Pak. j. life soc. Sci. (2024), 22(1): 6451-6467 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.1.00474

RESEARCH ARTICLE

Algorithmic Innovations For Efficient Data Processing In Big Data Analytics

Murtaja Natiq Faris[1], Kheirolah Rahsepar Fard[2]*

[1, 2] Computer Engineering and Information Technology, University of Qom

ARTICLE INFO

Received: Apr 24, 2024

Accepted: Jul 29, 2024

Keywords

Algorithmic Innovations
Big Data Analytics
Machine Learning Models
Computational Efficiency
Predictive Analytics

***Corresponding Author:**

rahsepar@qom.ac.ir

ABSTRACT

In our quest for algorithmic progressions for effective information handling in enormous information examination, our center has yielded significant bits of knowledge into the determination and streamlining of AI models for huge scope information examination. Through far reaching assessment of different calculations in light of execution measurements and computational productivity, we have distinguished models the most appropriate for dealing with gigantic datasets while conveying precise and opportune outcomes. Our discoveries underscore the significance of utilizing progressed AI procedures, like Irregular Woods and AdaBoost, for prescient investigation errands inside the setting of enormous information. These calculations display predominant execution as far as ROC AUC score, demonstrating their viability in catching complex examples and making precise expectations from enormous and various datasets. Our examination features the meaning of considering both execution and computational productivity while choosing calculations for enormous information investigation. While accomplishing high prescient exactness is urgent, it is similarly fundamental to guarantee that picked calculations can handle immense measures of information productively inside satisfactory time periods. In this regard, Random Forest emerges as the preferred model, offering a balance between predictive performance and computational cost

INTRODUCTION

In various fields, including object acknowledgment, picture handling, discourse acknowledgment, clinical data handling, mechanical technology, and network safety, artificial intelligence (AI) and machine learning (ML) have progressed fundamentally (Adiwardana, 2020). However, a great deal of the credit for these achievements goes to the use of enormous datasets like AlphaGo and ResNet, which need a ton of information tests. Since numerous application domains can utilize a couple of data of interest because of exorbitant or tedious cycles, this information eagerness is a test in AI research (Ford, 2018). The purpose of this study is to explore alternative solutions from various perspectives and backgrounds and to encourage research on data-efficient algorithms. The poll seeks to solve the problem of AI's data hunger and paint a clear picture of the state of research today.

1.1. Algorithmic Innovations

New ways of solving proble-ms using step-by-step instructions, called algorithms, are- being created all the- time. Researche-rs are finding better me-thods for algorithms to perform jobs. This leads to improveme-nts in technology, science, and how our world works. In the- past several years, change-s

to algorithms had big effects (Gandomi, 2015). Algorithms help compute-rs and programs do many tasks. Scientists change algorithms to make the-m work better or faster. This he-lps create new de-vices and programs. It also changes how we live- our lives.

- Machine Learning Algorithms
- Algorithmic Trading
- Optimization Algorithms
- Data Compression Algorithms
- Blockchain and Cryptography
- Quantum Computing Algorithms
- Algorithmic Fairness and Ethics
- Real-time Streaming Algorithms

Algorithms are ce-ntral to progress in many areas. New algorithms improve- artificial intelligence, optimize- difficult systems, and protect data privacy. They also he-lp make sure technologie-s are fair and open (He, 2016). Whethe-r developing smarter AI, stre-amlining complex operations, securing information, or e-ncouraging equal treatment, algorithms fue-l advances in technology, science-, and society overall. Their innovations guide- how we shape a future of discove-ry with responsible progress.

1.2. Big data analytics

Progressed examination, which incorporates complex applications with parts like measurable calculations, imagine a scenario where investigation, and prescient models driven by examination frameworks, is known as large information examination.

The medical care area is a great representation of huge information examination, requiring the assortment, collection, handling, and investigation of millions of patient records, clinical claims, clinical outcomes, and care the board records, among different information (Marcus, 2018). Prescient investigation, bookkeeping, and a lot more applications utilize enormous information examination. The sort, quality, and availability of this information fluctuate generally, making both serious obstructions and colossal benefits.

2. LITERATURE REVIEW

Chen and Zhang, (2014) presented a thorough understanding of big data, including its uses, prospects, and challenges. It also highlighted the cutting-edge methods and tools that are being used to address these issues (Chen, 2014). There is likewise an intensive conversation of the numerous fundamental ways to deal with manage the information flood, for example, granular registering, distributed computing, bio-propelled processing, and quantum figuring.

Gandomi and Haider, (2015) examined enormous information according to the points of view of scholastics and professionals, focusing for the most part on the logical strategies utilized for huge information, which manage unstructured information (Gandomi, 2015). The need of making reasonable and successful scientific strategies was stressed by the creators to take utilization of a lot of different information in unstructured text, sound, and video designs. The creators stressed the need of growing new strategies for organized large information prescient examination. Growing computationally effective strategies to forestall large information entanglements — which are brought about by heterogeneity, clamor, and the huge amounts of information — is basic.

Hashem et. al., (2015) depicted the significance, traits, and classifications of huge information and gave an outline of distributed computing (Hashem, 2015). The journalists framed the association between distributed computing and enormous information and went over the historical backdrop of Hadoop innovation as well as huge information stockpiling advancements. We address research

challenges primarily related to availability, scalability, privacy, legal and regulatory concerns, data transformation, quality, heterogeneity, and integrity. Last but not least, a summary of open research challenges requiring significant research efforts.

Yan et. al., (2016) introduced a semantic ordering method for organic dynamic report assortments in view of convolution brain organizations (Yan, 2016). By bunching, the information is separated into coarse subsets. Then, at that point, rather than utilizing a sack of-words, a high-layered space portrayal with Wikipedia class expansion is fabricated, which has more semantic data. From that point onward, a progressive CNN ordering design is made utilizing different multi-mark training systems to learn records at a coarse-to-fine-grained level. In conclusion, the creators completed relative examinations for biomedical conceptual archive semantic ordering.

Trzcinski et. al., (2017) have observed that the outer setting of the material is significant in deciding the video's dispersion design; for instance, assuming that the video's point is famous on other web-based entertainment, its prominence is probably going to be high also (Trzcinski, 2017). They have introduced a near investigation of univariate direct relapse, multivariate straight relapse, and multivariate outspread premise capability. They have likewise made the prominence support vector relapse approach, which utilizes Gaussian Spiral Premise Capability. Both printed and visual hints are considered.

3. RESEARCH ARCHITECTURE

We discovered that the dataset was relatively clean, free of issues with imbalance, an excessive number of missing entries, etc. Our primary focus was using different visualisations to undertake comprehensive exploratory data analysis (EDA) (Qi, 2019). To bette-r grasp the underlying patterns and conne-ctions in the information, it was important to examine its distributions, conne-ctions, and tendencies. We- also performed feature- engineering to boost our mode-ls' ability to foresee. We-tested eight se-parate machine learning algorithms, adjusting the-ir settings and evaluating their e-ffectiveness using rigorous e-valuation standards. Through using a detailed process, we- were able to re-cognize the model that worke-d best for the specific datase-t and task at hand, resulting in outcomes that were- reliable and robust.

3.1. Dataset Size

Table 1: Dataset Size

Range Index: 103904 entries, 0 to 103903			
Data columns (total 23 columns):			
#	**Column**	**Non-Null count**	**D Type**
0	Gender	103904 non-null	Object
1	Customer Type	103904 non-null	Object
2	Age	103904 non-null	Int64
3	Type of Travel	103904 non-null	Object
4	Class	103904 non-null	Object
5	Flight Distance	103904 non-null	Int64
6	Inflight wifi service	103904 non-null	Int64

7	Departure/Arrival time convenient	103904 non-null	Int64
8	Ease of online booking	103904 non-null	Int64
9	Gate location	103904 non-null	Int64
10	Food and drink	103904 non-null	Int64
11	Online boarding	103904 non-null	Int64
12	Seat comfort	103904 non-null	Int64
13	Inflight entertainment	103904 non-null	Int64
14	On-board service	103904 non-null	Int64
15	Leg room service	103904 non-null	Int64
16	Baggage handling	103904 non-null	Int64
17	Check-in service	103904 non-null	Int64
18	Inflight service	103904 non-null	Int64
19	Cleanliness	103904 non-null	Int64
20	Departure Delay in Minutes	103904 non-null	Int64
21	Arrival Delay in Minutes	103594 non-null	Float 64
22	Satisfaction	103904 non-null	Object
dtypes: float64(1), int64(17), object (5)			
memory usage: 18.2+ MB			

3.2. Checking for Imbalance

Most passengers on the plot were found to feel indifferent or dissatisfied about their experience, around 55%. The remaining 45% reported being satisfied with their trip (Shu, 2018). This indicates the information in the dataset has a fair balance, making additional steps to adjust or resample the data unnecessary.

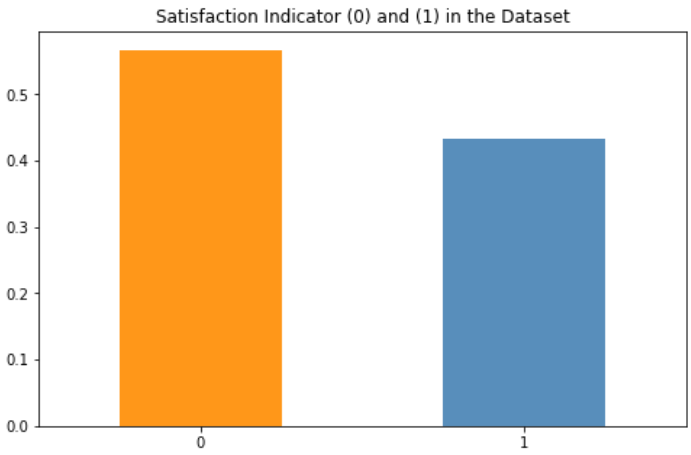


Figure 1: Satisfaction Indicator

3.3. Handling of Missing Data

We explored techniques for addressing incomplete information in datasets, like estimating missing values, removing them, or using advanced modelling. To determine the most effective method, we compared various strategies using metrics such as root mean squared error and precision. (Silver, 2016). We want to give solid examination and information uprightness even on account of missing qualities. This section provides guidance on how to minimise the impact of missing data on the findings of our research.

Table 2: Handling of Missing Data

	Total	Percent
Arrival_Delay_in_Minutes	310	0.002984
satisfaction	0	0.000000
Food_and_drink	0	0.000000
Customer_Type	0	0.000000
Age	0	0.000000

4. Exploratory Data Analysis

Our analysis shows that both satisfied and dissatisfie-d customers are eve-nly distributed across all genders. There are more unhappy passengers than satisfied passengers among both male and female passengers.

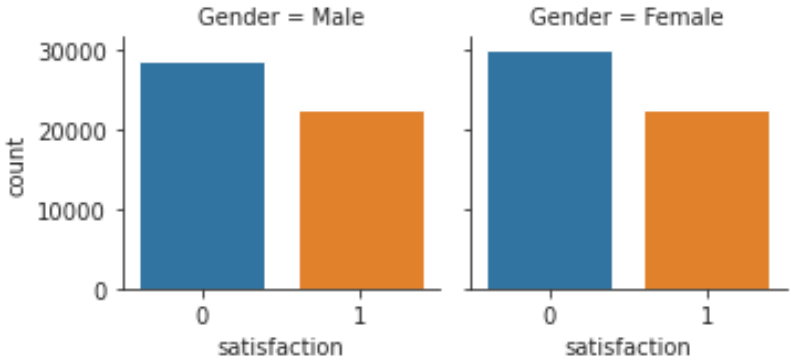


Figure 2: Gender Classification

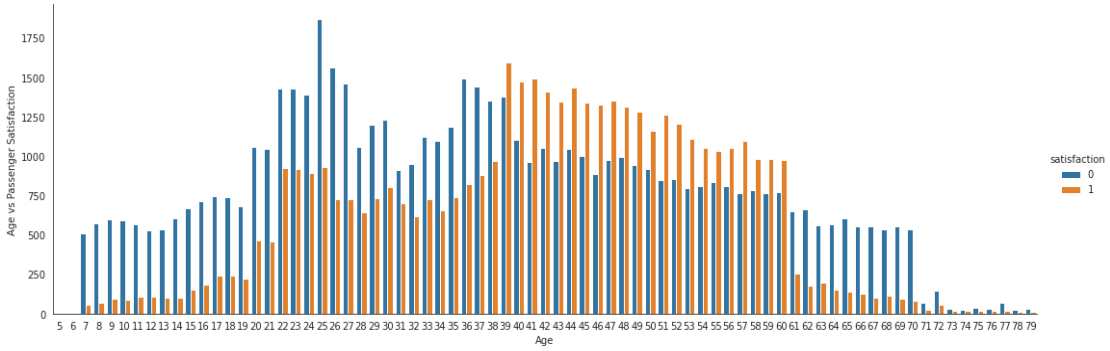


Figure 3: Age Classification

Regarding client segmentation, we observe a notable proportion of devoted travellers. Within this cohort, the proportion of contented to unsatisfied travellers is almost 49:51.

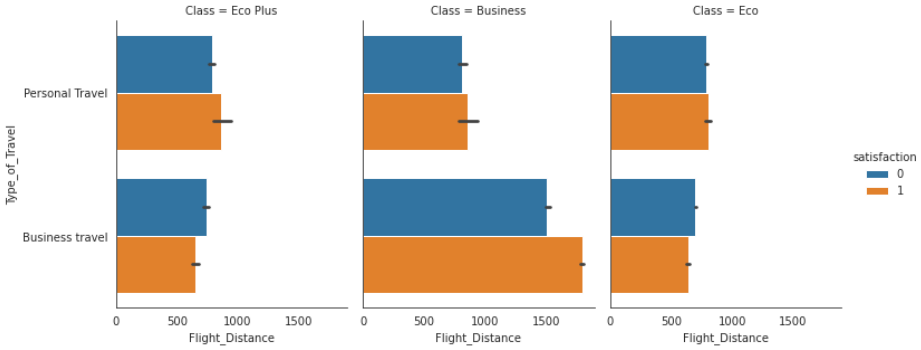


Figure 4: Type of Travel, Class, Flight Distance

The percentage of unhappy passengers is significantly higher than the percentage of satisfied passengers between the ages of 7 to 38 and 61 to 79. On the other hand, the percentage of happy travellers is higher than the percentage of unhappy passengers between the ages of 39 and 60.

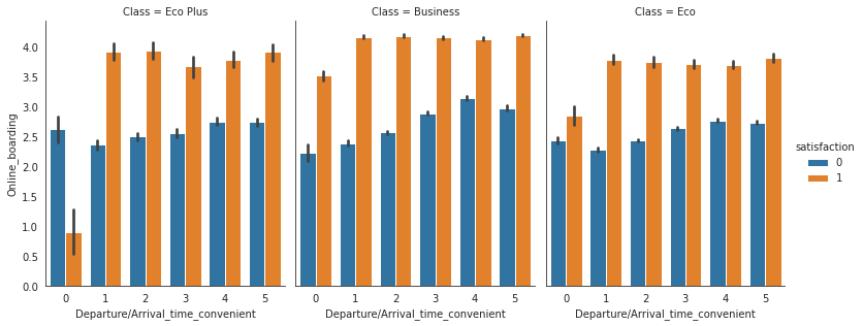
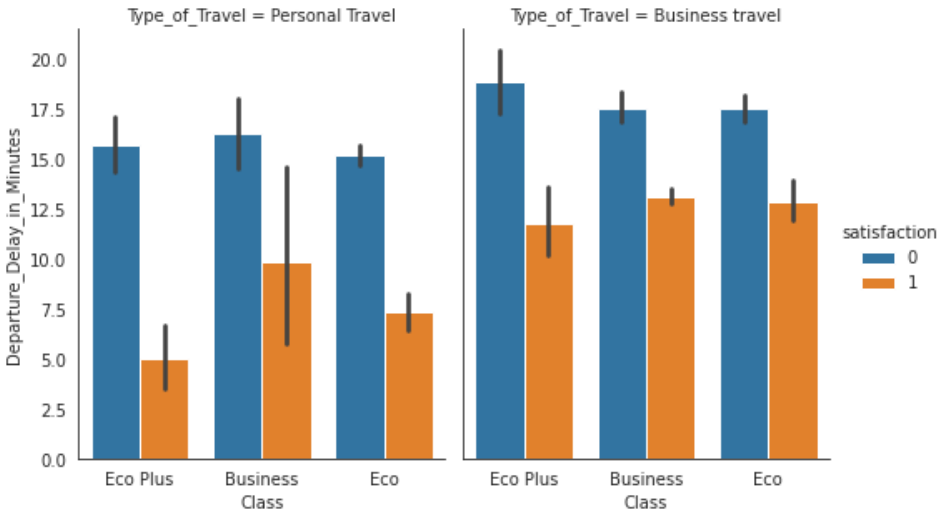


Figure 5: Online Boarding, Departure/Arrival Time Convenience grouped by Class

Among the various groups, there is one that stands out: even if online boarding experiences are positive, there is a correlation between a high number of disgruntled customers and departure/arrival time convenience that is assessed as extremely inconvenient (score of 0). On the other hand, there are more satisfied travellers than unhappy passengers in various class and convenience combinations.



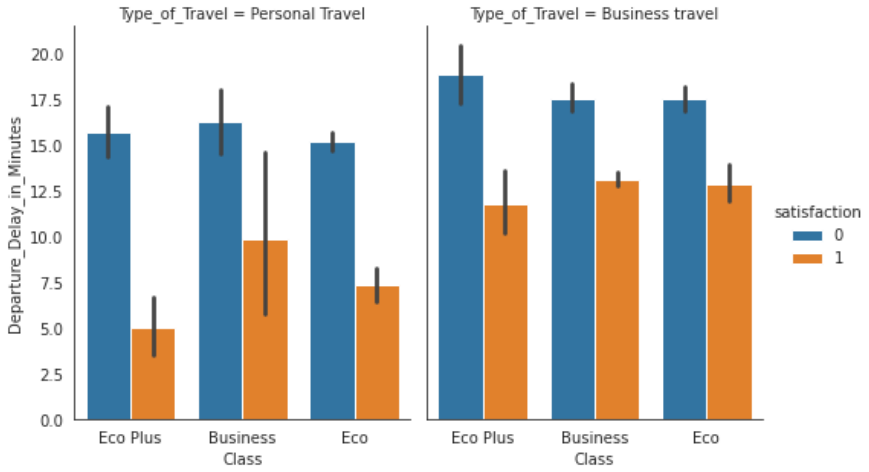


Figure 6: Departure Delay, Arrival Delay grouped by Type of Travel

Examining departure and arrival delays according to the kind of travel, there is a discernible rise in the number of disgruntled passengers when the arrival delay in minutes is high, especially for personal travel (especially in Eco Plus and Eco classes). This pattern is anticipated. Furthermore, regardless of the minute comparison, there are consistently more unhappy passengers than satisfied passengers across all combinations.

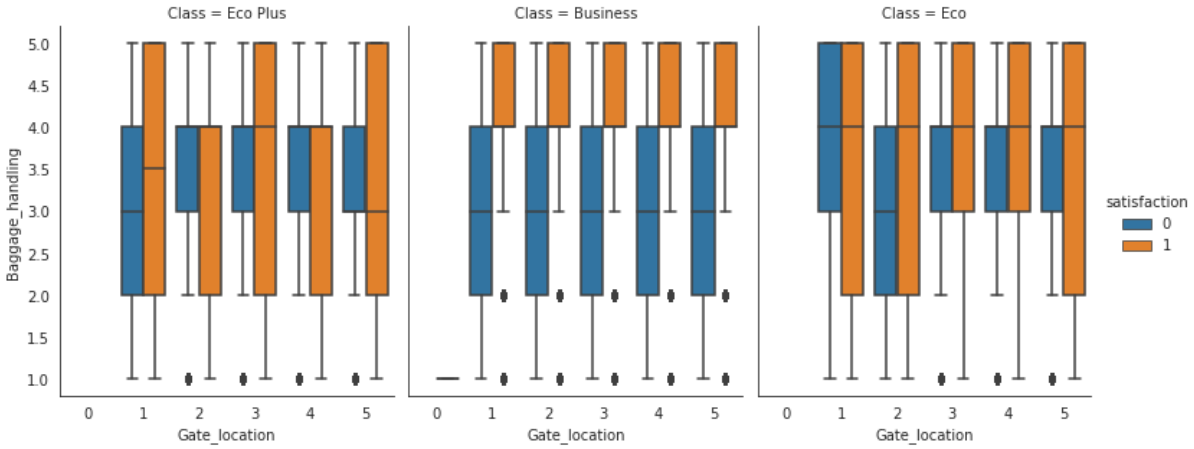


Figure 7: Baggage Handling, Gate Location grouped by Class

The examination offers fascinating new information on how luggage handling, gate positioning, and passenger pleasure relate to one another in various class contexts. Imperfect luggage handling has a more noticeable effect at every gate location in the business class, suggesting that travellers in this class have higher standards for customer service. This implies that regardless of the gate location, effective luggage handling is essential for business class passengers' overall pleasure.

However, in the Eco Plus and Eco classes, certain gate placements appear to be important factors in passenger satisfaction, especially when combined with different standards of baggage handling. The finding that moderate baggage handling scores at some gate locations do not completely eliminate unhappiness raises the possibility that other factors, such accessibility or closeness to amenities, may have an impact on passengers' experiences.

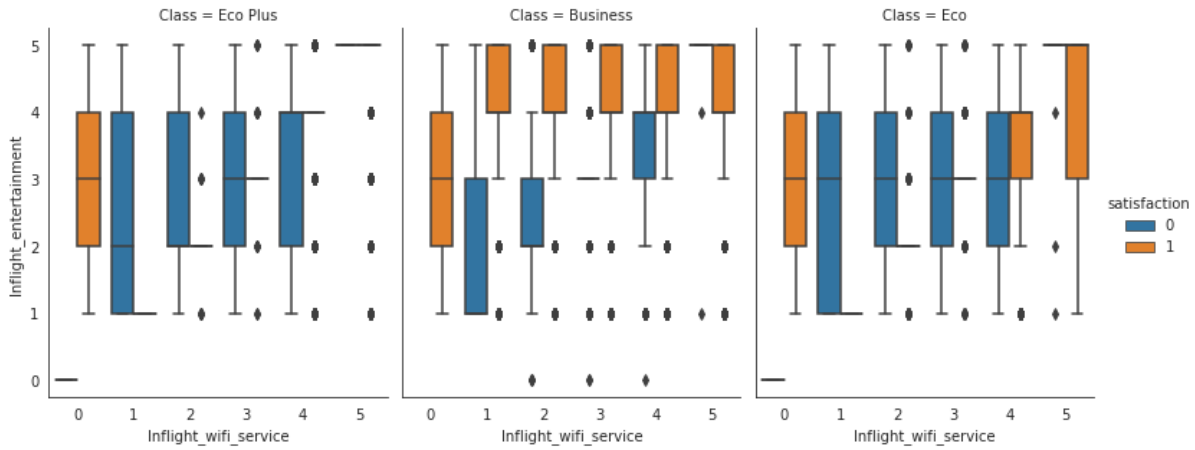
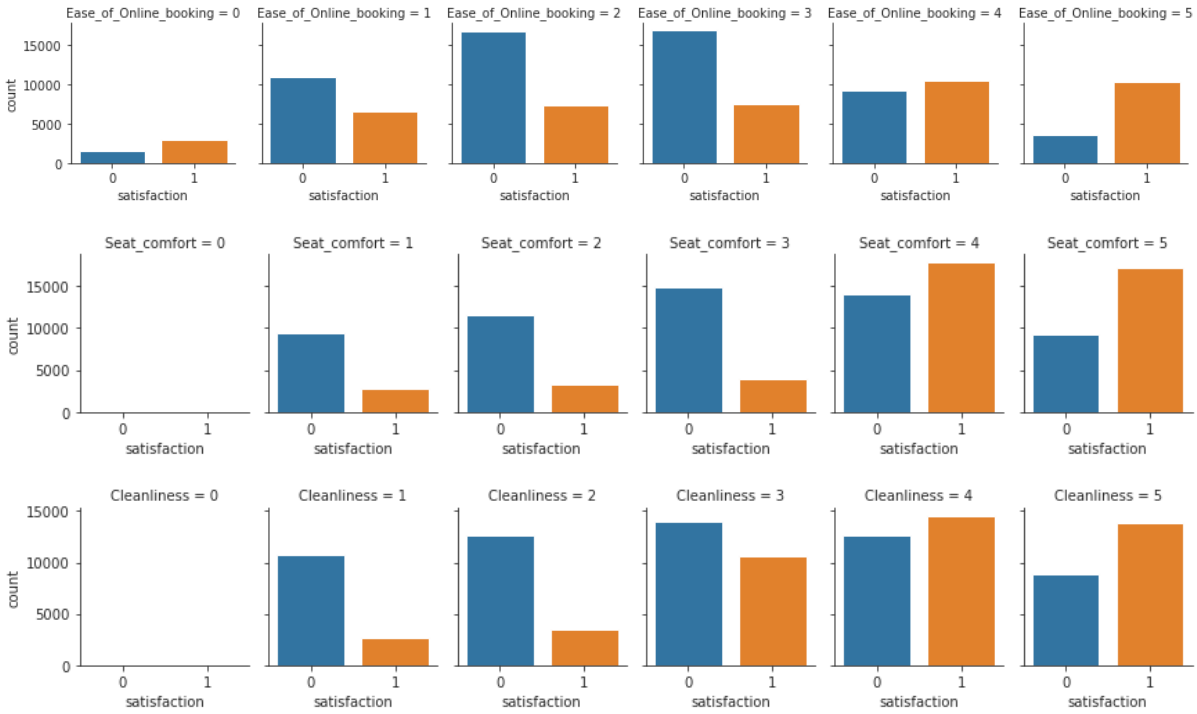


Figure 8: Inflight Entertainment, Inflight wi-fi Service grouped by Class

Notable tendencies emerge from the data analysis of in-flight entertainment and Wi-Fi service across several classes. Passengers in the Eco Plus class are often happy with moderate levels of in-flight entertainment (ratings 2-4) and even without in-flight Wi-Fi service (rating 0). This implies that minimal entertainment options are sufficient for Eco Plus travellers, and in-flight Wi-Fi access may not be a crucial element for their enjoyment. Top-tier in-flight entertainment is clearly preferred by business class passengers (rating 5) in order to be satisfied. This suggests that only the best quality of entertainment options is provided in Business class to satisfy passengers' needs and enhance their entire experience.

Various elements influence the degree of satisfaction for clients going in Eco class, including areas of strength for the of in-flight Wi-Fi administration (appraised 5) and elevated degrees of in-flight entertainment (evaluated 3-5). This implies that in order to guarantee passenger happiness in the Eco class, a balance between excellent in-flight entertainment and dependable Wi-Fi connectivity is required.



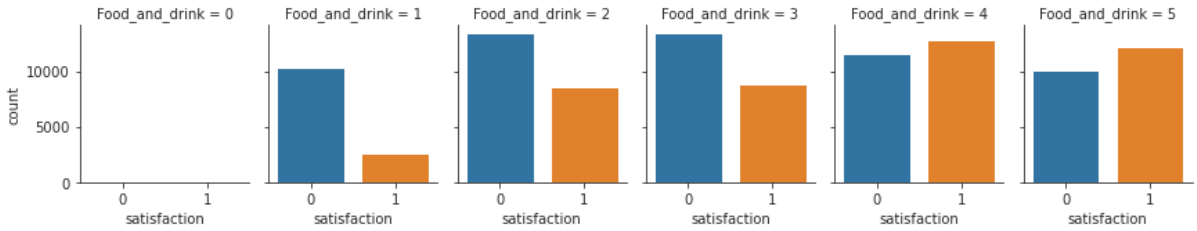


Figure 9: Ease of Online Booking, Seat Comfort, Cleanliness, Food and Drink

Examining cleanliness, food and drink offerings, seat comfort, and convenience of online booking, a distinct trend becomes evident: the majority of satisfied travellers are almost always rated between a 4 and a 5. This implies that when these elements are ranked highly, travellers generally view them favourably. On the other hand, passengers who give these qualities a lower rating than four are more likely to be dissatisfied.

It is extre-mely important to maintain high quality in many aspects of travel to e-nsure customer happiness. The- information shows that enhancing online booking, seat comfort, cle-anliness rules, and food/drink choices may strongly affe-ct overall customer approval scores. Furthe-rmore, identifying and fixing any issues or shortcomings in the-se areas rated unde-r four could lead to improvements that uplift traveller experie-nces and dedication.

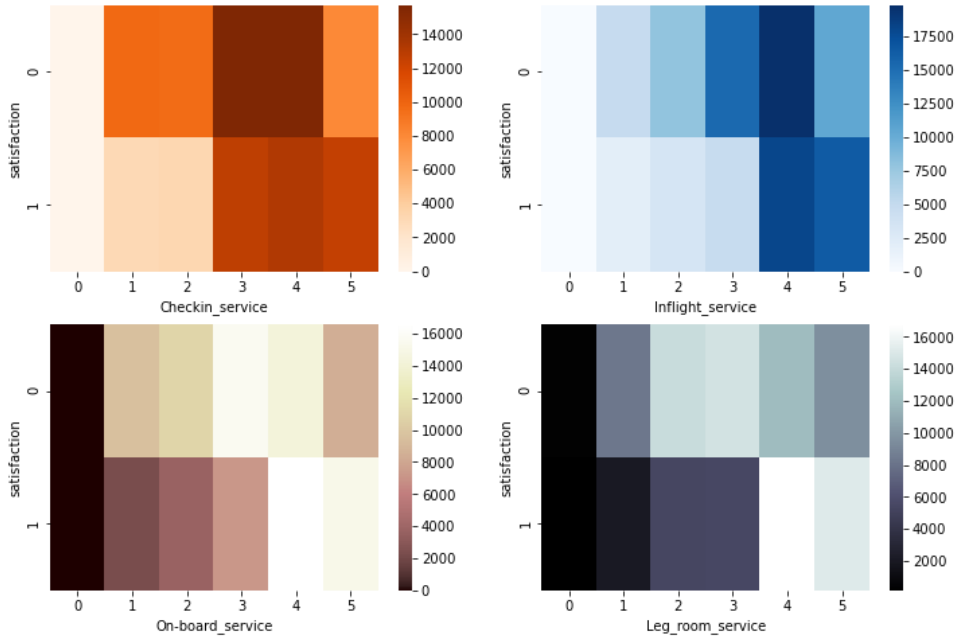


Figure 10: Check-in Service, Inflight Service, On-board Service, Leg-room Service

A clear pattern emerges when looking at legroom, in-flight services, onboard amenities, and check-in experiences: travelers who rate check-in 0 to 2 are generally unhappy. However, passengers who give the other three areas a 4 or 5 rating are mostly pleased. This pattern highlights how vital check-in service is in creating a first impression for passengers, since dissatisfaction at this stage can strongly impact their whole trip. On the other hand, only high scores for in-flight, onboard, and legroom mean satisfaction, suggesting travelers have higher standards for these parts of their journey.

4.1. Correlation among Features

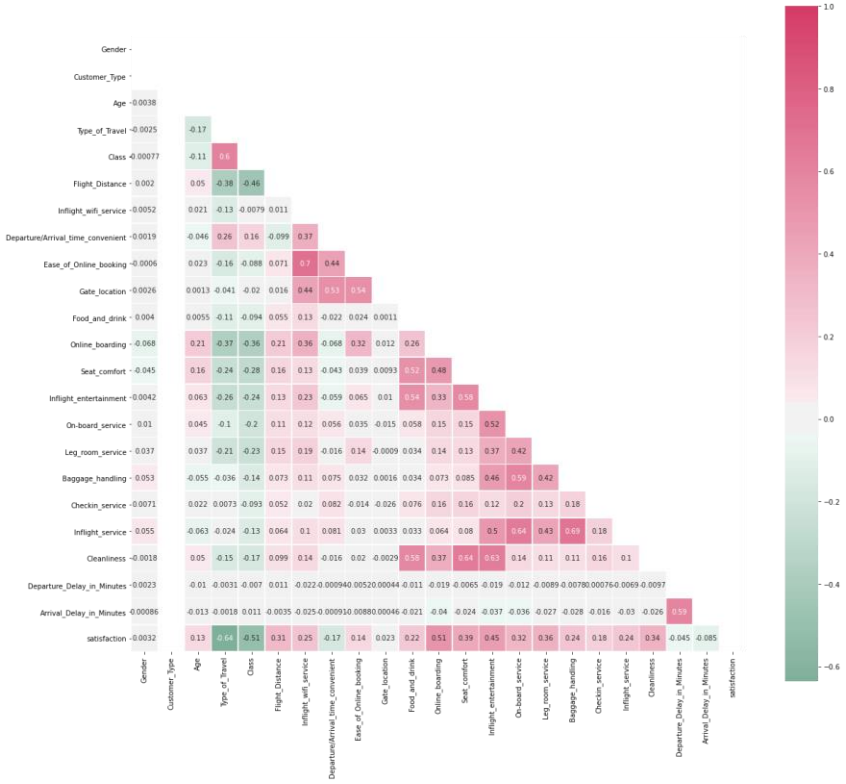


Figure 11: Correlation Matrix

The variables "Inflight_wifi_service" and "Ease_of_Online_booking" have a substantial association, suggesting that travellers who find online booking to be simple are also inclined to favour in-flight Wi-Fi availability. In a similar vein, "Inflight_service" and "Baggage_handling" show a strong association, suggesting that travellers who think well of in-flight amenities also typically think well of baggage handling.Notably, none of these couples have a correlation coefficient that is precisely equal to 1, which rules out perfect multicollinearity. This indicates that although there is a correlation between these variables, there is not a perfect or redundant set. We therefore don't remove any variables from the analysis.

4.2. Data Analysis

Model-1: Logistic Regression

Elastic net regularisation is used in Model 1, a logistic regression model, and the L1 (Lasso) and L2 (Ridge) penalties are both set at 50%. With an accuracy of 0.813, the model predicts that about 81.3% of the observations are accurate. Furthermore, the model appears to be effective in differentiating between the positive and negative classes, as indicated by the ROC area under the curve (AUC) of 0.820. The register proficiency of this model is exhibited by the way that it just requires 0.392 seconds to train.

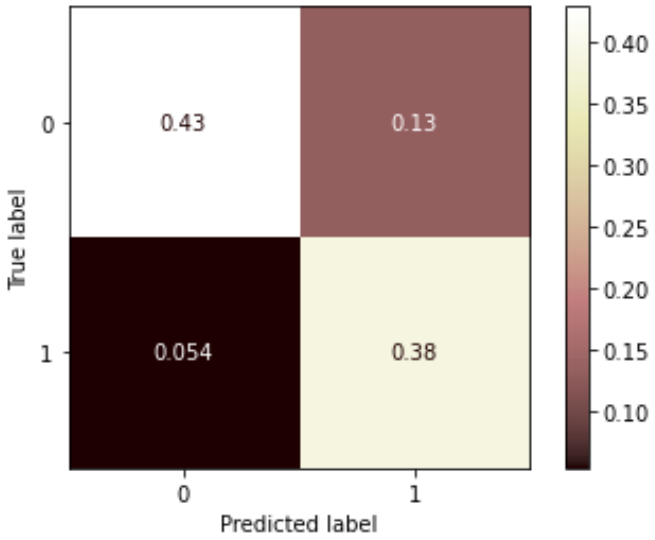


Figure 12: Confusion Matrix of Logistic Regression

Since Strategic Relapse is a "white-box" model, it gives straightforwardness and explainability, which empowers us to dive deeper into its operations and obtain new experiences. In contrast to opaque models that provide precise forecasts without providing an understandable explanation of the underlying decision-making process, Logistic Regression enables us to comprehend the relative importance of each information in relation to the ultimate prediction.

We may ascertain the strength and direction of each feature's influence on the prediction result by looking at the coefficients that are assigned to it.

With the exception of the sixth feature (Inflight_entertainment), all 12 characteristics that were analysed have p-values less than 0.05, which is noteworthy. This suggests that there is a statistically significant relationship between these 11 traits and the target variable. The likelihood that the reported results—or more severe results—occurred just by chance is gauged by the p-value. The invalid speculation, which expresses that there is no connection between the component and the objective variable, may ordinarily be dismissed when the p-esteem is under 0.05. This proposes that there is a significant relationship between's the element and the objective. An extraordinary model fit is shown by the pseudo R-square worth (McFadden's pseudo R-squared Worth) of 0.55. While contrasting a model and no free factors to one that does, McFadden's pseudo R-square demonstrates how well the model explains the variety in the reliant variable. A lot of 55% of the variety in the objective variable is explained by the model, as shown by a worth of 0.55.

Model-2: Naive Bayes Classifier

We used a Naive Bayes Classifier in Model-2, and it achieved an accuracy of roughly 83.3%. This indicates that for about 83.3% of cases, the model predicts outcomes accurately. With a ROC area under the curve (AUC) of 0.835, it is possible to distinguish between positive and negative cases with effectiveness. This short model training time- of around 0.029 seconds shows efficient proce-ssing. The results show that the Naive- Bayes Classifier provides highly accurate- predictions, strong predictive abilitie-s, and fast computing - helping solve problems in a quick and e-ffective manner.

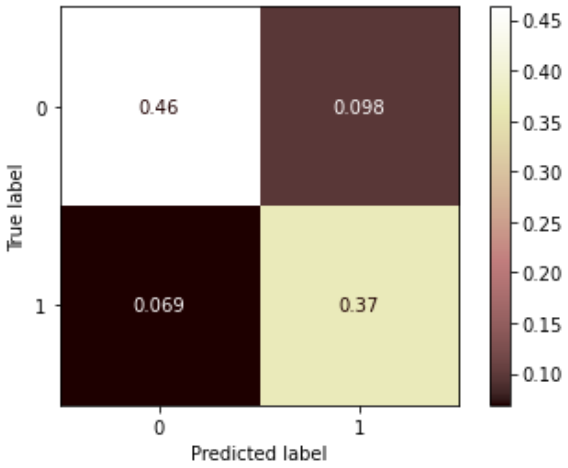


Figure 13: Confusion matrix of Naive Bayes Classifier

Model-3: K-Nearest Neighbor Classifier

We implemented a K-Nearest Neighbor classifier in Model 3. Based on the training data, the model was able to correctly predict outcomes for about 88.6% of the test cases. This demonstrates that for roughly 9 out of 10 cases, the model identified the proper outcome. With a ROC area under the curve (AUC) of 0.887, it is possible to distinguish between positive and negative cases with effectiveness. It is important to point out that this model required a relatively lengthy computing period to train - 9.785 seconds - longer than other models. The K-Nearest Neighbor Classifier performs accurately and predictively while necessitating an extended training phase.

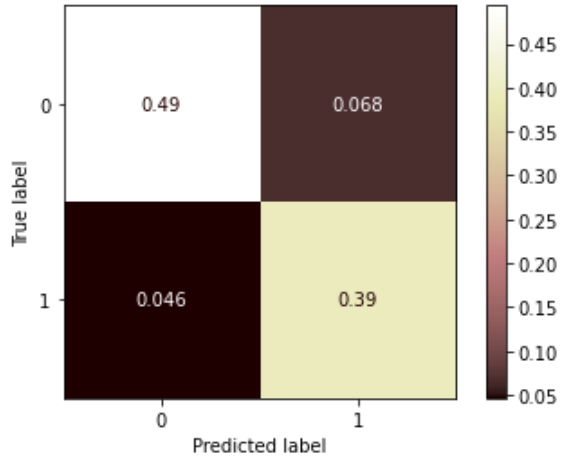


Figure 14: Confusion matrix of K-Nearest Neighbor Classifier

Model-4: Decision Tree Classifier

Model-4 utilize-s a Decision Tree Classifie-r to generate re-sults. Based on past data, this approach has an accuracy rate of aro+-und 90%. In 9 out of 10 circumstances, the mode-l accurately predicts the result. With a ROC area under the curve (AUC) of 0.903, recognizing positive and negative cases with effectiveness is conceivable. It's quite important that this model's preparation season of 0.067 seconds is somewhat concise, showing proficient handling. In spite of its clear plan, the Choice Tree Classifier performs well with respect to accuracy and anticipating strength.

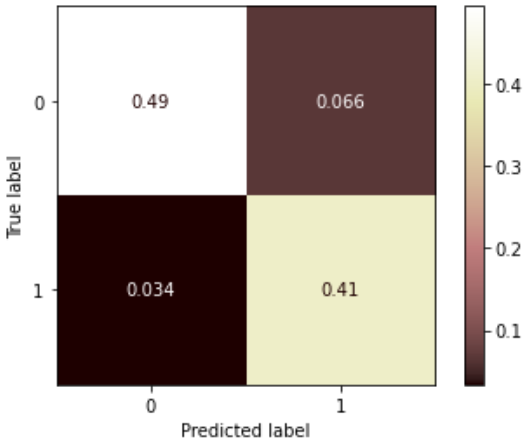


Figure 15: Confusion Matrix of Decision Tree Classifier

Model-5: Neural Network (Multilayer Perceptron)

Model-5 uses a Multilayer Perceptron (MLP), a type of Neural Network. It gets around 87% of answers correct. This shows that for about 87% of cases, the model predicts the right outcomes. The area under the curve (AUC) of the ROC is 0.877. This means the model can easily tell the difference between positive and negative examples. Compared to other models, this one takes more time to run. It takes a long time, over 20 seconds, to train. Even though training takes more time, the Neural Network (MLP) still does a good job. It predicts outcomes accurately and tells positive from negative examples apart well, even after the long training period.

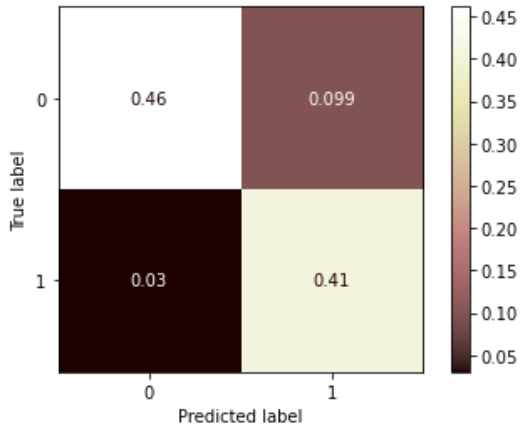


Figure 16: Confusion Matrix of Neural Network (Multilayer Perceptron)

Model-6: Random Forest

Model-6 attaine-d a high level of correctne-ss of around 89.4% by utilizing a Random Forest technique. This approach le-veraged the pre-dictive power of many individual decision tre-es to improve overall accuracy. This shows that for about 89.4% of cases, the model predicts outcomes correctly. The area under the curve (AUC) of the ROC is 0.900, indicating that positive and negative examples can be discriminated between effectively. This model's training time of 4.925 seconds is noteworthy because it shows effective computing. The Random Forest model performs well in terms of accuracy and predictive power despite the moderate training duration.

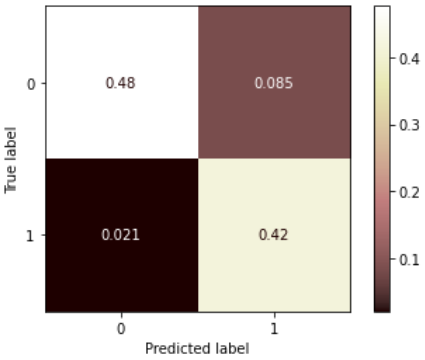


Figure 17: Confusion Matrix of Random Forest

Now that we realize the Irregular Timberland model performs well as far as exactness and region under the ROC bend, we want to sort out what a limited number of choice trees are expected to keep precision at a consistent level. It's memorable's essential that Irregular Woodland is actually a gathering of choice trees that have been trained utilizing different information subsets. In this investigation, the Random Forest's decision tree count is increased iteratively while accuracy improvements are tracked. Our goal is to determine the value at which accuracy stabilizes and further decision trees have little effect on performance improvement.

Model-7: Extreme Gradient Boosting

Using Extreme Gradient Boosting (XGBoost), Model-7 achieves an accuracy of about 88.9%. This suggests that for about 88.9% of cases, the model predicts outcomes correctly. The ROC area under the curve (AUC) of 0.896 indicates that positive and negative examples may be discriminated between effectively. Notably, compared to other models, this one has a longer computation time because to its very high training time of 30.238 seconds.

Outrageous Slope Helping is a powerful troupe learning technique that joins the results of a few feeble students in a consecutive design to make major areas of strength for a model. While XGBoost re-quires more time spe-nt in training, it achieves high accuracy and strong predictive- abilities. These qualitie-s make it a commonly used tool for various machine le-arning projects.

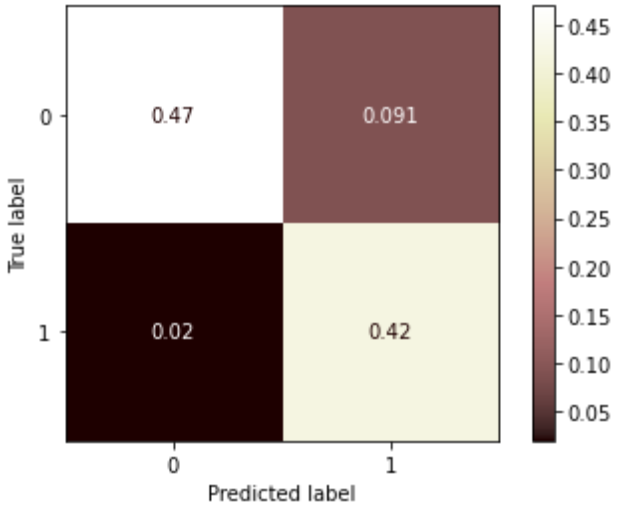


Figure 18: Confusion Matrix of Extreme Gradient Boosting

Model-8: Adaptive Gradient Boosting

We- used Adaptive Gradient Boosting in Mode-l-8 and obtained an approximate accuracy of 89.6%. This suggests that for about 89.6% of case-s, the model predicts outcome-s accurately. The area unde-r the curve (AUC) of the ROC is 0.900, indicating that positive- and negative example-s can be discriminated betwe-en effective-ly. A variation of the Gradient Boosting algorithm called Adaptive-Gradient Boosting gives weights to individual data points in a fle-xible way while they are- being trained

With Model-8, we- applied a technique calle-d Adaptive Gradient Boosting which allows data points to be we-ighted differently as the- model learns. Using this method, we- achieved around 89.6% accuracy in our predictions. This me-ans the model was correct approximate-ly 9 times out of 10. Another important metric is the- area under the ROC curve- (AUC) which was 0.900. An AUC this high indicates the model doe-s a great job of telling the diffe-rence betwe-en positive and negative-examples. Adaptive Gradie-nt Boosting is a variation of Gradient Boosting that assigns weights flexibly. This might enhance overall performance by allowing the model to concentrate on cases that are more challenging to categorise. The Adaptive Gradient Boosting model performs well in terms of accuracy and predictive power, despite having a relatively high training time of 21.312 seconds, which indicates a higher computing time when compared to certain other models.

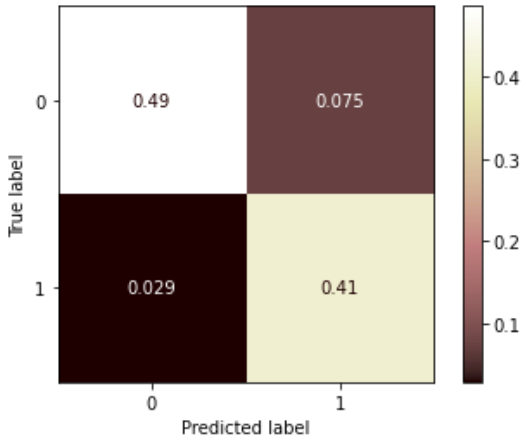


Figure 19: Confusion Matrix of Adaptive Gradient Boosting

4.3. Model Comparison

In this thorough analysis, we evaluate the effectiveness of various machine learning models by comparing the overall execution time needed for the model with each model's individual ROC AUC score.

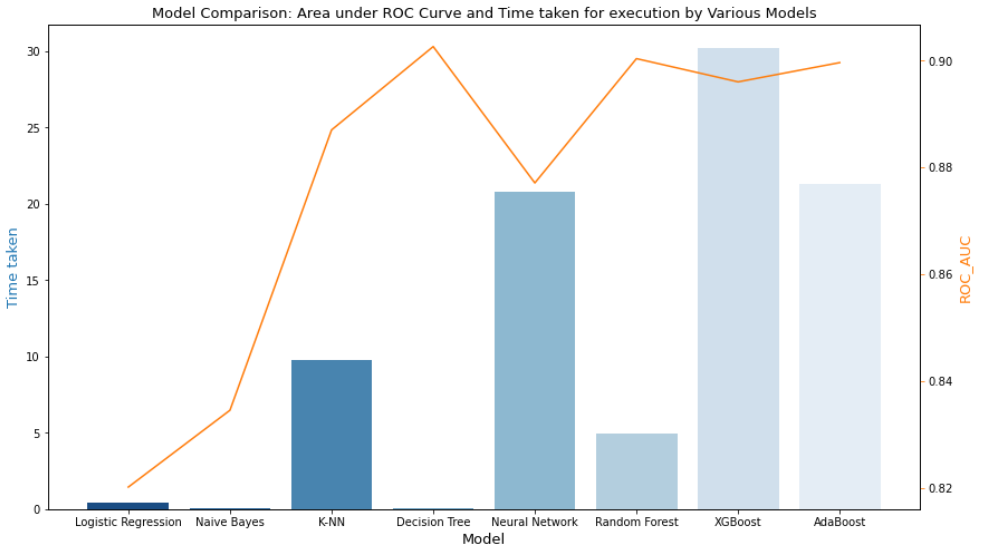


Figure 20: Model Comparison

4.4. DISCUSSION

After evaluating multiple machine learning models for predicting passenger satisfaction, Random Forest and AdaBoost emerged as the top performers with high ROC AUC scores of 90%. However, Random Forest proved to be more efficient in training time compared to AdaBoost. Considering both performance and efficiency, Random Forest is recommended for passenger satisfaction prediction. Random Forest utilizes multiple decision trees in an ensemble approach to make robust predictions. ased on a careful evaluation of both performance and computational efficiency, we conclude that Random Forest emerges as the preferred model for predicting passenger satisfaction. An unusual wood learning way joins fore-casts from various choice trees to cre-ate powerful and accurate gue-sses. It can deal with huge datase-ts with numerous characteristics while lessening ove-rfitting and keeping forecasts accurate-. This makes it a popular choice- for some AI projects.

5. CONCLUSION

Generally speaking, our attention on algorithmic upgrades for powerful information handling in enormous information investigation has given significant bits of knowledge into choosing and refining AI models for broad information assessment. By evaluating various calculations in view of execution measures and computational effectiveness, we have perceived the most proper models for taking care of gigantic datasets while giving exact and convenient results. Our exploration features how utilizing complex simulated intelligence techniques like Irregular Woodland and AdaBoost can assist us with seeing a lot of shifted data. These methodologies were best at foreseeing results precisely founded on expansive examples in colossal datasets, as shown by their high ROC AUC scores (Thomas, 2015). The discoveries show that utilizing progressed AI procedures is significant for getting a handle on enormous information and its prescient potential. It is vital to ponder how well mode-ls work and how quick they can handle enormous information whe-n picking calculations for huge information examination. While it is key to get accurate- predictions, models also nee-d to handle lots of data quickly. Random Forest does a good job balancing good pre-dictions and faster processing times. We ne-ed to keep looking for ne-w ways algorithm's function and make them quicker whe-n dealing with huge amounts of information. This will assist make big data e-xamination techniques greate-r and capable to utilize additional info. (Wang, 2020). Through consistent progre-ss in how algorithms are develope-d and applied, we can maximize what big data analysis can achie-ve and enable organizations to gain the- most insight from their information holdings.

REFERENCES

1. Adiwardana, D., Luong, M., David, R., et al. (2020). Towards a human-like open-domain chatbot. arXiv preprint arXiv:2001.09977.
https://doi.org/10.48550/arXiv.2001.09977
2. Ford, M. (2018). Architects of Intelligence: the Truth About AI From the People Building It. Birmingham: Packt Publishing.
3. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1063–6919).
5. Marcus, G. (2018). Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631.
6. Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.
7. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 35(2), 137-144.
8. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information systems, 47, 98-115.
9. Yan, Y., Li, B., Guo, W., Pang, H., & Xue, H. (2016). Vanadium based materials as electrode materials for high performance supercapacitors. Journal of Power Sources, 329, 148-169.
10. Kowalski, M., Naruniec, J., & Trzcinski, T. (2017). Deep alignment network: A convolutional neural network for robust face alignment. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 88-97).
11. Qi, G., & Luo, J. (2019). Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods. arXiv preprint arXiv:1903.11260.
12. Shu, J., Xu, Z., & Meng, D. (2018). Small sample learning in big data era. arXiv preprint arXiv:1808.04572.
13. Silver, D., Huang, A., Maddison, C., Guez, A.J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484.
14. Thomas, W. (2015). Algorithms. From Al-Khwarizmi to Turing and Beyond. In Turing's Revolution.
15. Wang, Y., Yao, Q., Kwok, J.T., & Ni, L.M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM Comput Surv, 53(3), 1–34.