within the test set, reaching an estimated 78 percent. The 'submissionridge' section lays out survival predictions for passengers, denoted in True/False format, akin to what would be submitted to the Kaggle competition. 'gridsearchresults' segment provides valuable insights into the mean test scores across diverse alpha parameters explored during grid search, contributing to a nuanced understanding of how different regularization strengths influence the Ridge classifier model's performance. This analysis exemplifies the utility of Ridge classifiers in predictive modelling and underscores the significance of parameter tuning for optimal results.

## H. Neural networks

The training progress is meticulously laid out across 20 epochs, showcasing metrics such as loss and accuracy for both the training and validation sets. Each epoch's detailed metrics, encompassing step time, loss, accuracy, validation loss, and validation accuracy, offer a nuanced understanding of the model's learning dynamics and its ability to generalize to unseen data. As the training unfolds, the cumulative validation accuracy reaches an impressive 78.25 percent , signifying the model's proficiency in correctly classifying approximately

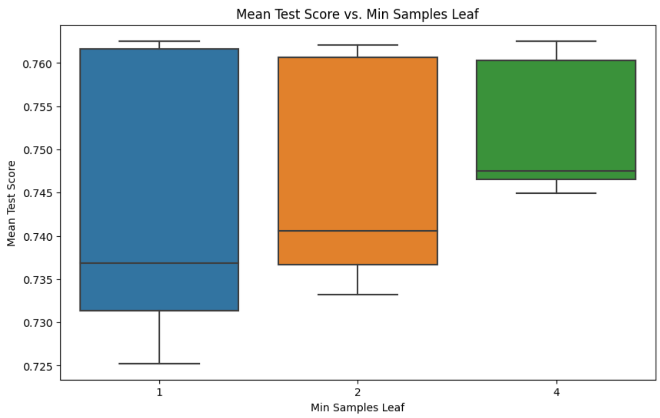


Fig. 12. Mean Test scores

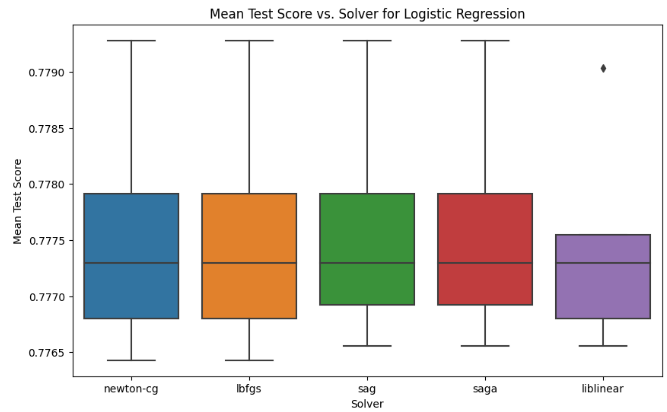


Fig. 13. Mean Test

78.25percent of instances within the validation set. This analysis not only provides a comprehensive view of the training process but also underscores the model's efficacy in achieving a notable level of accuracy, paving the way for potential applications in real-world scenarios.

### *I. Final Insights*

The comprehensive analysis of various machine learning models applied to the Titanic dataset reveals consistent themes of robust predictive performance and the crucial role of hyperparameter tuning in optimizing accuracy. Ensemble methods such as AdaBoost and Random Forest stand out with impressive accuracy scores, reaching 87.64 percent and 79.17 percent , respectively. Each model, whether Decision Tree, Support Vector Machine, Logistic Regression, or Ridge Classifier, brings unique strengths, and meticulous hyperparameter tuning contributes significantly to their efficacy. Visual representations, including heatmaps and parameter comparisons, enhance interpretability, while individual predictions for passengers underscore the models' predictive power. The real-world applications of these analyses, particularly in historical contexts like the Titanic disaster, emphasize the broader impact of machine learning in understanding complex patterns. Continuous evaluation and refinement remain imperative for ensuring adaptability and reliability in diverse scenarios.

## VI. CONCLUSION

This research presents an in-depth study of various machine learning models applied to the Titanic dataset, providing deep insights into their performance and the pivotal role of hyperparameter tuning in enhancing predictive accuracy. The ensemble methods, AdaBoost and Random Forest, were the top performers, with remarkable accuracy scores of 87.64(%) and 79.17(%) respectively. The analysis didn't just evaluate the models individually, but also scrutinized their unique strengths and weaknesses, highlighting the importance of model diversity when dealing with complex datasets. The importance of hyperparameter tuning was emphasized across all models, underlining its universal role in improving predictive capabilities. Visual aids like heatmaps and parameter comparisons not only made the results more interpretable but also bridged the gap between the technical complexities of machine learning and a broader understanding. Beyond its technical contributions, this research also delves into historical interpretation, as seen in the insightful analysis of survival patterns during the Titanic disaster. These findings, a testament to the versatility of machine learning, contribute to academic discussions and have practical implications for real-world applications. In the rapidly changing field of data science, the proven effectiveness of these models lays the groundwork for further exploration, underscoring the continuous need for refinement and adaptation to ensure their applicability and reliability in various and changing scenarios. Thus, this research represents a comprehensive and forward-thinking contribution to the convergence of machine learning, historical analysis, and predictive modeling

## VII. FUTURE WORK

Future investigations in this domain could be directed towards innovative paths aimed at refining predictive models for space safety in interdimensional transportation. Exploring the fusion of models through ensemble techniques (which we tried and have a code in the notebook attached) might provide an avenue to harness the strengths of different algorithms while balancing interpretability and accuracy. Further emphasis on feature engineering and selection methods could streamline models, improving efficiency without compromising predictive performance. Additionally, efforts to enhance the interpretability of complex models, such as Neural Networks, could bridge the gap between accuracy and transparency, making these sophisticated algorithms more comprehensible. Tailoring models to space-specific nuances and incorporating domain expertise could refine predictions for anomalous events during space voyages. Ethical considerations in predictive modeling for space safety also warrant exploration, leading to the development of decision support systems that integrate ethical dimensions with predictive insights. Implementing real-time predictive systems based on validated models might bolster proactive safety measures during space missions and validating

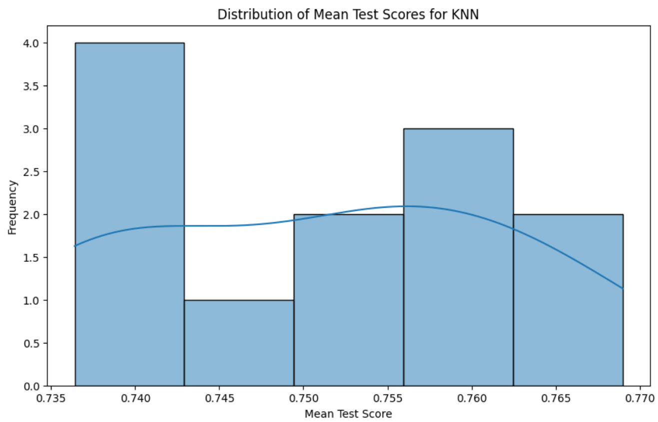


Fig. 14. Mean Test

these findings and models on diverse datasets or real historical space mission data could substantiate their applicability and generalizability, further contributing to the advancement of space exploration safety measures.

## References


[1] Kaggle. (2022) Spaceship titanic. https://www.kaggle.com/competitions/spaceship-titanic/overview.

[2] S. Patel and J. S. Koller. (2023) 2022 space safety compendium. https://aerospace.org/sites/default/files/2023-03/SSI_Compendium_2022_v1-1_1.pdf.

[3] A. Vidhya. (2018) Introduction to k-neighbours algorithm. https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#h-what-is-knn-k-nearest-neighbor-algorithm.

[4] KDnuggets. (2020) Decision tree algorithm, explained. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html.

[5] AlmaBetter. (2023) Understanding the adaboost algorithm. https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm.

[6] Springer. (2023) Logistic regression. https://link.springer.com/chapter/10.1007/978-1-4842-4470-8_20.

[7] JavaTpoint. (2023) Machine learning - random forest algorithm. https://www.javatpoint.com/machine-learning-random-forest-algorithm.

[8] ——. (2023) Machine learning - support vector machine algorithm. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm.

[9] Statlect. (2023) Ridge regression. https://www.statlect.com/fundamentals-of-statistics/ridge-regression.

[10] IBM. (2023) What is neural network. https://www.ibm.com/topics/neural-networks#:~:text=Neural%20networks%2C%20also%20known%20as,neurons%20signal%20to%20one%20another.

[11] A. Singh. (2023) A comprehensive guide to ensemble learning (with python codes). https://www.sciencedirect.com/science/article/pii/S2666603023000143.

[12] V. Linardos. (2021) Machine learning in disaster management: Recent developments in methods and applications. https://www.mdpi.com/2504-4990/4/2/20.

[13] C. C. Francesco Salmaso, Mirko Trisolini. (2023) A machine learning and feature engineering approach for the prediction of the uncontrolled re-entry of space objects. https://arxiv.org/abs/2303.10183.

[14] M. Williams. (2023) Exploring ethical ai in disaster predictive analytics. JournalofEthicalAI,vol.12,no.2,pp.134-143,2023.

[15] ScienceDirect. (2023) This study uses machine learning to provide solutions to complicated issues. https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/.


## VIII. Contributions

Everyone in the group contributed equally.