e-PROCEEDINGS

of The 5ᵗʰ International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings
of the 5ᵗʰ International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

ANALYSIS OF THE PASSENGERS' LOYALTY AND SATISFACTION OF AIRASIA PASSENGERS USING CLASSIFICATION

Ee Jian Pei[1], Chong Pui Lin[1] and Nabilah Filzah Mohd Radzuan[2]
[1] Centre for Emerging Technologies in Computing (CETC),
Faculty of Information Technology, INTI International University
[2] Faculty of Computing, Universiti Malaysia Pahang
([1]i17012981@student.newinti.edu.my, [1]puilin.chong@newinti.edu.my,
[2]nabilahfilzah@ump.edu.my)

"Business rules" of airline is to earn profits by providing air transportation services and flights to the travel passenger. Passengers purchase airline tickets base on their different requirement to choose a satisfied flight from different airlines, and become a loyal passenger to the airline due to satisfaction. The airline can establish a long-term win-win relationship with the loyal passengers which is long-term purchase company's flights as the passenger can have a satisfied flight, while the airline can earn the long-term profit. This research proposes the dashboard system in finding the passengers' purchase behavior of loyalty and satisfaction from the hidden data in order to make a better business strategy in stand out from other competitors, and visualize the report. The classification method will be used included random forest, logistic regression and lightgbm. The result will identify the various possibilities of information, and contribute prediction of passenger loyalty and satisfaction.

Keywords: analytic, classification, random forest, logistic regression, lightgbm, airline

1. Introduction

Airline industry has played a significant role in the rapid development of the world economy because airlines not only providing air transportation services for passengers to drive the development of the tourism industry around the world and also increase trade activities around the world. "Scheduled airlines carried more than 2.5 billion passengers last year and 40% of international tourists now travel by air. According to the IATA's (the International Air Transport Association) air travel has grown by 8% per year." (Jahmani, 2020).

This phenomenon had highly increased the level of competition among the airlines and also the demand of the airline services. At the same time, it had also caused each of the airlines to constantly update the technology and services in order to provide better services for passengers. The main source of revenue for an airline depends on how many passengers take the airline's flight, so passengers are very important to the airline. Even the airlines also provide low-cost airfare as well as the basic service for the requirement of passengers who desire to have a cheaper airfare on their trip. This can allow the airlines can better expand and improve their business. Therefore, many airlines companies started to focus on improving the level of the quality service, which can participate in the competition between the low-cost airline and the full-cost airline.

In Malaysia, there are some of the popular airlines can always be seen which including Malindo Air, Firefly, Malaysia Airlines, MASwings, and AirAisa. For this study, the AirAsia company is used as case study to propose the analysis of the relationship between the standard of passengers' loyalty and satisfaction. AirAsia is one of the famous low-cost carrier (LCC) which also the largest airline in Malaysia. The headquarter of AirAsia is located in near Kuala Lumpur, Malaysia as well as its main hub is at Kuala Lumpur International Airport 2 (KLIA2).

The service of AirAsia plays an important role in the passengers' satisfaction and loyalty. This is because a good service can satisfy the passenger and it may create a loyal customer to the airline. The loyal passenger normally will repurchase more often than ordinary customers and even spend more. Sometimes, satisfied passengers will share their flight's experience with others and it indirectly helps the airline to do free advertising in order to increase potential customers. (Sandada and Matibiri, 2016).

Other than that, AirAsia always consistently remain inspires humanity to listen to the passengers' requirements and deliver better service for the passengers' flight experience. For example, AirAsia as a low-cost carrier would like to allow the passengers to have good flight experience, there has provided the most legroom in coach, free Wi-Fi in flight, inflight entertainment, name-brand snacks or hot meals, and also friendly service. Hence, the services of AirAsia will directly affect the passenger's satisfaction and also loyalty.

Even though there were many studies about the airline industry, the study of the relationship between the standard of passengers' satisfaction and loyalty to the services provided remains inadiguate. For instance, there is a study about passenger satisfaction towards service quality in Malindo Airlines. (Hassan, Sapri, Mustafa, Mohammad, Rajamanikam, and Mahmood, 2019) Sometimes, there can see that the passengers may be satisfied with the service but they are not loyal, while in another hand, the passengers maybe not satisfied with the service but they are loyal as the satisfaction and loyalty are not substitutes for each other.

Therefore, this paper proposed a study to help AirAsia to find out the factor that will affect passenger satisfaction and loyalty in order to predict passenger satisfaction to support decision-making in service improvement. The factors that may affect the passenger satisfaction and loyalty can be the flight punctuality, seat comfort, type of travel, class of the cabin, inflight service, cleanliness, leg room, and so on. The way of AirAsia participates in the competition between flights is providing passengers with low-cost air tickets and good service, so the quality of service is very important to AirAsia. Usually, when each passenger purchases the airline tickets, they will be based on their expectation to select the satisfied flight from different airlines, and the passenger may become a loyal passenger to the airline due to satisfaction on the airline service

2. Related Work

In this study, AirAsia was the airline used as case study. AirAsia is a leading low-cost carrier not only in Malaysia but also in ASIA and ASEAN. It had connected the people and places with 388 routes, especially 104 of the routes are the unique routes. The entire business model of AirAsia is surrounding by a lowcost concept which aims to run the business in simple and efficient. (AirAsia, 2020) AirAsia had taken a few strategies to achieve this concept which is the high utilization of aircraft, low airfare, and the point-to-point network. Factors that affect the competitiveness of airline are included service quality, customer satisfaction, and customer loyalty,

2.1 Service Quality

Service is the foundation of service marketing, and service quality is the core of service marketing. Whether the enterprise with tangible products or intangible service industry, service quality is one of the important elements keys to survive in a competitive environment. Service quality is the overall view of the difference between the customer's expectation of the service and the perception of the service. (Suki, 2014) For the aviation industry, it has been classified as an intangible service industry. The services provided by the airline is totally differenced to other service industry as the airline mainly provide a flight for the passengers to their destination and also the services during the flight experience. (Abdullah and Noor, 2007) If its service quality is not within the expectations of passengers, there is no doubt that it will indirectly affect passenger satisfaction and loyalty, as well as the overall revenue of airlines. This is because passengers usually make a comparison based on the service quality from various airlines to see whether the service is bad or good in order to find flights that meet their needs. For instances, the services of the airline that always used to be compared such as inflight Wi-Fi service, punctuality of aircraft, the convenient time of the flight, ease of online booking, food or beverage, online boarding, seat comfort, inflight entertainment, on-board service, leg room service, baggage handling, checking service, inflight service, and the cleanliness. Therefore, airlines have to always keep improving their service quality which has become a requirement in the market in order to survive from the competition.

There are also many researchers who often wisely use SERVQUAL method to study the service quality. Of course, it is also including the study of the airline service. By applying the SERVQUAL

method in the airline industry, the 5 factors that affect the service quality have been identified by the study (Aydin and Yildirim, 2012) which are tangibles, reliability, responsiveness, assurance, and empathy. Meanwhile, the study in (Saadat, Tahbet and Mannnan, 2018) found that the service quality of one airline in Malaysia is constructed based on 5 components which are flight attendants, tangible features, food service, online service, and ground staff. The passenger is always the person who buys the service quality, and the airline is the one who acts as a service provider. Hence, there is a majority of researchers think that delivering a high quality of service can highly increase the chances the passengers purchase the flight from the airline.

2.2   Customer Satisfaction

Customer satisfaction is an overall response of complex customer experience on the services to see whether the customer is happy or unhappy, which can also be known as the evaluation of customer experience. (Ban and Kim, 2019) This is a comparison between the customer's perception and the expectation during customer evaluation. This is due to reason customer satisfaction is a very essential thing that the service industry should always keep top of mind and it can be used to have a clear understanding of what actually the customer expectation in order to allow the customer to judge the performance of the company. The conceptualizations of customer satisfaction can be separated into two which are considered as transaction-specific and cumulative. (Rahim, 2016) The transaction-specific satisfaction means that the customer satisfied with the specific service or goods, while the cumulative satisfaction is the customer's overall evaluation of the service process encountered. This means customer satisfaction is known as an important way to cultivate and meet customer preferences as well as expectations to enhance value creation. Normally, only when the customer is satisfied, the organization can succeed and profit from it. There is no way for the organization to start a business without a customer. Therefore, customer satisfaction will only exist, when the customers' perception of service is higher than customers' expectations.

2.3   Customer Loyalty

Customer loyalty is the final outcome of customer satisfaction toward the product or services of the organization. The customer loyalty is a concept that defines the behaviour of customer which repurchase the product or services from the organization. (Al-Reface, Bata and Jalham, 2014) It is considered the mind-set of the customer toward the trusty of product or service about the organization and will tend to help the organization recommend to others. Even though customer satisfaction is impossible for the organization to rely on it in order to bring higher success in the business, but it will occur a positive financial result when the customer repurchases the service regularly which is customer loyalty toward the organization. (Gures, Arslan and Tun, 2014) Once the organization has clear evidence that proves the relationship with the loyal customer, this will cause both sales and profitability of the organization to increase.

   This is because customer loyalty is based on the quality of the product or service delivered by the organization in order to fulfil customer satisfaction to establish the relationship. Moreover, the cost that used in discover new customers is always exceed than the cost to retain the current customer. Thus, this has made every organization more concerned about how to retain the current customer and change them into loyal customers. Meanwhile, this why the reason passenger loyalty is important for the airline industry to remain competitive. In the study of (Vilkaite-Vaitone and Papsiene, 2016), the frequent flier program FFP is the first loyalty program established by the airline to ensure passenger loyalty and found that the frequent flier program can effectively help to increase the number of passenger loyalty via repurchasing the flight. Consequently, this has caused every airline to have its own loyalty program as the loyalty program is the way to retain the market share.

2.4   Current Airline Dashboard System

The dashboard is a module that realizes data visualization that is possessed by general business intelligence and is a data virtualization tool that shows the status of measurement information and

Key Business Indicators (KPIs) to enterprises. (Klipfolio, 2020) Generally, it will show the summary status of the analysis data of the organization in order to deliver more information and insights for the department or organization to improve the business. The dashboard combines and organizes digital, metric, and performance scorecards on a simple screen. They adapt to specific roles and display metrics specified by a single perspective or department. Undoubtedly, the key feature of the dashboard is to obtain real-time data from multiple data sources, and it is a customized interactive interface. The dashboard provides a better experience for data with a rich and interactive visual interface.

Every system will have its own limitations, which also includes the BI dashboard system. One of the limitations that can occur in the BI dashboard system is the messy design which shows a lot of detailed information of analysis without understanding what actually target audience need from the dashboard. Since the dashboard system is based on the construct of time, so there maybe will not display the information for the current most critical of the plan. Besides that, it is hard for the dashboard system to attach the supporting data as the data will be automatically refreshed in failure. In addition, there is not a standard rule for the organization to use the metrics of the dashboard. Therefore, it causes the employee of the organization to have different ways of using the dashboard's metrics to generate reports.

2.5   Classification Techniques

Classification is known as the predictive data mining technique (machine learning) that used to assign a part of the attributes (subpopulation) from the collection in order to build a model that classifies the record of the entire collection. (Oracle, 2020). It is the supervised learning process because it includes the learning step (training) and classification step. In learning step, it is based on the training data set to develop a model to learn the input and output data, while in the classification step, there will have a test data set used to predict the class and test the accuracy of the model. (Neelamegam and Ramaraj, 2013) Classification is separated in two categories: • Binary: There only have two possible values - Example: Gender, male or female. • Multiclass: There have more than two values - Example: State: Selangor, Kedah, Johor, etc. In classification, there have many algorithms that commonly used to construct the model which including decision trees, Bayesian classifiers, neural networks, k-nearest neighbour, random forest, logistic regression, lightBGM, and etc. In this study, the random forest, logistic regression and lightBGM were used to compare the accuracy and select the most accurate model.

2.5.1   Random Forest

Random Forest is the algorithm based on the basic building block of decision trees, and the fundamental idea of the random forest is the combination of different decision trees into a single model. (CFI, 2020). This is because sometimes the predictions made by a single decision tree may be inaccurate, so there is required the random forest. The random forest can combine a large number of decision trees to aggregate the predictions and get more knowledge to make the information network bigger, in order to make more comprehensive predictions. (Yiu, 2019) There is no association between different decision trees. In the simple word, the random forest will consist of a larger number of individual decision trees constructed at the training time and outputting a class prediction as well as the most votes of the class will become the model prediction.

Advantages are included it can come out with very high-dimensional (many features) data, and there is no need for dimension reduction and feature selection, it can judge the importance of features, can determine the interaction between different features, not easy to overfit, the training speed is relatively fast, and it is easy to make a parallel method, relatively simple to implement, for unbalanced data sets, it can balance errors, and if a large part of the features is lost, the accuracy can still be maintained.

Limitation are included random forests have been shown to overfit on some noisy classification or regression problems, and for data with attributes with different values, attributes with more value

divisions will have a greater impact on the random forest, so the attribute weights produced by the random forest on this data are not credible

2.5.2 Logistic Regression

Logistic regression is a machine learning method used to solve the binary classification (0 or 1) problem, used to predict the possibility of something. (JavaTpoint, 2018) It conducts a suitable regression analysis based on a given set of independent variables to predict the dependent variable is binary. The outcome of logistic regression is either categorical or discrete value which displays in probabilistic values which between 0 and 1, instead of given exactly 0 and 1. For example, the possibility of a user purchasing a product (buy/not buy), raining today(yes/no), etc. The logistic regression is different compared to linear regression as the logistic regression is for the classification problem, while the linear regression is for the regression problem. There is using the "S" shaped logistic function in logistic regression.

Therefore, the data and the relationship between one dependent binary variable and one or more independent variables can be described and explained by logistic regression. The advantages are included simple to implement and widely used in industrial issues, the amount of calculation during classification is very small, the speed is fast, and the storage resources are low, convenient observation sample probability score, for logistic regression, multicollinearity is not a problem, it can be combined with L2 regularization to solve the problem, and the calculation cost is not high, easy to understand and implement.

Limitation are included when the feature space is large, the performance of logistic regression is not very good, easy to underfit, generally the accuracy is not too high, cannot handle a large number of multi-type features or variables well, can only deal with two classification problems (softmax derived on this basis can be used for multiple classification), and must be linearly separable, and for non-linear features, conversion is required.

2.5.3 LightGBM

LightGBM (Light Gradient Boosted Machine) is a fast and high-performance gradient boosting framework based on the decision tree algorithm. (Nitin, 2020) It splits the tree wisely in the most appropriate way, while other enhancement algorithms split wisely the depth or level of the tree instead of splitting the leaves. This is because lightGBM is used the histogram algorithm to train the decision trees in the training phase.

And LightGBM uses a node expansion method (Leaf-wise) with depth restrictions to improve model accuracy as it can reduce training error and get better accuracy. If simply use Leaf-wise, it may grow a deeper tree, and which may cause overfitting on a small data set. The advantages are included faster training speed and higher efficiency, lower memory usage, higher accuracy (compared to any other lifting algorithms), ability to process big data, support parallel learning, and sparse feature optimization based on histogram.

3. Result and Discussion

Once the evaluation of the models is completed, there is required to conduct model comparison based on the performance of those models in order to find the most accurate and good performance model to use in this study. Table 1 shows the comparison of models with the evaluation of metrics. By comparing the performance of the model, it was found that the lightGBM model is the best model to this study. Therefore, there will select the lightGBM model as the most suitable model to use in the dashboard system in order to perform the prediction on passenger satisfaction.

Table 1: Model Comparison

Evaluation Metrics	Models		
	Random Forest	Logistic Regression	LightGBM
Accuracy	0.887	0.815	0.895
Error Rate	0.113	0.186	0.105

Precision	0.823	0.745	0.833
Recall / Sensitivity	0.943	0.874	0.95
F1-Score	0.880	0.806	0.888
ROC-AUC Score	0.893	0.822	0.901
Time Taken (s)	12.49	0.74	1.84

3.1 System Testing

In the dashboard, the overall of the charts are able to show the correct information, and the add-drop function for the charts is able to work well. While the "Basic Information" button on the side bar is able to redirect the page to the basic information page. In the basic information page, the combo box for the user to select the variable is working functionally, and the "Analysis Description" link can redirect to the analysis description of basic information. The charts of the basic information page can display well once selected specific variables, and the add-drop function of the chart can be able to work. For example, the charts of gender and flight distance can clearly see that it can work functionally. The "Service Detailed" button on the sidebar is able to redirect to the service information page. The combo box in service information page is also can work probably to show the list of the services for user to select. While the "Analysis Description" link under the combo box also can work well which show the analysis description of service information page.

The charts of the specific service (Departure/arrival time convenient) that selected which are able to display functionally, and the value of each charts are correct as well as the add-drop function for the charts also can work probably. The "PSAT Prediction' on the sidebar is able to redirect to the prediction information page. If the user leaves the blanks for each input fields and proceed to the prediction, it can be able to display the notification to prompt the user to enter the value for each input fields. Once the input fields are completed, the results (The passengers should be satisfied/ neutral or dissatisfied) are able to display for the user as shown in Figure 1 and Figure 2.
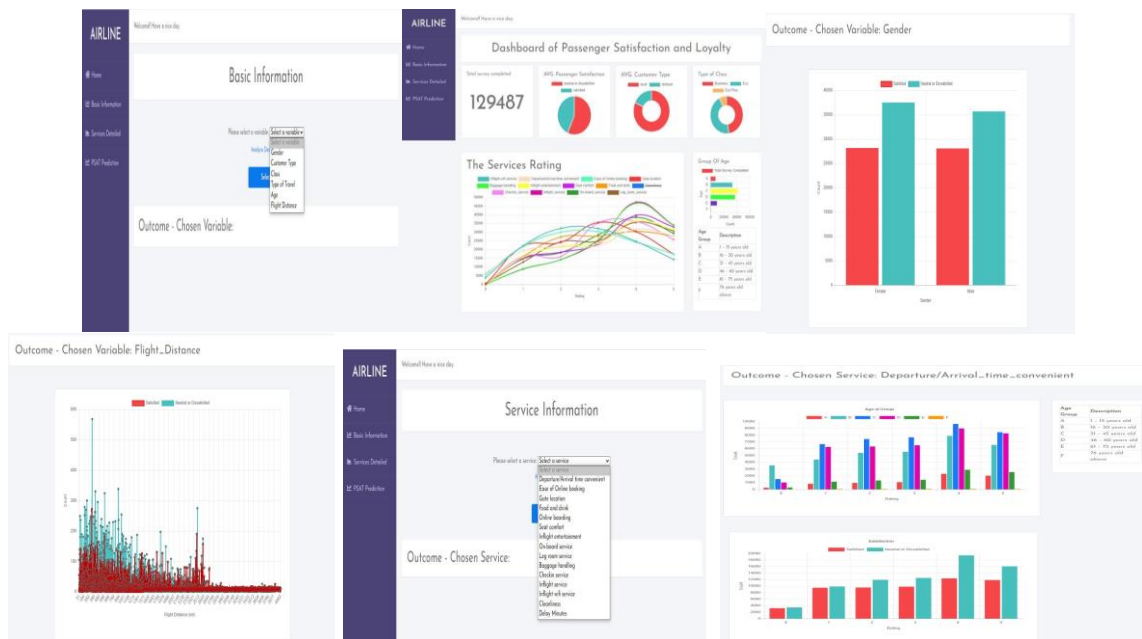

Figure 1: Performance of dashboard

Figure 2: Performance of Dashboard-Passenger satisfaction

4. Conclusion

As conclusion, clear explained on how to evaluate the model and what things will be used to evaluate the performance of the model was done. In the model evaluation, there will not only look into the accuracy of the model, but also will into the classification report (precision, recall, f1-score & support), ROC Curve & AUC, time is taken to completely build the model as well as the confusion matrix. After conducting the comparison of the models, the author had found that the most accurate and well performance model which is the lightGBM model. Therefore, the lightGBM model was selected in the development of the dashboard system in order to perform the prediction on passenger satisfaction and the overall of the dashboard is able to work functionally.

Acknowledgment

References

Abdullah, K., Manaf, N.H.A. and Noor, K.M. (2007). Measuring the service quality of airline services in Malaysia. International Journal of Economics, Management and Accounting, 15(1).

AirAsia, G. (2020). *Corporate Profile*. [online] Available at: <https://ir.airasia.com/corporate_profile.html>.

Al-Refaie, A., Bata, N., Eteiwi, D. and Jalham, I. (2014). Examining Factors that Affect Passenger's Overall Satisfaction and Loyalty: Evidence from Jordan Airport. *Jordan journal of mechanical & industrial engineering*, *8*(2).

Aydin, K. and Yildirim, S. (2012). The measurement of service quality with servqual for different domestic airline firms in Turkey. *Serbian Journal of Management*, *7*(2), pp.219-230.

Ban, H. and Kim, H. (2019). Understanding Customer Experience and Satisfaction through Airline Passengers' Online Review. *Sustainability*, 11(15), p.4066.

CFI, E., 2020. *Random Forest - Overview, Modeling Predictions, Advantages*. [online] Corporate Finance Institute. Available at: <https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>.

Gures, N., Arslan, S. and Tun, S.Y. (2014). Customer expectation, satisfaction and loyalty

relationship in Turkish airline industry. *International Journal of Marketing Studies*, *6*(1), p.66.

Jahmani, A. (2020). *THE EFFECT OF ROYAL JORDANIAN AIRLINE SERVICE QUALITY ON PASSENGERS' SATISFACTION*. 18th ed. [ebook] International Journal of Business and Society, pp.519-530. Available at: <http://www.ijbs.unimas.my/images/repository/pdf/Vol18-s3-paper7.pdf>.

JavaTpoint. (2018). *Logistic Regression In Machine Learning - Javatpoint*. [online] www.javatpoint.com. Available at: <https://www.javatpoint.com/logistic-regression-in-machine-learning>.

Klipfolio, I. (2020). *Data Dashboards. Definition, Design Ideas Plus 3 Examples*. [online] Klipfolio.com. Available at: <https://www.klipfolio.com/resources/articles/what-is-data-dashboard>.

Neelamegam, S. and Ramaraj, E. (2013). Classification algorithm in data mining: An overview. *International Journal of P2P Network Trends and Technology (IJPTT)*, *4*(8), pp.369-374.

Nitin. (2020). *Lightgbm Binary Classification, Multi-Class Classification, Regression Using Python*. [online] Medium. Available at: <https://medium.com/@nitin9809/lightgbm-binary-classification-multi-class-classification-regression-using-python-4f22032b36a2#:~:text=LightGBM%20Binary%20Classification%2C%20Multi-Class%20Classification%2C%20Regression%20using%20Python,-Nitin&text=LightGBM%20is%20a%20gradient%20boosting,uses%20tree-based%20learning%20algorithms.&text=While%20LightGBM%20can%20handle%20a,faster%20training%20speed%20and%20efficiency> .

Nor Laila Hassan, R.S., Aziam Mustafa, N.N.M. and Misalini Rajamanikam, S.M.M. (2019). Passengers Satisfaction Towards Service Quality in Malindo Airlines.

Oracle. (2020). *Classification*. [online] Docs.oracle.com. Available at: <https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#DMCON004>.

Rahim, A. (2016). Perceived Service Quality and Customer Loyalty: The Mediating Effect of Passenger Satisfaction in the Nigerian Airline Industry. *International Journal of Management and Economics*, 52(1), pp.94-117.

Saadat, M., Tahbet, T.R. and Mannan, M.A. (2018). Factors That Influence Customer Satisfaction in Airline Industry in Malaysia. *IOSR JBM*, *20*(8), pp.1-6.

Sandada, M. and Matibiri, B. (2016). An Investigation into the Impact of Service Quality, Frequent Flier Programs and Safety Perception on Satisfaction and Customer Loyalty in the Airline Industry in Southern Africa. *South East European Journal of Economics and Business*, [online] 11(1), pp.41-53. Available at: <https://www.researchgate.net/publication/305677104_An_Investigation_into_the_Impact_of_Service_Quality_Frequent_Flier_Programs_and_Safety_Perception_on_Satisfaction_and_Customer_Loyalty_in_the_Airline_Industry_in_Southern_Africa>.

Suki, N.M. (2014). Passenger satisfaction with airline service quality in Malaysia: A structural equation modeling approach. *Research in transportation business & management*, *10*, pp.26-32.

Vilkaite-Vaitone, N. and Papsiene, P. (2016). Influence of customer loyalty program on organizational performance: A case of airline industry. *Engineering Economics*, *27*(1), pp.109-116.

Yiu, T. (2019). *Understanding Random Forest*. [online] Medium. Available at: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>.

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING,
MATHEMATICS AND STATISTICS