Machine Learning Models for Airline success factors after COVID-19, Earnings prediction of farely distribution and car prices dynamics of UAE automobile industry

Muddassir Ahmed
School of Computing
National College of Ireland
Dublin, Ireland
x23138688@student.ncirl.ie

***Abstract*—In our research we have selected three datasets. First is related to airline passenger's satisfaction. Airline business has affected in Covid-19 and many airlines got restrictions. Airlines suffer loss and stop operations. When Covid-19 finished airlines were expecting normal operations. In this paper we analyze the factors of competition between different airlines and their success factors. We selected a dataset of airline passengers and perform classification models such as random forest and Logistic regression and after they compared each other. Results shows random forest algorithm get accuracy 99 percent and main factor was wifi services during flight. In Second dataset we are trying to predict income based on available attributes of the dataset. Equal wealth distribution is major concern of all over the world and developed countries has overcome this problem and many developing countries are trying to balance between their population. In this paper we have implemented two algorithms and classification use to predict income. Thirst dataset we have selected automobile industry. Due to large number of sale and purchase of automobiles, car prices always remain hot topic in developing countries because most of the people wanted to purchase used cars. In this research we were trying to check car prices depends on some attributes. We have selected a dataset of UAE cars; first we check null values of our dataset then identify duplicate values of dataset. In this research we use random forest and Linear regression for training. In random forest we got 95 percent as MSE is 0.025, 0.0378 RMSE and 0.0008 MAE. Linear Regression and random forest score are 88 percent. Dataset was divided into 80,20.**

***Index Terms*—Random Forest, Logistic Regression, Linear Regression, SVM, Naive bayes**

I. Introduction

Now service industry is replacing with manufacturing due to globally change [1]. Customer satisfaction become major aspect for increasing business growth also many researches analyze quality play vial role in industry. Airline industry has big competition all over the world [2]. Nowadays, customer analyze airline by their services which they are providing such as food services, internet services and many more. Some of the passengers prefer food and service quality of airline [3]. In this article, we implement different algorithms on dataset and check accuracy of algorithms. In our second problem, Income distribution become main concern in last few decades as population growth as increased. It became main concern of each country that wealth distribution must be equal and fair in the society. In this paper we highlight main factors which effect income distribution. For this research we selected a dataset and perform preprocessing for better use of dataset. We also use visualization for better understanding of dataset. Then we split our dataset for training and testing purpose, after that we train our models and check the results then provide conclusion. In our third problem, Transportation is backbone industry of every country for its economy, UAE is very rapid growth in automotive industry. There are each company operating in UAE and their sales are improving day by day (Rizvi, 2019). According to GCC report there are 1.49 million units were sold and in 2021 its growth was 10 percent increased so it is world largest growing market. Now a days every one want to purchase new brand car and bank and other insurance companies providing affordable rates. Prediction of car price is most demanding factor because any person how has less knowledge of car prices can get quote. [**?**].

II. Related work

A. Quality of Services

Service quality has strong relationship with customer because if service of airline match with standard of customers then service quality meet and customer will be happy [11]. It is very hard to fulfill all standards of each customer, but some of airlines are trying hard to meet all requirements of customers. Quality of services is also have big challenge of cost, if services increase then cost of ticketing will increase but customer always want to purchase lowest cost ticket with maximum services. It is very difficult of each airline after Covid 19 to give maximum services with lower price.

B. Airline company's services

Characteristics of services are not same of each airline, some are fix and other are versatile. Some airlines are providing comfortable seats, large weight of baggage, extra free meal and many more [2]. If any customer feels that some other airline providing these extra services in free of cost, definitely customer will choose maximum services provider airline.

Table. 1. Four types of service business

Service Work & Provider	Routinized	Knowledge
Integrated	Service Factory	Service Shop
Decoupled	Service Store	Service Complex

Fig. 1. Image

Different researches has done to predict the car prices and they use different approaches and their results are between 50 percent to 95 percent [4]. Machine learning technique was used to predict car prices by using decision tree, multiple regression and k nearest neighbors based on newpaper which achieve accuracy 60-70 percent [5]. German e commerce use different dataset and use different techniques and measure result by using MEA [Monburinon, et al., 2018]. Car sale prediction plays very important role in automotive industry and price prediction using machine learning algorithms [9]. Dubai old car price prediction using machine learning [3]. Now a days leasing companies offering very handsome car price packages and different banks also has car finance department. They are using different applications which is integrated with AI and ML models [7]. There are many techniques using for this purpose and some of techniques are defined in ML algorithms [10]. Different approaches are using to predict next year salary [1]. Employee of every organizations has major concern about annual bonus and salary increment [8]. Classification models can be used to predict income [12]. support vector machine, logistic regression and naïve bayes can use for better income prediction [6]. Author suggest hot coding rather than label encoding for categorical data. In Pakistan, researcher achieved higher accuracy on pak wheels automobiles website by using multiple liner regression [Noor Jan, 2017].

III. Methodology

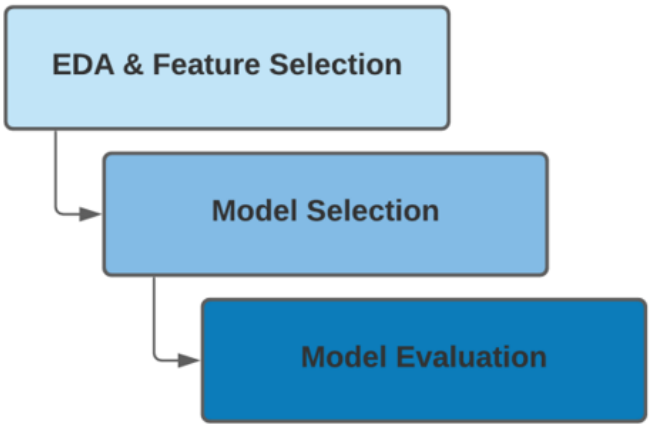

Fig. 2. Image

We have selected a dataset which contains approximately 26,000 customer's data and it is has 24 attributes. There are fourteen columns which contains the result of survey collected from passengers which using flight experience. This dataset has one to five scale of rating, one mean less satisfied and 5 is maximum satisfied. There are some zero values founds, its mean customer did not answer these questions. After cleaning the dataset our data shows below:

A. Cleaning dataset

	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	Satisfaction
0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	3	1	5	3	5	5	4	3	4	4	5	5	25	18.0	neutral or dissatisfied
1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	3	3	1	3	1	1	1	5	3	1	4	1	1	6.0	neutral or dissatisfied
2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	2	2	5	5	5	5	4	3	4	4	4	5	0	0.0	satisfied
3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	5	5	2	2	2	2	2	5	3	1	4	2	11	9.0	neutral or dissatisfied
4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	3	3	4	5	5	3	3	4	4	3	3	3	0	0.0	satisfied
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
129875	78453	Male	disloyal Customer	34	Business travel	Business	526	3	3	3	1	4	3	4	4	3	2	4	4	5	4	0	0.0	neutral or dissatisfied
129876	71167	Male	Loyal Customer	23	Business travel	Business	646	4	4	4	4	4	4	4	4	4	5	5	5	5	4	0	0.0	satisfied
129877	37675	Female	Loyal Customer	17	Personal Travel	Eco	828	2	5	1	5	2	1	2	2	4	3	4	5	4	2	0	0.0	neutral or dissatisfied
129878	90086	Male	Loyal Customer	14	Business travel	Business	1127	3	3	3	3	4	4	4	4	3	2	5	4	5	4	0	0.0	satisfied
129879	34799	Female	Loyal Customer	42	Personal Travel	Eco	264	2	5	2	5	4	2	2	1	1	2	1	1	1	1	0	0.0	neutral or dissatisfied

Fig. 3. Image

B. Dataset analysis

Data analysis showing total count of each variable. It is calculating mean of all attribute. There are id, age, flight distance, inflight services, departure/arrival time convenient, ease of online boking, gate location, food and drink, online booking, seat comfort, inflight entertainment, on board services, leg room services, baggage handling, checkin services, inflight services, cleanliness, departure delay in minutes and arrival delay in minutes.

	id	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort
count	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000
mean	64958.335169	39.428761	1190.210662	2.728544	3.057349	2.756786	2.976909	3.204685	3.252720	3.441589
std	37489.781165	15.117597	997.560954	1.329235	1.526787	1.401662	1.278506	1.329905	1.350651	1.319168
min	1.000000	7.000000	31.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	32494.500000	27.000000	414.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	64972.000000	40.000000	844.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000
75%	97415.500000	51.000000	1744.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	5.000000
max	129880.000000	85.000000	4983.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Fig. 4. Image

This analysis shows statistical view of data and it shows some of attributes are useful and some are irrelevant. We will remove extra/unnecessary attributes and apply label encoding. Removing departure features. Final form of our data shows below:

Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes
129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000	129487.000000
3.358067	3.383204	3.351078	3.631886	3.306239	3.642373	3.286222	14.643385	15.091129
1.334149	1.287032	1.316132	1.180082	1.266146	1.176614	1.313624	37.932867	38.465650
0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2.000000	2.000000	2.000000	3.000000	3.000000	3.000000	2.000000	0.000000	0.000000
4.000000	4.000000	4.000000	4.000000	3.000000	4.000000	3.000000	0.000000	0.000000
4.000000	4.000000	4.000000	5.000000	4.000000	5.000000	4.000000	12.000000	13.000000
5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	1592.000000	1584.000000

Fig. 5. Image

IV. Exploratory Data Analysis

A. Airline dataset

For visualization we need to identify the features where we want to apply modeling. We identify the feature for visualization and then deleted some features gender, age, gate location.

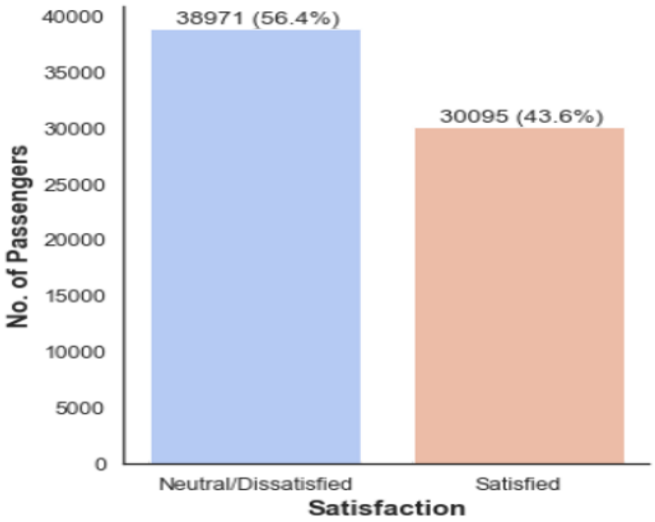

Fig. 6. Image

This graph show satisfaction of customers. There are two types of customers in our dataset. One type is personal traveler and second one is business class traveler. Dissatisfied personal traveler are 19177 and business traveler are 19794. Satisfied personal traveler are 1709 and business traveler are 28386. It shows that business class is higher satisfied as compare to other class.

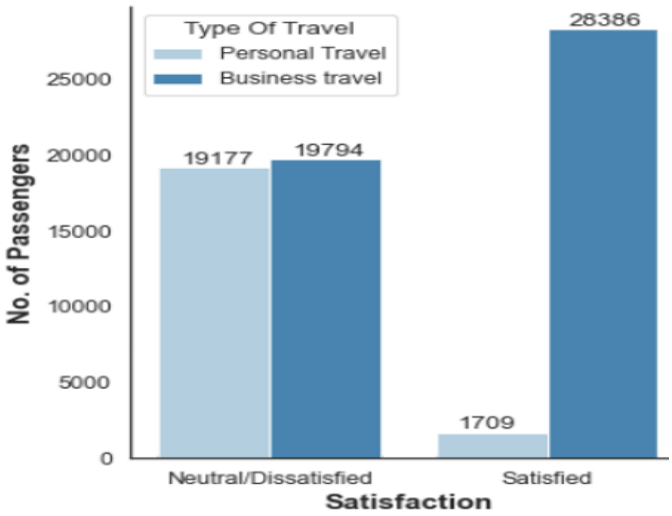

Fig. 7. Image

Disappointed passengers are 56.4 percent and happy customers are 43.6 percent. As we know first time of traveler has neutral opinion because that customer could not compare experience with any other at first time. Regular customer can give different feedback as compare to new customer because frequent traveler had experience with multiple airlines and know all services providing by different airlines.

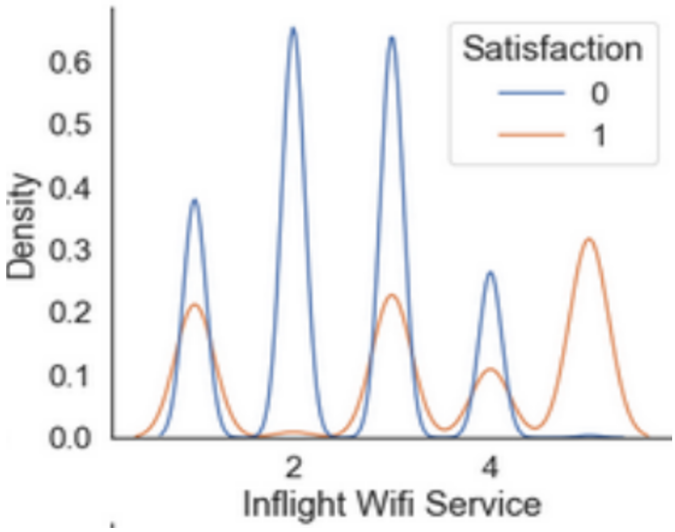

Fig. 8. Image

We train our model into 80 percent dataset and test into 20 percent dataset.

Logistics Regression (C = 0,04)
Random Forest (Max Depth= 17)

Overall random forest model is best as it has 0.97 precision, and 0.94 recall with 0.99 AUC.

Logistic Regression	AUC	0.957475
Logistic Regression	Precision	0.882574
Logistic Regression	Recall	0.868921

Fig. 9. Image

Logistic regression is used for classification problems. It is always predicting the answer of outcome. In our dataset we have found logistic regression for AUC value is 0.957475 and got precision value 0.8825 and recall value is 0.8689. Table shows the result:

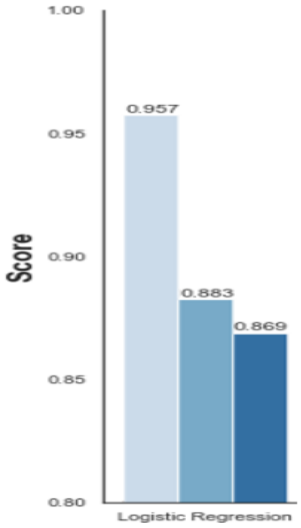

Fig. 10. Image

Random Forest	AUC	0.992917
Random Forest	Precision	0.972766
Random Forest	Recall	0.935849

Fig. 11. Image

Business class customers are less than personal class customers in random forest. Personal class shows 0.993 and business class shows 0.936.

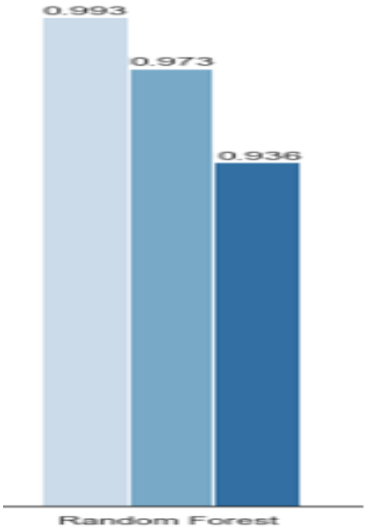

Fig. 12. Image

Precision, recall and F1 score machine learning evaluations for measuring accuracy of the model. Precision is also called positive predictor. It is used to measure positive prediction values. It is calculated by ratio of true prediction to sum of true positive and false positive.

Recall is known for true positive and it is also called sensitivity. It is used to measure ability of model for capture all positive instances of the dataset.

F1 score use to provide a balance measure of the model performance by considering both false positive and false negative.

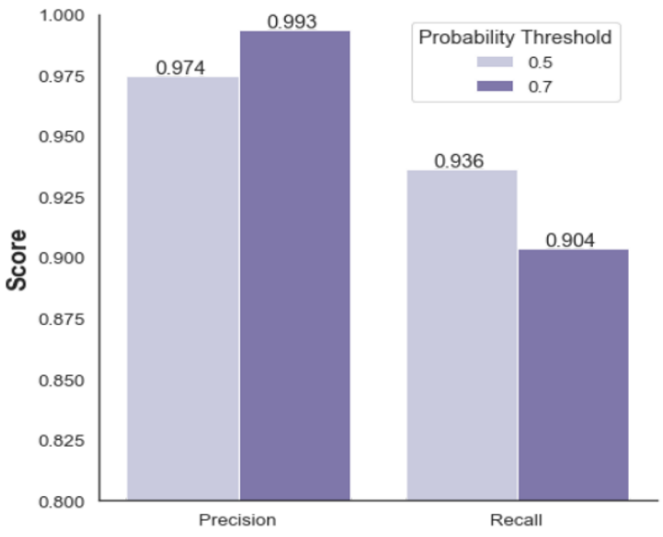

Fig. 13. Image

	Scoring	Threshold	Score
0	Precision	0.5	0.974476
1	Precision	0.7	0.993353
2	Recall	0.5	0.936335
3	Recall	0.7	0.903746

Fig. 14. Image

Random forest default probability is defining as 0.5 and precision were improving 0.97 to 0.99 when random forest probability is defined as 0.7. As we have selected random forest with 0.7 threshold.

B. Income Dataset

Income dataset is for classification which predict income that is less than fifty thousand and greater than fifty thousand. First rows of dataset shows in diagram, it contains numerical values and categorical data as well. This dataset show only data, and attribute names are missing. We have to add attribute name in our dataset

```
  V1                V2     V3         V4 V5                  V6
1 39         State-gov  77516  Bachelors 13       Never-married
2 50  Self-emp-not-inc  83311  Bachelors 13  Married-civ-spouse
3 38           Private 215646    HS-grad  9            Divorced
4 53           Private 234721       11th  7  Married-civ-spouse
5 28           Private 338409  Bachelors 13  Married-civ-spouse
6 37           Private 284582    Masters 14  Married-civ-spouse
      V9     V10  V11 V12 V13           V14    V15
1  White    Male 2174   0  40 United-States  <=50K
2  White    Male    0   0  13 United-States  <=50K
3  White    Male    0   0  40 United-States  <=50K
4  Black    Male    0   0  40 United-States  <=50K
5  Black  Female    0   0  40          Cuba  <=50K
6  White  Female    0   0  40 United-States  <=50K
> shape <- dim(data)
> # Display the shape
```

Fig. 15. Image

After addition of attributes names of each column. Variable name we can use for further processing of our dataset.

```
  age       workclass fnlwgt  education education_num      marital_status         occupation
1  39       State-gov  77516  Bachelors           13       Never-married       Adm-clerical
2  50 Self-emp-not-inc  83311  Bachelors           13  Married-civ-spouse    Exec-managerial
3  38         Private 215646    HS-grad            9            Divorced  Handlers-cleaners
4  53         Private 234721       11th            7  Married-civ-spouse  Handlers-cleaners
5  28         Private 338409  Bachelors           13  Married-civ-spouse     Prof-specialty
6  37         Private 284582    Masters           14  Married-civ-spouse    Exec-managerial
    relationship  race    sex capital_gain capital_loss hours_per_week native_country income
1  Not-in-family White   Male        2174            0             40  United-States  <=50K
2        Husband White   Male           0            0             13  United-States  <=50K
3  Not-in-family White   Male           0            0             40  United-States  <=50K
4        Husband Black   Male           0            0             40  United-States  <=50K
5           Wife Black Female           0            0             40           Cuba  <=50K
6           Wife White Female           0            0             40  United-States  <=50K
```

Fig. 16. Image

After applying proper naming of each column we are now checking null values of our dataset.

```
          age     workclass        fnlwgt     education  education_num marital_status
            0             0             0             0             0              0
   occupation  relationship          race           sex  capital_gain  capital_loss
            0             0             0             0             0              0
hours_per_week native_country        income
```

Fig. 17. Image

C. Car Sale Dataset

A dataset was selected which contains approximately 10,000 rows of data and has 20 attributes.

Below diagram show brand wise car data. In our dataset Nissan car are maximum as compare to other cars. Nissan car is showing maximum due to many its factors. It is depending on price of the car, features of the car, reliability of the car, petrol consumption, comfortability and many more feature. Customer always focus in all attributes before buying the cars.

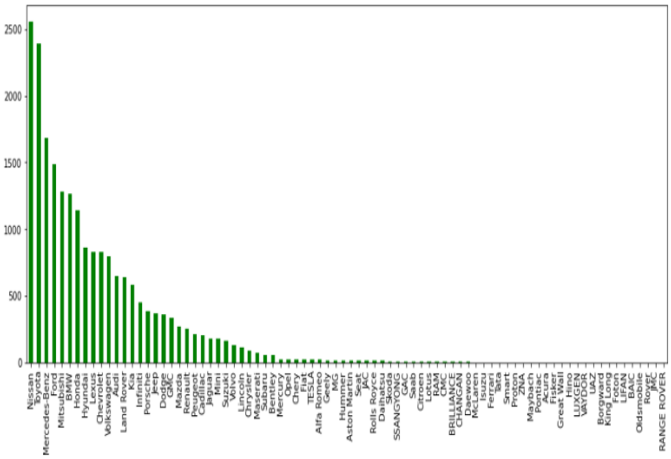

Fig. 18. Image

In below diagram shows white car is most occurred in dataset and black is at second number and so on. White car

most in use due to its resale and cleanliness. Customer always think about resale of the car before buying. There are more than 6000 customers who prefer white car buying and it is also very easy to clean white car. Black car people like due to its shine and beauty, so that's why it is at second number for resale.

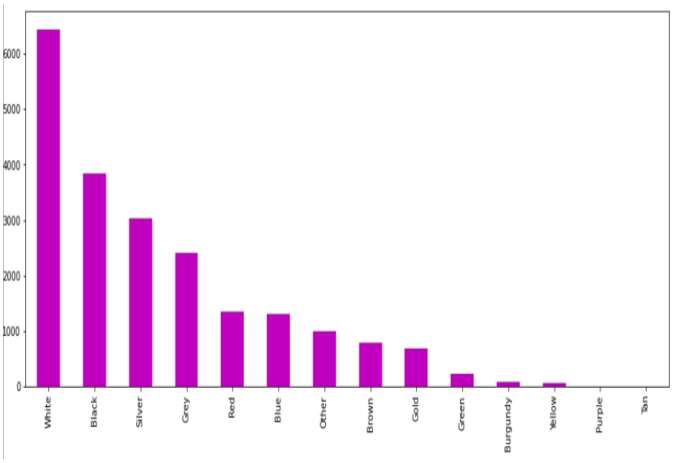

Fig. 19. Image

D. Preprocessing

In this step, first checked null values of dataset then applied label encoding after checking. Removed all unnecessary attributes which are not usable for our research question. This is very crucial part of any research because preparation and well understanding of dataset related to problem is very important before applying modelling into data. Skewness and Kurtosis was performed and outliers were identified. Then outliers were removed from dataset. Box plot outliers of dataset so year wise outliers shows as below:

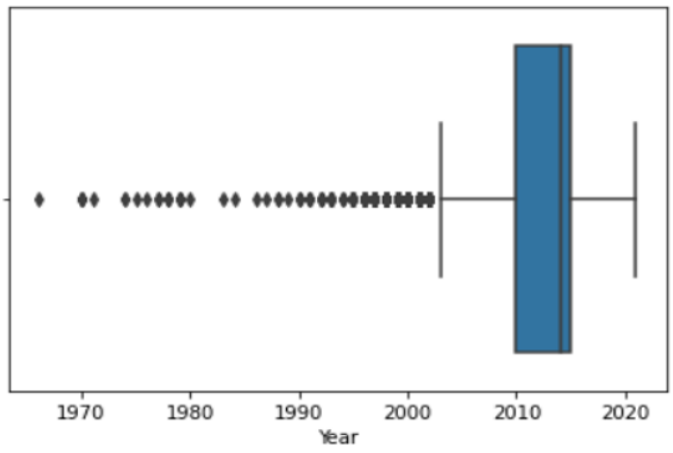

Fig. 20. Image

E. Modelling

Applied two algorithms on dataset e.g. random forest and Linear regression for training purpose. This is all done for prediction of car prices and its values of used cars in UAE. In first phase we trained our algorithms on dataset and test to analyse the result. Both models were evaluated by MSE, RMSE and MAE. We analysed random random forest regressor performed well.

Model Evaluation Parameters:

Following parameters can be used to evaluate model

1: Mean Square Error (MSE): The Mean Squared Error measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss. Mean square error is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a function.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} |yi - y|^2$$

Fig. 21. Image

2: Root Mean Square Error (RMSE): It is used to evaluate quality of most frequent predictions. It is used for prediction and actual values by using square root.

3. Mean Absolute Error (MAE):

It shows average between actual and prediction in dataset.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |yi - y|$$

Fig. 22. Image

V. Results and Model Evaluation

After taking the result of trained dataset on 80 percent, then test the results on remaining 20 percent. Result of AUC is 0.99, precision is 99.1 percent and recall is 91.2 percent.

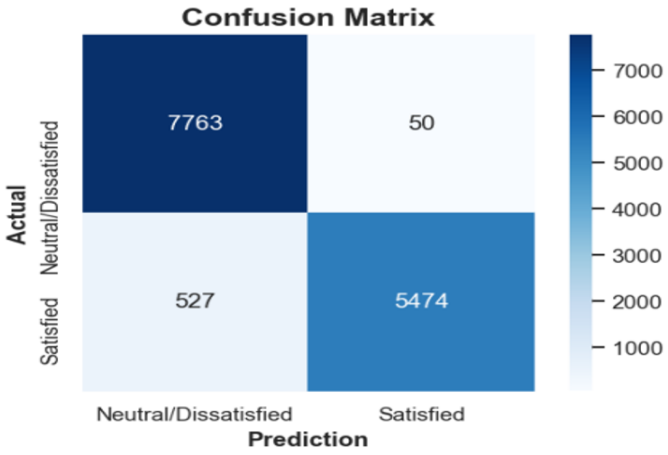

Fig. 23. Image

Passengers shows satisfaction and it indicates 99.1 percent.

	features	importance
2	Food And Drink	0.015048
9	Checkin Service	0.026856
10	Inflight Service	0.027592
8	Baggage Handling	0.030065
11	Cleanliness	0.031272
1	Ease Of Online Booking	0.035996
6	On-board Service	0.040355
12	Customer Type_Returning Customer	0.045322
7	Leg Room	0.052875
4	Seat Comfort	0.061945
5	Inflight Entertainment	0.072244
14	Class_Economy	0.110338
13	Type Of Travel_Personal Travel	0.116912
0	Inflight Wifi Service	0.151804
3	Online Boarding	0.181375

Fig. 24. Image

In our income dataset after removal of irrelevant variables of dataset then it is divided for training and testing purpose. We selected 70 percent of dataset for training purpose and remaining 30 percent will use for testing purpose. Naive Bayes and SVM model use for training purpose. Then making prediction on dataset, after assessing of performance evaluated. After that accuracy is measured

```
Confusion Matrix and Statistics

          Reference
Prediction    1    2
         1  676   14
         2 6297 1961

               Accuracy : 0.2947
                 95% CI : (0.2853, 0.3043)
    No Information Rate : 0.7793
    P-Value [Acc > NIR] : 1

                  Kappa : 0.042

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.09695
            Specificity : 0.99291
         Pos Pred Value : 0.97971
         Neg Pred Value : 0.23747
             Prevalence : 0.77928
         Detection Rate : 0.07555
   Detection Prevalence : 0.07711
```

Fig. 25. Image

```
                  Kappa : 0.042

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.09695
            Specificity : 0.99291
         Pos Pred Value : 0.97971
         Neg Pred Value : 0.23747
             Prevalence : 0.77928
         Detection Rate : 0.07555
   Detection Prevalence : 0.07711
      Balanced Accuracy : 0.54493

       'Positive' Class : 1
```

Fig. 26. Image

SVM model results

```
Confusion Matrix and Statistics

          Reference
Prediction    1    2
         1 6806 1355
         2  167  620

               Accuracy : 0.8299
                 95% CI : (0.822, 0.8376)
    No Information Rate : 0.7793
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3697
```

Fig. 27. Image

```
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9761
            Specificity : 0.3139
         Pos Pred Value : 0.8340
         Neg Pred Value : 0.7878
             Prevalence : 0.7793
         Detection Rate : 0.7606
   Detection Prevalence : 0.9120
      Balanced Accuracy : 0.6450

       'Positive' Class : 1
```

Fig. 28. Image

In our third dataset we have implemented two models. Random Forest

This model is used for classification and regression purpose. Many researchers suggest random forest is best as compare to decision tree. Below result shows MSE, MAE and RMSE.

```
**********Evaluation Scores of Random Forest Regressor**********
Score of Random Forest Regressor is :  0.9587547249635755
MSE of Random Forest Regressor is :  0.025504
MAE of Random Forest Regressor is :  0.000844
RMSE of Random Forest Regressor is :  0.03784
```

Fig. 29. Image

Linear Regression

In this we use two variables, one is dependent and other one is independent variable. Car prices are dependent on year of model. If the model of car is increase, then car price will be increase and vice versa. In this dataset linear regression is not much suitable. Model was trained and then RMSE, MSE and MAE used to check model.

```
**********Evaluation scores of Linear Regression**********
R-squared training data:  0.855024
Mean Square Error:  0.135463
Mean Absoulte Error:  0.11563
Root Mean Square Error:  0.04213
```

Fig. 30. Image

Results comparison

SR#	Algorithm	Accuracy	MSE	MAE	RMSE
1	Random Forest Regressor	0.95	0.025	0.0008	0.03
2	Linear Regression	0.85	0.13	0.11	0.04

Fig. 31. Image

VI. Conclusion

In this research we built classification model which identify factors of customer satisfaction for airline. Random forest and logistic regression model use for this classification. Random forest (AUC 0.99, Precision 0.97, Recall 0.93) gives better results as compare to logistic regression (AUC 0.95, Precision 0.88, 0.86). Majority of the customers were satisfied by in flight wifi services. Business class of passengers were already getting these services and economy class were highly satisfied by this inflight wifi services. Furthermore, passengers were satisfied by online booking and comfort level while traveling. In income dataset we used naïve bayes and svm model. SVM (AUC 0.89, accuracy 84 percent) gives the better result as compare to naïve bayes (AUC 0.67, accuracy 74 percent). We collected UAE cars dataset and trained our two model random forest and linear regression. Random forest accuracy we calculated 95 percent which is best as compare to other.

References

[1] Navoneel Chakrabarty and Sanket Biswas. A statistical approach to adult census income level prediction. In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pages 207–212. IEEE, 2018.

[2] Jeewon Choi, Hyeonjoo Seol, Sungjoo Lee, Hyunmyung Cho, and Yongtae Park. Customer satisfaction factors of mobile commerce in korea. *Internet research*, 18(3):313–335, 2008.

[3] Prashant Gajera, Akshay Gondaliya, and Jenish Kavathiya. Old car price prediction with machine learning. *Int. Res. J. Mod. Eng. Technol. Sci*, 3:284–290, 2021.

[4] Taqwa Hariguna, Wiga Maulana Baihaqi, and Aulia Nurwanti. Sentiment analysis of product reviews as a customer recommendation using the naive bayes classifier algorithm. *International Journal of Informatics and Information Systems*, 2(2):48–55, 2019.

[5] Rahim Hussain, Amjad Al Nasser, and Yomna K Hussain. Service quality and customer satisfaction of a uae-based airline: An empirical investigation. *Journal of Air Transport Management*, 42:167–175, 2015.

[6] Alina Lazar. Income prediction via support vector machine. In *ICMLA*, pages 143–149, 2004.

[7] Mariana Listiani et al. Support vector regression analysis for price prediction in a car leasing application. *Unpublished. https://www. ifis. uni-luebeck. de/˜ moeller/publist-sts-pw-andm/source/papers/2009/list09. pdf*, 2009.

[8] Mrs S Nagavali, R N Lakshmi Sruthi, Md Arifa, M Revathi, and K Durga Sravani. Income prediction using machine learning. *Journal of Engineering Sciences*, 14(07), 2023.

[9] Saamiyah Peerun, Nushrah Henna Chummun, and Sameerchand Pudaruth. Predicting the price of second-hand cars using artificial neural networks. In *The Second International Conference on Data Mining, Internet Computing, and Big Data*, number August, pages 17–21, 2015.

[10] Sameerchand Pudaruth. Predicting the price of used cars using machine learning techniques. *Int. J. Inf. Comput. Technol*, 4(7):753–764, 2014.

[11] K Singaravelu and VP Amuthanayaki. A study on service quality and passenger satisfaction on indian airlines. *Journal of Commerce and Trade*, 12(2):106–115, 2017.

[12] Ghatkesar Yamnampet. Comparative analysis of classification models on income prediction. *International Journal on Recent and Innovation Trends in Computing and Communication*, 5(4):451–455.