Critical Analysis and Prediction of Airline Passenger Satisfaction

Tasdeed Haq, 210001917
School of Mathematics, Computer Science and Engineering
City, University of London
London, United Kingdom
taz.haq@city.ac.uk / tasdeed.o.haq@gmail.com

***Abstract*— The aim of this study was to analyze which airline services contributed most to passenger satisfaction and which services should be emphasized to boost satisfaction and increase loyalty. Twelve services were deemed important overall, subsets of these being more important accounting for class, where significantly more services were vital in business class compared to lower tiers. Clustering techniques were applied to identify subgroups of satisfied passengers and machine learning models were implemented to assess how well satisfaction could be determined.**

Analysis showed airlines should focus on improving services for those in lower classes who were easily satisfied but suffered poor satisfaction rates before improving services for business class where rates were already high. Airlines should further incorporate incentive programs to holidaymakers alongside. The optimal model provided excellent predictive capabilities with 96% accuracy hence a beneficial tool for airlines to quickly analyze satisfaction given large passenger numbers.

Keywords—Passenger Satisfaction

I. Introduction

The airline industry is an extremely competitive one, both pre- and post-pandemic, given the large number of major carriers operating in the field. Therefore, the delivery of high-quality services is crucial for airline survivability. High quality services produce a knock-on effect in the sense that it hopefully leads to increased passenger loyalty hence an increased market share resulting in higher profits [1], which is important because whilst the number of passengers kept rising, roughly a 5-7% increase yearly pre-pandemic (2015-19) [2], the net profit per passenger was declining, "$10 for 2015, $9 for 2016-17" [3] and with marginally less in 2018-19 [4]. As airlines are now bouncing back, we expect this trend to continue. This leads to airlines having to deal with slight profitability [5] due to the tight margins of having to cover costs flights whilst keeping fares low enough to maintain competitiveness [6].

To worsen things, the pandemic occurred which led to total airline revenue falling so low in 2020, it matched total airline revenue 18 years before, in 2002 [7], wiping out nine years of industry profit [8] – this was as a result of international border restrictions, significant decreases in the number of passengers wanting to fly (a world decrease in 60% [9]), and flights flying with reduced passengers to maintain social distancing [9]. Further combined with increased costs such as fuel and labor, many carriers have struggled for example Virgin Atlantic which required a government bailout [10]. As a result, it is expected that airlines will increase ticket prices by 3% in a bid to recoup their debts [7]. Due to this, even pre-pandemic, airlines have now placed more emphasis on service quality instead [11] as a means of attracting more passengers, where the dataset we will be using considered as grounds for understanding passenger satisfaction.

II. Analytical Questions and Data

With increased importance in understanding what factors affect passenger satisfaction, many airlines now carry out surveys to assess what services satisfy their passengers most [12]. Much literature available focus on the relationship of how services relate to satisfaction, hence less available literature relating demographical factors to satisfaction. Sivesan et al. [13] investigated this with customers from banks and found that factors such as gender did not affect satisfaction and factors such as age did. We can explore this further, and see how this affects specifically airline passenger satisfaction, hence we pose the question:

1. Do personal demographic factors such as Age and Gender contribute to Passenger Satisfaction?

A natural hypothesis is that Business Class passengers are more satisfied, which is an interesting question to investigate and carry out further analysis on. This may help airlines decide whether to improve services for passengers in lower tiers:

2. Does the class of a passenger indeed have an effect on satisfaction? If so, can we investigate what services are most important between Economy and Business Class?

In general, we would like to examine the drivers between passenger satisfaction, thus this research aims to answer the following which give airlines a sense of what services need to be emphasized:

3. Can we identify the most important services that contribute to satisfaction?
4. Do subgroups appear amongst satisfied passengers in which different services are more important?
5. Is it possible to devise a reliable model to predict passenger satisfaction? How much contribution do the services we deemed important have?

The dataset that we have is suitable for answering these questions, as it contains personal demographical and class data as well as more than 15 kinds of airline services. With a large dataset (>120,000 observations), this will allow for a detailed analysis of airline passenger satisfaction.

III. Data (Material)

The dataset is extracted from Kaggle [14] consisting of 129,880 observations further containing measurements such as passenger loyalty, reason for travel, how long flight delays were in minutes for departure and arrival and flight distance. Each service such as cleanliness were all measured using ratings between zero and five.

There were two additional features, 'Unnamed: 0' and Passenger ID; the former being meaningless, whilst Passenger ID was used to identify the presence of duplicate reviews which the dataset did not; as this then held no relation to satisfaction after, both features were then dropped.

Feature	Contents
Class	'Eco', 'Eco Plus', 'Business'
Gate Location	0, 1, 2, 3, 4, 5
Check-In Service	0, 1, 2, 3, 4, 5
Customer Type	'Loyal', 'Disloyal'
Age	Continuous - 7 to 85 years old
Satisfaction	'satisfied', 'neutral or dissatisfied

Fig. 1. Sample Overview of Features Contained in Data

With all non-numerical features containing no erroneous values, for example passenger ratings were indeed rated from 0-5 etc., outlier analysis was carried out on continuous features such as 'Age' through histograms and comparison of quartiles to extremes which highlighted the distributions as well as potential outliers (Fig. 2).

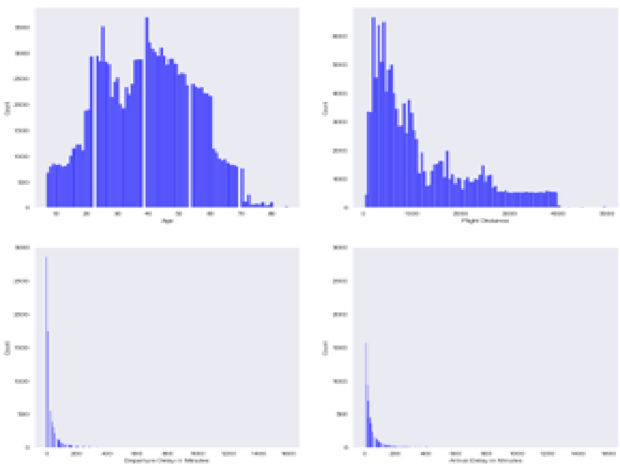

Fig. 2. Initial Distributions of Continuous Features, clear outliers skewing the axis except Age

Whilst 'Age' showed a roughly Gaussian distribution, Departure and Arrival Delay Times exhibited extreme right skew, which is expected, as most flights depart and arrive on time. In terms of Flight Distance, this showed less right skew but highlighted that the data contained more short-medium haul flights. All three skewed features showed outliers - due to their non-Gaussian nature, bootstrapping techniques were employed to remove outliers rather than methods utilizing interquartile ranges [15].

Investigation of missing values revealed arrival time delays contained 377 null values; without further information on the specific flight in terms of where it's flying between, and carrier, it would be inaccurate to impute these values by other means and were therefore dropped. Compared to the original dataset, overall, 3% of data was removed – leaving 125,934 observations.

Imbalances in the amount of 'satisfied' to 'neutral or dissatisfied' passengers (our predictor variable) were investigated and showed to pose no issues with a 45% to 55% ratio.

IV. Analysis

A. *Exploratory Data Analysis (EDA)*

Before answering our questions, EDA was implemented. Univariate analysis showed imbalances amongst passenger classes, loyalty and type of travel. Many services looked evenly distributed across ratings with some showing increasing numbers as rating increased such as Onboard Service:

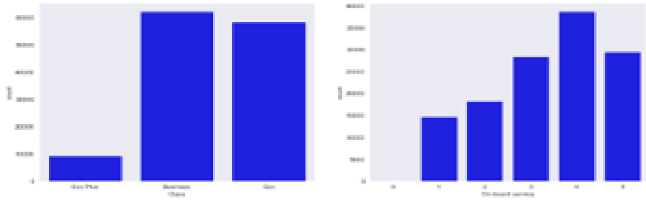

Fig. 3. Sample Univariate Analysis on Class and On-Board Service

Correlation analysis was carried out assessing feature relations and to give ideas of which features could be important towards satisfaction.

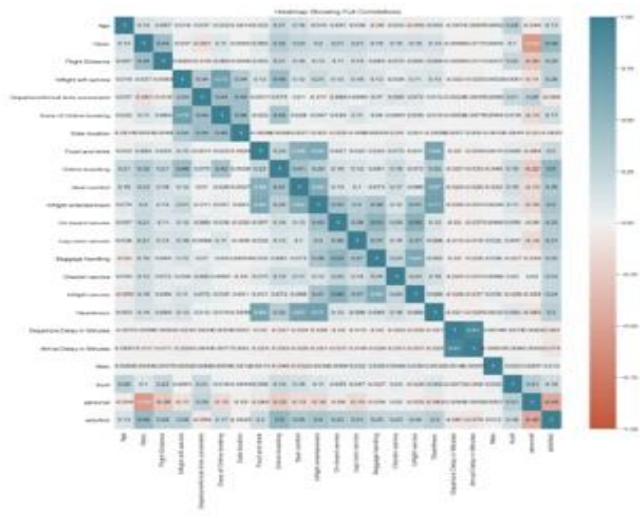

Fig. 4. Correlation Heatmap Highlighting Strength of Linear Relationships Between All Features

B. *Question 1*

Fig. 4 suggested age had little effect on satisfaction. With age being roughly Gaussian, this was split into four groups using quartiles. An initial test to see if age was related to satisfaction was comparing satisfaction by age group in Fig. 5:

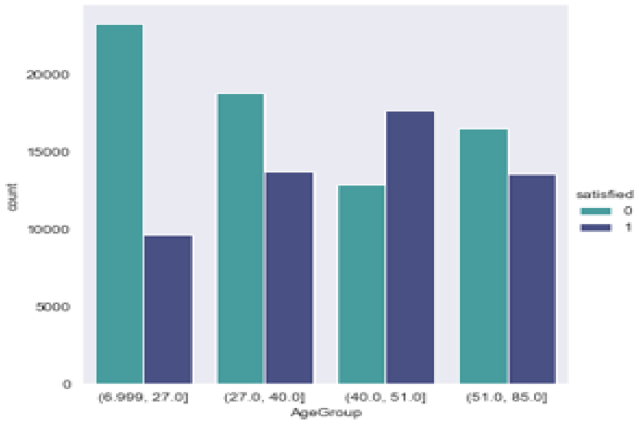

Fig. 5. How Age Groups Varied with Satisfaction

Passengers between 40-51 looked more satisfied unlike other groups. However, this does not give the full story, i.e., this could just be a quirk from the way age was split. For a clearer view of age's relevance, bivariate distributions of the groups with other services were examined accounting for satisfaction to assess whether distributional changes appeared for differing groups however there was not, show in Fig. 6.

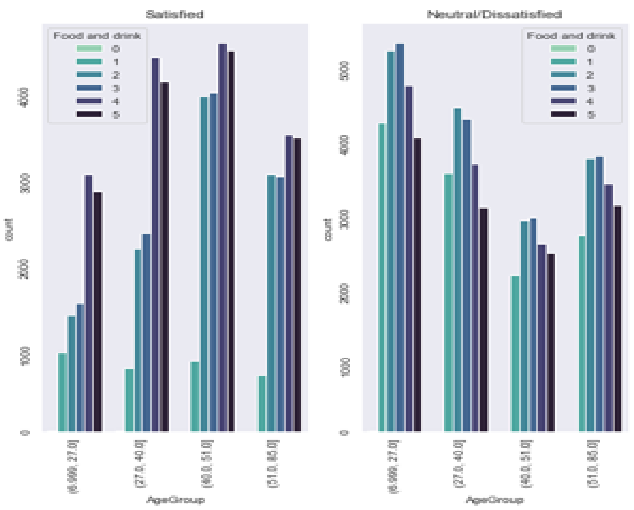

Fig. 6. Sample Plot - Highlighting No Changes in Bivariate Distributions for All Age Groups Even When Accounting for Satisfaction

With age having low correlations with satisfaction and all features via Fig. 4, combined with the fact that accounting for satisfaction, the bivariate distributions of different age groups with other features remained unchanged deemed age unimportant. A similar conclusion held for Gender using similar methodology– hence, both were removed.

C. Question 2

Analysis of class showed roughly 19% of Economy, 25% of Economy Plus and 70% of Business Passengers were satisfied. Whilst this answers the first part of the question, the huge jump in satisfaction motivated the second part. To answer this, we considered differences between Economy and Business only, due to small differences in satisfaction between Economy and Economy Plus, and further because these were the most populated classes hence a fairer comparison.

Similar approaches were taken to Question 1, however we only considered satisfied passengers i.e., looking at distributional changes of services amongst different classes, for example in Fig. 7:

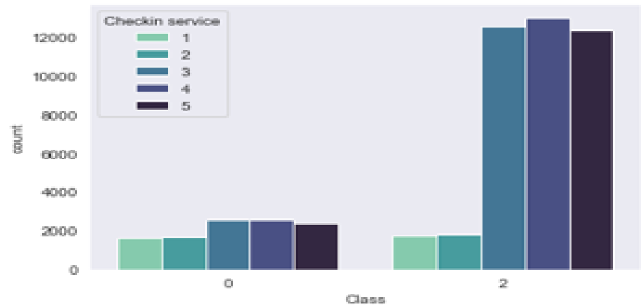

Fig 7. Sample Plot Highlighting Importance of Check-In Service amongst Satisfied Economy and Business Passengers

A summary of which services were more important to each class:

Class	Important Services
Economy	In-Flight Wi-Fi Service, Ease of Online Booking, Food and Drink, Online Boarding, Seat Comfort (slightly), In-Flight Entertainment (slightly), Baggage Handling (slightly), In-Flight Service (slightly) and Cleanliness
Business	Food and Drink, Online Boarding, Seat Comfort, In-Flight Entertainment, On-board Service, Leg Room, Baggage Handling, Check-In Service, In-Flight Service and Cleanliness

Fig 8. Importance of Services to Biggest Classes of Passengers

The services that held significantly more importance to Business Passengers than Economy were Seat Comfort, Inflight Entertainment, Onboard Service, Leg Room, Baggage Handling, Check-In Service and Inflight Service. Naturally, it takes more to please Business Class passengers.

D. Question 3

Whilst Fig. 4 gave ideas on related features to satisfaction, this is not the best representation. We deeper examined this by looking at how the proportions of satisfied passengers changed for each feature:

On-Board Service	Proportion Satisfied	Gate Location	Proportion Satisfied
0	0.000	0	1.000
1	0.195	1	0.496
2	0.250	2	0.463
3	0.315	3	0.345
4	0.533	4	0.388
5	0.648	5	0.568

Fig 9. Sample Overview of How Proportions of Satisfied Passengers Vary per Rating

Implementation of this for all services showed that for many services, there were general increasing trends in satisfaction levels as rating increased, suggesting contribution to satisfaction, however there were few which did not follow this trend for example 'Gate Location' (Fig. 9) and 'departure/arrival time convenient'. Further examination of these using approaches and conclusions in Question 1 found that these were deemed unimportant – hence were removed.

For numerical features, Fig. 4 showed high correlations (0.93) between Departure and Arrival Delay in Minutes hence we dropped the latter, leaving ‘Flight Distance’ and ‘Departure Delay in Minutes’ – analysis of the former produced Fig. 10 highlighting its importance, with many dissatisfied passengers flying shorter distances whilst satisfied passengers showed noticeably large amounts of satisfaction over longer distances:

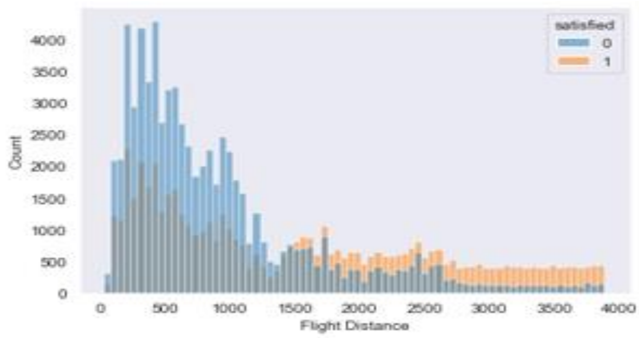

Fig. 10. How Satisfaction Varied with Flight Distance

Due to high skew in ‘Departure Delay in Minutes’, KDE plots were used, which showed marginal differences in terms of satisfaction; similar approaches in Question 1 showed this was the case hence removed. Overall, 8 features were dropped.

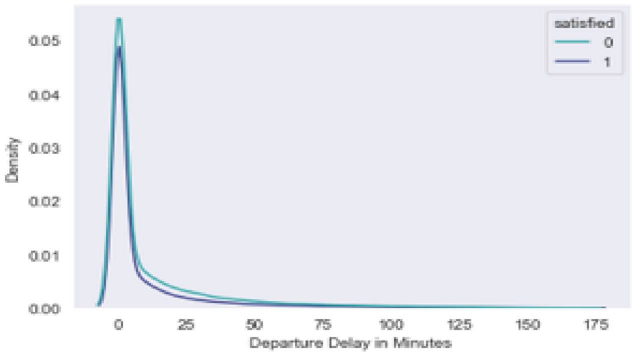

Fig. 11. How Satisfaction Varied with Departure Delay

Fig. 4 showed three apparent groupings of features, falling into the categories convenience, comfort, service:

Convenience	In-Flight Wi-Fi Service, Departure/Arrival Time Convenient, Ease of Online Booking and Gate Location
Comfort	Food and Drink, Online Boarding, Seat Comfort, In-Flight Entertainment and Cleanliness
Service	In-Flight Entertainment, On-Board Service, Leg Room Service, Baggage Handling and In-Flight Service

Fig. 12. Notable Groupings from Correlation Heatmap

These contained some services we deemed unimportant; however, groupings may cause more contribution to satisfaction – hence a new dataset was created with the combined features aggregating via averages yielding a reduced dataset containing eight features against satisfaction to be used as comparison with the ungrouped dataset later.

E. *Question 4*

K-Means Clustering was performed on only satisfied passengers in hopes we could identify subgroups of passengers where different services had different importance’s. This was not performed on the grouped dataset due to uncertainty on how the new features performed.

Both datasets were scaled via Min-Max Scaling as resulting datasets would still be interpretable and due to the non-normality of Flight Distance.

Two heuristics were used in determining number of clusters, elbow plots and average silhouette scores. The former suggested three clusters whilst the latter suggested two, however two yielded major imbalances in size, hence three was determined optimal with higher numbers giving worsened results, resulting in Fig. 13.

Features:	Cluster 1	Cluster 2	Cluster 3
Class	1.982	0.145	1.983
Flight Distance	1675.107	678.430	1784.329
In-Flight Wi-Fi Service	4.034	3.916	1.383
Ease of Online Booking	4.033	3.236	1.448
Food and Drink	3.503	3.514	3.535
Online Boarding	4.211	3.717	3.982
Seat Comfort	4.064	3.505	4.146
In-Flight Entertainment	4.114	3.365	3.993
On-Board Service	4.123	3.269	3.913
Leg Room Service	4.067	3.232	3.875
Baggage Handling	4.245	3.399	3.966
Check-In Service	3.795	3.165	3.764
In-Flight Service	4.237	3.431	3.986
Cleanliness	3.789	3.504	3.843
Loyal (1 = Loyal)	0.880	0.840	0.964
Personal (1 = Personal Travel)	0.007	0.301	0.006
Satisfied (1 = Satisfied)	1.000	1.000	1.000
Proportion of Overall Satisfied Passengers	0.45	0.23	0.32

Fig. 13. Clustering Results with 3 Clusters Applied to Scaled Data, Which Were Then Transformed Back to Unscaled Counterparts and Aggregated via Mean

Clusters 1 and 3 were nearly completely Business passengers whilst Cluster 2 made up nearly the entirety of the remaining classes. The major difference between the two Business clusters was that Inflight Wi-Fi Service and Ease of Online Booking had low importance in satisfaction with far below average ratings. Cluster 1 rated all services very highly with average ratings nearly all above 4 showing that this contained passengers which were hardest to please, making up 45% of satisfied passengers. Cluster 2 showed average ratings for nearly all services, yet passengers were still satisfied suggesting it takes less to satisfy those in lower classes, although it is noted that average flight distances were short– many services have less importance on shorter flights.

F. *Question 5*

With two proposed datasets, both scaled - multiple models, including some with hyperparameter tuning (HT) via basic grid searches (with 10-Fold Cross Validation) were tested – summarized in Fig. 14:

	GNB	L-SVC (HT)	LR (HT)	DT (HT)	RF
Cross-Val Mean Acc.	0.867	0.873	0.874	0.953	0.960
Cross-Val std.	0.003	0.002	0.002	0.001	0.0005
Train Acc.	0.867	0.873	0.874	0.968	1.000
Test Acc.	0.860	0.867	0.867	0.954	0.960
Precision	0.816	0.828	0.828	0.927	0.933
Recall	0.855	0.859	0.859	0.966	0.971
F1 Score	0.835	0.843	0.843	0.946	0.952
ROC-AUC	0.920	N/A	0.920	0.983	0.992

Fig. 14. Metrics for All Models Fit, GNB - Gaussian Naive Bayes, L-SVC - Linear SVC, LR - Logistic Regression, DT - Decision Trees, RF - Random Forests, (HT) - Hyperparameter Tuned. NOTE: Results ONLY for ungrouped data, due to poor performance from grouped data

As satisfaction is balanced, accuracy was deemed a good metric, which was used alongside Precision, Recall, F1 score and AUC. 10-Fold Cross Validation was used initially to assess expected goodness-of-fit via mean cross-validated training accuracy. We stopped using the grouped dataset after fitting the initial two models. It consistently underperformed on the simplest models compared to our ungrouped dataset which was performing significantly better hence had no use in more complex models. Our groupings did not help in identifying satisfaction.

Clearly, the default Random Forest model outperformed all other models in every metric with impressive values and therefore our ideal model. We could see in the below plot, that all features we deemed important were being utilized supporting all earlier conclusions.

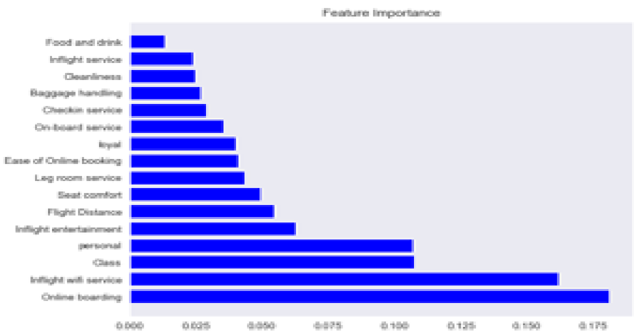

Fig. 15. Feature Importance's for Random Forests Model

V. FINDINGS, REFLECTIONS AND FURTHER WORK

Based on the data, we have shown that personal demographical factors such as age and gender did not play a role passenger satisfaction. Via questions one and two, we successfully managed to extract the most relevant services overall and combined with question two also identified the most important services specifically to the biggest classes: Economy and Business. This proves immensely helpful for airlines; by improving targeted services for both classes this would hopefully yield increased loyalty. We found big differences in satisfaction for type of travel, this would require further investigation, however in the meantime, airlines should utilize incentive programs to help counteract this and improve satisfaction.

Our K-Means Clustering Model successfully managed to separate satisfied passengers in Economy and Economy Plus from satisfied passengers in Business (Fig. 13). Whilst no distinct subgroups were identified in the former, two large clusters were identified within satisfied Business Class passengers, where we saw Inflight Wi-Fi Service and Ease of Online Booking received far below average ratings in one cluster, yet still led to satisfaction. Whilst a strange phenomenon, this supported conclusions from question two where these services were not deemed important to Business Class. We further identified clusters of passengers who were easiest and hardest to satisfy which in fact were Economy and Business Class passengers respectively. Given that satisfied Economy and Economy Plus passengers were easier to please, they however only accounted for 23% of satisfied passengers, hence airlines should place further emphasis on satisfying these passengers before moving on to improving quality of those in Business as the current quality of services here already led to an admirable 77% satisfaction.

We saw the default Random Forest model outperformed all other models across all metrics, with notably a 96% accuracy. Comparison of mean cross-validated accuracy and test accuracy showed the model was not overfitting. We further saw all features we deemed relevant were all important in the model supporting analysis from earlier sections. This model will prove to be a reliable tool within airlines to quickly analyze passenger review data given large passenger numbers.

One must note however the limitations with our data. Satisfaction, whilst only taking two outcomes, was not distinct i.e., 'satisfied' and 'neutral or dissatisfied'. Hence, by modelling whether a passenger was not satisfied, we implicitly assumed this meant the passenger was dissatisfied which is not accurate. Lack of multiple sources of information on airline carrier, arrival or departure times, where flights were flying from and to or plane model (and therefore seat capacity etc.) were missing. This raises questions on the validity of the data and whether it was an accurate representation of passenger satisfaction. Hence, more interesting analysis could not be done for example identifying whether newer airplane models, which have more functionalities yielded more satisfied passengers, identifying which clusters of airlines yield the most dissatisfied passengers (hypothetically, one would assume low-cost airlines [16]), a spatial analysis on satisfaction and more.

Further lines of enquiry aside from fixing above limitations, include a thorough investigation of why reason for travel affected satisfaction significantly and improving our optimal model by hyperparameter tuning. We could try using other complex ensemble models such as XGBoost and Bagging Classifiers to further improve models. Advantages of using these is that they can handle 'evolving data streams' [17]. It is fair to assume that over time passenger preferences change and hence importance of different services will change with it, therefore old data becomes unrepresentative. As airlines collect more data over large periods of time, this is where these models outperform Random Forests, with its ability to handle and highlight changing importance's in services.

VI. WORD COUNTS

Excluding the Title, Student Details, Headings and Subheadings and References.

Abstract	150/150
Introduction	300/300
Analytical Questions and Data	300/300
Data (Material)	300/300
Analysis	1000/1000
Findings, Reflections and Further Work	600/600

REFERENCES

[1] J.-W. Park, R. Robertson, and C.-L. Wu, 'The effect of airline service quality on passengers' behavioural intentions: a Korean case study', Journal of Air Transport Management, vol. 10, no. 6, pp. 435–439, Nov. 2004, doi: 10.1016/j.jairtraman.2004.06.001.

[2] https://www.icao.int/annual-report-2019/Pages/the-world-of-air-transport-in-2019.aspx

[3] H.-J. Ban and H.-S. Kim, 'Understanding Customer Experience and Satisfaction through Airline Passengers' Online Review', Sustainability, vol. 11, no. 15, Art. no. 15, Jan. 2019, doi: 10.3390/su11154066.

[4] https://www.statista.com/statistics/655375/net-profit-per-passenger-of-commercial-airlines-worldwide

[5] J. Namukasa, 'The influence of airline service quality on passenger satisfaction and loyalty: The case of Uganda airline industry', TQM Journal, vol. 25, no. 5, pp. 520–532, 2013, doi: http://dx.doi.org/10.1108/TQM-11-2012-0092.

[6] https://money.com/airline-profit-per-passenger/

[7] https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/back-to-the-future-airline-sector-poised-for-change-post-covid-19

[8] https://www.bloomberg.com/news/articles/2021-10-04/airline-losses-from-covid-to-exceed-200-billion-industry-says

[9] https://www.icao.int/sustainability/Documents/Covid-19/ICAO_coronavirus_Econ_Impact.pdf

[10] https://www.bbc.co.uk/news/business-53406604

[11] S. Tsafarakis, T. Kokotas, and A. Pantouvakis, 'A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement', Journal of Air Transport Management, vol. 68, pp. 61–75, May 2018, doi: 10.1016/j.jairtraman.2017.09.010.

[12] N. Halpern and D. Mwesiumo, 'Airport service quality and passenger satisfaction: The impact of service failure on the likelihood of promoting an airport online', Research in Transportation Business & Management, p. 100667, Jun. 2021, doi: 10.1016/j.rtbm.2021.100667.

[13] S. Sivesan and M. Karunanithy, 'Personal Demographical Factors and their Influence on Customer Satisfaction from Customer Perspective', European Journal of Business and Management, vol. 5, no. 20, pp. 42–45, Aug. 2013.

[14] https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction

[15] https://suplari.com/calculating-outliers-spend-data-bootstrapping/

[16] https://www.theguardian.com/business/2015/dec/05/easyjet-complaints-compensation-claims-denied

[17] H. M. Gomes et al., 'Adaptive random forests for evolving data stream classification', Mach Learn, vol. 106, no. 9, pp. 1469–1495, Oct. 2017, doi: 10.1007/s10994-017-5642-8.