

Human Satisfaction and Suicide Rate Analysis

Kishore Lakshmanan

x20253583

MSc in Data Analytics

Data Mining and Machine Learning 1

National College of Ireland, IRELAND

Abstract—with the development of internet technologies, business need a proper feedback of their product from the people to understand the deliverable standards. So, there are many companies which focus on the opinion and suggestions from the humans to improve their business based upon the supplied answer provided by them. In some way, the spread of bad comments and feedback about the product in internet causes worst impression against the companies. Because of this reason, company needs a proper feedback from all customers after every transaction. Due to this feedback process, large data is accumulated to work upon it. Another important case would be the raising of more number of suicide without knowing any reason. Even though we have several new technologies, we are shamed to not having the proper insights to show the victim of suicide. In this paper, we are examining different Machine Learning algorithms to predict the human satisfaction and suicide rate to overcome all those issue. The performance matrices is analysed through various methods and graph to showcase the model summary and accuracy of each and every model.

Index Terms—Internet technologies, business, feedback, suicide, Machine Learning algorithms, human satisfaction and suicide rate, performance, accuracy

I. INTRODUCTION

In the internet era, most people are accustomed to sharing their reviews in a public arena for the benefit of others. The business also wants people to give more feedback on their product to analyse and rectify their mistakes to produce more profits. There are many factors to note in these area whether the opinion provided by the customer is acceptable or not and sometimes it might be false one. The Machine Learning techniques, Big Data, IoT, Cloud and Artificial Intelligence are all emerging as new technologies. These new technologies have shown to be effective in resolving difficult human issues [1]. So the most important thing is to choose the right elements to help our model produce better outcomes. In this assignment, we'll look at several machine learning algorithms to determine how well they function and what downsides they have in a certain field. We'll look at the datasets utilized in this analysis, as well as the approaches used to clean and modify our data, and compare the best model with a variety of machine learning algorithms to arrive at the best result.

A. Datasets

In this project we are using 3 datasets to perform our analyzes. All the datasets are having more than 10K records with than 10 columns and they are related in some ways (gender and age) to compare the better machine learning models. All

the data sets are retrieved from kaggle and were publicly available to get use of it. These datasets are structured one. The following table 1 describes our complete datasets. First data set is about airline passenger satisfaction which consists of gender, customer type, age, travel type, class, flight distance and some other factors with satisfaction score. Second data-set is about booking satisfaction in hotel industry which consists of id, age, Gender, Purpose of travel, Type of travel and booking and other attributes with their feedback score. The third data set is master data with suicide rate which consists of Year, sex, age, suicide number, population, suicides rate per 100k population, Country, HDI for year, GDP per capita etc. With these three data sets we are going to find below research questions as use cases.

Metadata	Datasets		
	Airline_data	Hotel_data	Suicide_data
Name of the Dataset	Airline_Passenger_satisfaction	Hotel_Booking_satisfaction	master
Source	https://www.kaggle.com/	https://www.kaggle.com/	https://www.kaggle.com/
Repository	United States	Europe	World Bank
Access Rights	Public	Public	Public
Records	1,29,881	1,03,905	27,821
Format	Structured (CSV file)	Structured (CSV file)	Structured (CSV file)
Created Date	20-02-2020	05-06-2020	01-12-2018
Modified Date	20-02-2021	05-06-2020	01-12-2018

Fig. 1. Description of Data

B. Research Questions

The below stated use cases are the research questions which is used to analyse and compare the different machine learning algorithms in this project. Our research question is how is the satisfaction and suicide rate based on gender and we see some of the use cases involved below.

- 1) RQ1: Predict the customer satisfaction on travel from airline and hotel industry.
- 2) RQ2: Predict the customer satisfaction rate based on gender and other factors.
- 3) RQ3: Predict the suicide rate based on gender and age.

C. Motive of this research paper

The main aim of the project is to help government especially on the business improvement and healthcare. All the 3 datasets are related to each other to make the better prediction based on the common parameter such as gender and age. We can easily

estimate the prediction based on the outcome of the each and every model used here. We have implemented various machine learning models to find whether which algorithm will be best suitable for that prediction using the performance notation. Some of the model will work best and some will not. So our aim of this research is to find out the appropriate model which is well suited for our prediction analyzes.

II. RELATED WORK

User satisfaction is the important aspect of each and every business [2]. Will see the previous work on different research papers to differentiate its benefits and drawbacks of their findings.

A. Customer Satisfaction on Airlines

C. Baydogan [3] is interested to solve various NLP problems. So he research with various machine learning methods to analyse the performance. But his study only shows the 10-fold cross validation. T. Tusar [4] were used bag-of-words techniques to measure various machine learning algorithm and he haven't justified the proper performance notation to justify it. V. Varsha [5] was developed his model with deep networks and neural networks to help the airline institution and the rescheduling of flights. His method can provide an accurate prediction of whether or not an aircraft arrival delay is to be expected. But he is not able to give insights about the customer satisfaction summary. T. Parbat [6] suggested hybrid machine learning (ML) models to extract data from these qualities. This work currently presents regression approaches for machine learning to estimate prices. When compared to the existing method, this study effort shows that the suggested algorithm's accuracy is extremely high 98 percent and that the execution time (sec) is quite short. He only states the model for price prediction and not for customer satisfaction.

H. Basker [7] had gathered data on the best arrival time of flight, different airport in 43 major cities in the country, and queue time waiting time of the customer ratings. Additionally, he used the Logistic Model Trees machine learning approach to predict customer happiness based on parameters such as an current report of an airport, the number of flights, ranking of the current flight performance, delays in the flight arrival by its average, and waiting time of the customer. He turned customer satisfaction to happiness based approach which will give less performance of analysis. V. Natarajan [8] used a logistic regression model is utilized to determine the state of the delay, as opposed to a decision tree model for evaluating the delay's performance. The suggested study compares the efficacy of the decision tree algorithm with that of logistic regression. The results of this simulation show probable delays in major airports based on time, day, weather, and other factors. Only delays have been analyzed with his study.

B. Customer Satisfaction on Hotel Booking

B. Agorakis [9] has a creation of a new schema for manipulating hotel reviewer comments based on NoSQL databases, as well as the relevant approach based on the data mining

and neural networks for the hotel dataset for his analysis. He wasn't goes deep into the machine learning models that's the drawback of his model. A. Eid [1] provide information on the function of machine learning and its related technologies in the hotel business. However he doesn't show the performance evolution of all the models.

H. N. T. Thu [10] provide an online review system for measuring guest satisfaction in Vietnamese hotels. The happiness of guests was assessed using natural language processing technology and machine learning based on six factors. He only analyses his review based on those six criteria separately. J Augustin [11] proposed Machine-learning approaches and artificial neural networks optimized with genetic algorithms which were resulting in a 98 percent cancellation rate. The suggested technique enables us to determine not just cancellation policy and the count, also which person are always about to cancel the bookings more. A Nuno [12] approach is same like a previous author Augustin's approach. But he only find that who likely to cancel the bookings by his research.

C. Study on Suicide prediction

S. Renjith [13] suggests using a combination of computer neural networks model to assess the public postings for analysing any suicide intention occurs or not. But Social media post might be fake in some cases. So it is not a well suggested model. B. I. Seraphin [14] model composes tree based classification which has an 91 percent accuracy and the MLP classification model provide them 90 percent accuracy. N. Norden [15] goal is to use machine learning algorithms to made a model which is useful for predicting the attempts of suicide among people if they have any kind of problems. It also includes a comparison of models based on predictions and shows the model for distinguishing the person suicide with a problem and without a problem. Python is used in 75 percent of the assessed studies to develop machine learning-based models by C. S. Gema [16]. He researched many paper and conclude with svm model has more accuracy. The use of Bayesian estimation provides for a larger degree of causality prediction than ordinary spline regression models, as seen by the suggested model's comparably least mean square value of every estimations which is said by T. Ananya [17]. T. Agarwal [18] made a comparison of three machine learning methods for suicide prediction: logistic regression, random forest, and Nave-Bayes. The goal of his study is to figure out how successful these algorithms are at preventing future suicides. But within 3 model we are not able to conclude the best fit.

III. METHODOLOGY

This research has been developed to understand the positive and negative aspects of machine learning algorithms. For that we have selected 3 datasets with the aim of finding the research questions with performance evolution of machine learning models. The important part of the project before applying machine learning model is to prepare and clean the data. So we have implemented the CRISP-DM technique to all

the datasets for getting a proper knowledge. Based on the Cross Industry Standard Process for Data Mining process, we need to know the business and prepare the data, after which algorithmic model can be constructed, evaluated before any data preparation. In this section we see the various CRISP-DM process, data preparation, Machine learning algorithms and modelling progress used in the project are shown in detail.

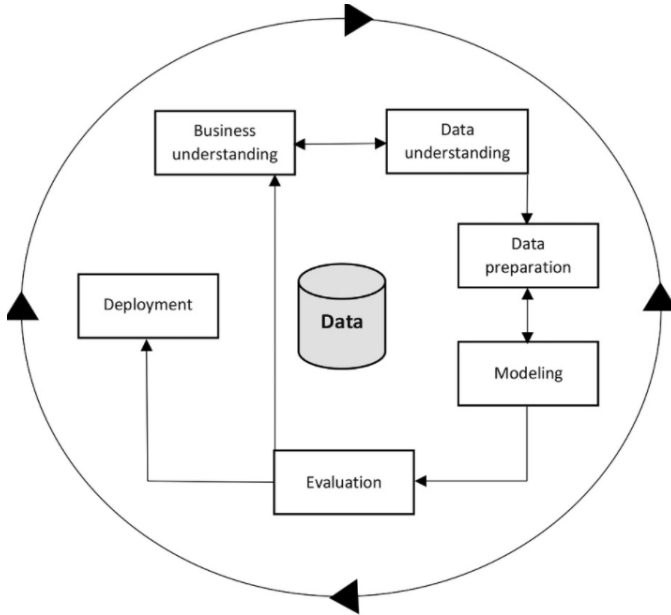


Fig. 2. CRISP-DM

A. CRISP-DM Processes

The CRISP-DM is the process of making the data useful to the business by doing several works in between them such as understanding the data, preparation of data, modelling progress, evaluation and then deployment for making the criteria satisfied. The important stage of this process is understanding and preparing the data which will take more time. The data needs to be deployed once it is properly discovered. In this section, we are going to validate the data by doing several test steps by cleaning the data which has missing values and transforming the data to get proper insights.

The following figure 3 shows the more detailed concept of data mining which have performed in our research paper.

- 1) **Business Requirement:** Understanding the business is focused in this phase by analysing the research's aim. In this stage we are analysing our 3 datasets and find the exact business requirement by setting up the proper research question. Excluding this activity, the other tasks are standard project management that applies to all research which are analyse business aim, tackle situation and predicting data mining goals and draft the plan.
- 2) **Understanding of Data:** In the data understanding stage, it strengthens the previous part of business requirement by aiming to identify, do proper collection of data by analysing its source. The stages in this process are

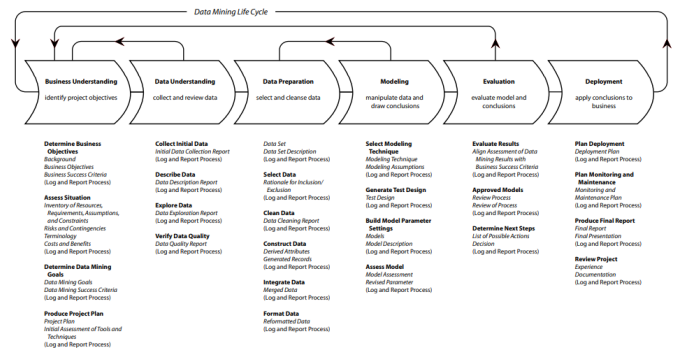


Fig. 3. Phases in CRISP-DM

collecting initial data, describe and explore the data and then verify its quality. We will discuss in detail in below sections.

- 3) **Preparation of Data:** The quality of data is being checked at this point, like missing values to be rectified / removed, make away from the unwanted variables from the dataset. Data preparation is the major part of the CRISP-DM process. It takes around 3/4 of the project time into data preparation stages. In this research, we have to convert the string to categorical variables to some of the dataset and cleaned them by removing null records, adjusting column and converting string to categorical variable etc. We will discuss in detail in below sections.
- 4) **Modelling:** In this phase, We need to start building the machine learning model by using various ML algorithms. It consists of several steps to achieve our model. We need to split the data into test and train, select the suitable model and then start building the model. After that we need to analyse whether the model build is best fitted to our research or not. If it is not then we need to iterate with different model or make changes to it to get the good model. CRISP-DM method will work best by iterating the model for business. We will discuss in detail in below sections.
- 5) **Evaluation of Model:** The evaluation of model is makes the model performance analyses for our research question. The model which fits better would be picked for the business deployment stage. This phase has several stages like performance evaluation, review the process and analysing the next process.
- 6) **Project Deployment:** This stage is comparatively more easier than all other steps like creating report about the model used and data mining stages gone through are filed for the business depends upon the objectives. The model won't be good if the research are not good enough for the customer taught. There are several stages in this deployment phase also as last phase of our data mining process. It consists of making final report and review the project once it is deployed. After its getting moved to

the production in the company, it needs to be monitored occasionally.

B. Understanding and Preparation of Data

We have selected three related data set based on age and gender in different sectors like airplane, hotel and medical field. All of them having different business requirement. So based upon the research questions, we need to look at the data and prepare them to have a better data for model selection. For airplane dataset, the business requirement is the find the satisfaction rate of customer based on gender by using several parameters like type of travel, customer class, seat comfort and baggage handling etc. The progress on the hotel business is to predict the satisfaction rate of customer based on gender with several attributes like how will be the hotel wifi score, hotel location, food and staff etc. For Suicide data, we need to analyse which column should we consider like country, age, sex, year, population and suicide per 100k population etc. for the business requirements. We have to predict the suicide rate based on the gender, age and population. Because we need a lot of info about the column was used, which have the missing value which is to be deleted. Additionally, because the dataset was designed to be built in a basic and quick manner, it is not needed to frequently access the old data, resulting in a more perfect operation in means of speed and resources. For this sort of study, the quantity and type of variables available are critical.

In the Preparation phase, first we need to check the missing value and to fill with mean / median or remove them if not necessary for our analysis. We will look at the missing values of each dataset in order.

Missing data in Airline Dataset: The following picture 4 illustrates the missing records in each column which will be useful for data cleaning.



Fig. 4. Missing values in airline data

There are 393 missing values in arrival delay in minutes column and have x column which is not required as it contains only serial numbers. We can also see the missingness map using missmap function in below figure 5

After Analysing the missing values, we have to do selection of variables by checking outliers and correlation between them. The below diagram 6 shows the correlation plot of airline data.

Missing data in Hotel Dataset: The following picture 7 illustrates the missing records in each column which will be useful for data cleaning.

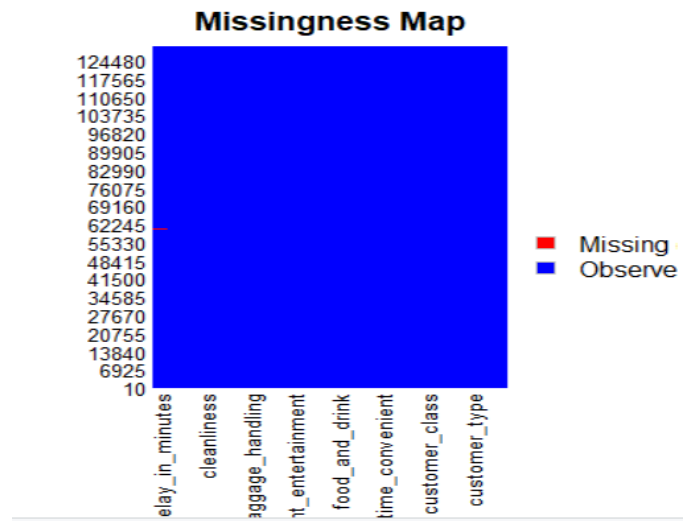


Fig. 5. Missingmap fuction for airline data

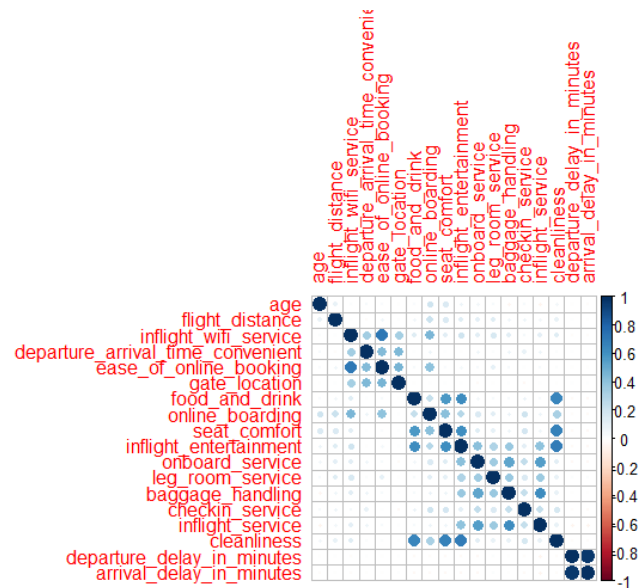


Fig. 6. Correlation Plot for airline data

There are no missing value found in hotel booking data set. We can also see the missingness map using missmap function in below figure 8

Same like previous one, we have to do selection of variables by checking outliers and correlation between them after Analysing the missing values. The below diagram 9 shows the correlation plot of hotel data.

Missing data in Suicide Dataset: The following picture 10 illustrates the missing records in each column which will be useful for data cleaning.

There are 19456 missing values in HDI for year column and if we remove those rows means we have to lose more than 10K records n our dataset. So we have removed the HDI for year column to analyse the better prediction. We can also see the

```
sapply(hotel,function(x) sum(is.na(x))) #checking for null attributes
      id      Gender      Age      Type.of.Travel      Type.of.Booking
      0          0          0          0          0
purpose_of_travel      0      Departure.Arrival..convenience      Ease.of.Online.booking
      0          0          0          0
Hotel.wifi.service      0      Food.and.drink      Stay.comfort
      0          0          0          0
Hotel.location      0      checkin.checkout.service      other.service
      0          0          0          0
Common.Room.entertainment      0      satisfaction
      0          0
Cleanliness      0
```

Fig. 7. Missing values in Hotel data

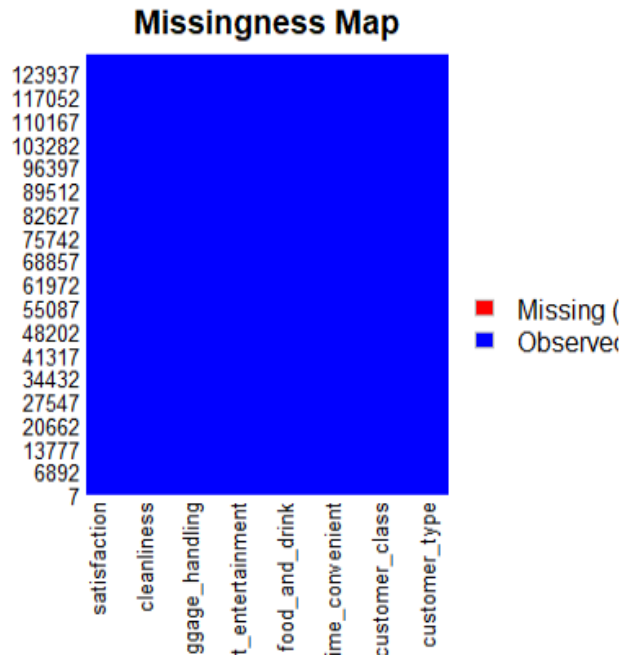


Fig. 8. Missingmap fuction for hotel data

missingness map using missmap function in below figure 11

Like same as before datasets, we have to do selection of variables by checking outliers and correlation between them after Analysing the missing values. The below diagram 12 shows the correlation plot of suicide data.

C. Machine Learning Algorithms

Machine Learning algorithms are broadly classified into 3 types based on the outcome variable. They are supervised learning, unsupervised learning and reinforcement learning. Lets see about supervised learning first. Training data is data that has a known label or outcome, such as spam/not-spam or a current property price. A model is created through a training procedure that requires it to generate predictions and corrects it when those predictions are incorrect. The model is trained until it reaches the appropriate degree of accuracy on the training data. Classification and regression are two examples of supervised learning issues. Logistic regression and decision trees are two examples of algorithms. Lets see about the unsupervised learning. The input data is not tagged and has no known outcome. Deducing structures from the incoming data is used to create a model. This might be for the purpose of extracting broad rules. It might be done using a mathematical

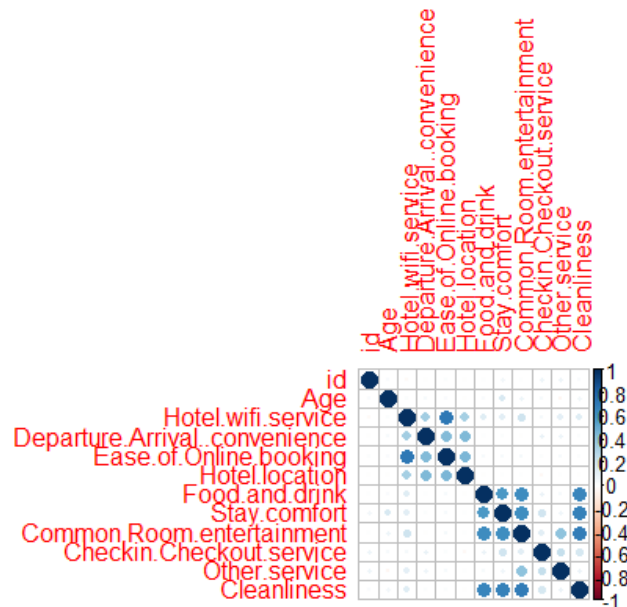


Fig. 9. Correlation Plot for hotel data

```
> sapply(suicide,function(x) sum(is.na(x)))
      country      year      sex      age      suicides_no
      0          0          0          0          0
population      suicides.100k.pop      country.year      HDI_for_year
      0          0          0      19456      gdp_for_year....
gdp_per_capita...      generation      0
```

Fig. 10. Missing values in suicide data

procedure to systematically minimize redundancy, or it could be done manually sorting data by similarity. Clustering, dimensionality reduction, and association rule learning are examples of unsupervised learning tasks. The Apriori algorithm and K-Means are two examples of algorithms. Lets see about the semi structured learning. The input data consists of both labeled and unlabeled instances. There is a desired prediction problem, but the model must learn the structures to both organize and forecast the data. Classification and regression are two examples of reinforcement learning. Example algorithms are flexible extensions of previous approaches that make assumptions about how to describe unlabeled data. Semi-supervised learning approaches in domains like picture classification, where there are big datasets with few labeled instances, are a hot issue right now. When crunching data to predict business choices, supervised and unsupervised learning approaches are commonly used.

IV. EVALUATION

In this section, we are going to implement our model building process to achieve our research question motive in exact quality. If the model is not perfect we need to iterate the training data or picking different model based on the conditions to retrieve the best fitted model. After applying the proper model, we need to analyze its performance measure to compare the algorithms used to finalize the research question.

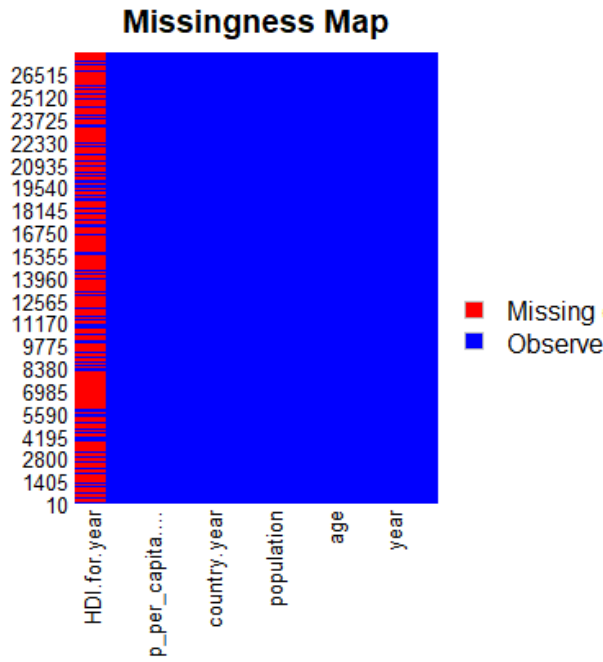


Fig. 11. Missingmap fuction for Suicide data

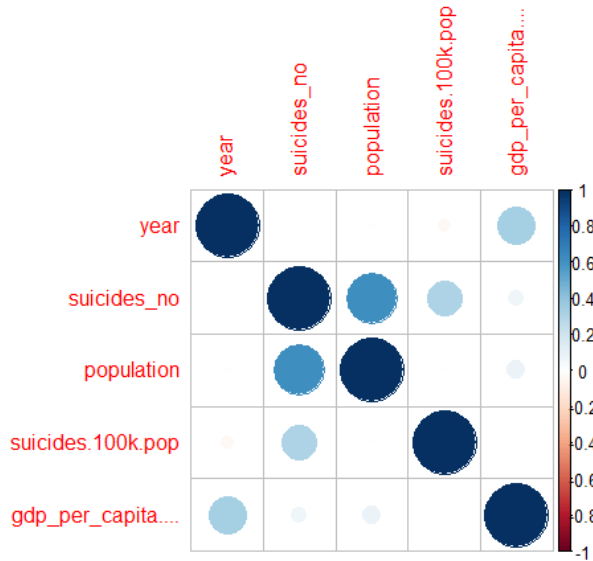


Fig. 12. Correlation Plot for Suicide data

Along with that we are going to visualize the prediction with some parameterised part of our project approach. We will discuss about the success and drawback and limitation of the applied model.

A. Model Implementation and Evaluation

In this part, we are going to provide the best fitted model and to evaluate it. We have used R Studio application do all our data preparation, modelling, visualization and performance evaluation. First of all we need to articulate our problem before hand by Knowing what we want to forecast can assist in determining which data is more beneficial to collect. Conduct data exploration and attempt to think in the areas of classification, clustering and regression that we discussed in our research paper on commercial application of machine learning when phrasing the challenge. The tasks are spliced into different ways that are classification, clustering and regression. We need at least 10k samples to proceed with the test. so in this step we have passed that all 3 dataset contains more than that. In our airline dataset we have 100k records, so we have to check if the resultant variable would be based on the three models whether it is classification, regression, clustering. The variable contained in this was a labelled data. Hence, classification model is best fitted to it. Same like the previous, hotel booking data also consist of labelled data. So classification or clustering only applies to this dataset. Clustering only applies to data value of 10k samples with known labels. Because of that reason, we have considering the classification model for each datasets. In this we have splitted the dataset into two parts, train and test stages by 80 and 20 percent respectively. We have converted the output variable as a string to categorical variable as 0 or 1 for better result and performance evaluation. After that we have applied 2 models logistic regression and random forest algorithm in the airline dataset. KNN algorithm had been applied to hotel booking dataset. Each model has its own function and with this summary we are happy to analyse its performance. Last dataset is suicide data. It consists of 30k records. We have to make classification or regression model to get better result. As our research is based on continuous variable, so we are intended to select the regression model to find out the problem. In this regression models, we have used 3 types of algorithm such as linear regression, support vector machine and decision tree. The decision tree and SVM both will act as either regression or classification problem based on its constraints. After successfully applied the model, we need to visualize it with various parameter and evaluate its performance by comparing 6 different models we have used in this project.

B. Visualization and Performance Measure

The data visualization and performance measure is act as a important role of the project to analyse the use cases and compare the performance notation of applied machine learning models. There are many in-build packages to perform this activity. In this project, we have chosen ggplot function to visualize our data. The performance metric can be done in two ways based on the model. If the model is belongs to the classification part, then we use the confusion matrix to showcase our result. Confusion matrix consists of four segments such as false positive, true positive, false negative and true

negative. Based on this segmentation the result of the model will be calculated. If the model is comes under regression then we use R2 and RMSE after the prediction part is done. Several other useful packages are there like performance function etc. to show the AUC, accuracy and other notations.

Below are the use cases which has been implemented to show the visualization and performance measure of the model.

- 1) Predict how would be satisfaction rate in airline and hotel industry based on gender.
with the industrial sector, we can search for several parameters to take an different approach to make our research better perspective. Having the airline and hotel booking data set we are finding the satisfaction score of the passenger / customer based upon the gender.

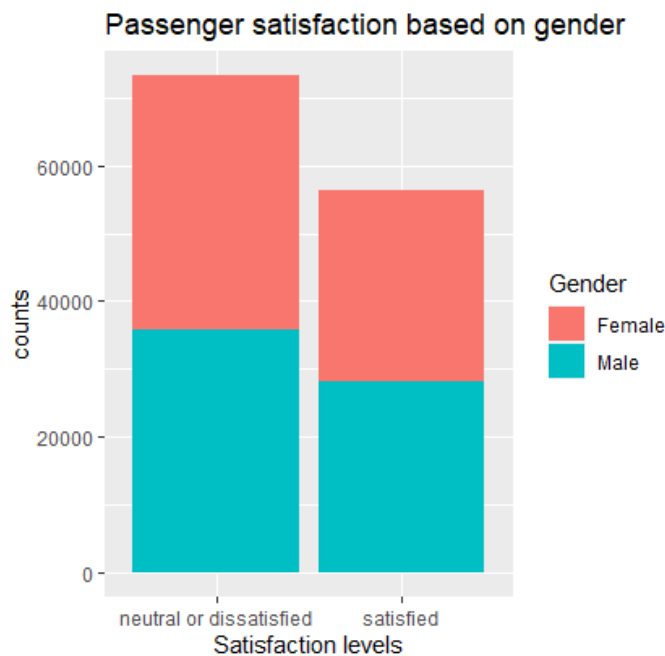


Fig. 13. Satisfaction rate in Airline industry

The above figure 13 shows the satisfaction score of passenger based on gender. This predicts that the satisfaction rate is highly favour to the female participant due to high voting. We can see that female candidate is more dissatisfied as well as more satisfied in this study. The following graph 14 shows the overall satisfaction score of the customer. This shows the neutral or dissatisfied count is more compared to the satisfaction level. As we stated above, we have applied three classification model in this 2 dataset which is shown below. The performance measure taken from the confusion matrix has shown in the table below 15. We can justify that the random forest perform well in the airline dataset which has 99 percent accuracy rate and KNN would be the one dataset which is also acted as a suitable model to the hotel industry.

- 2) Predict how would be suicide rate based upon the gender and age. With the third dataset, We can estimate the

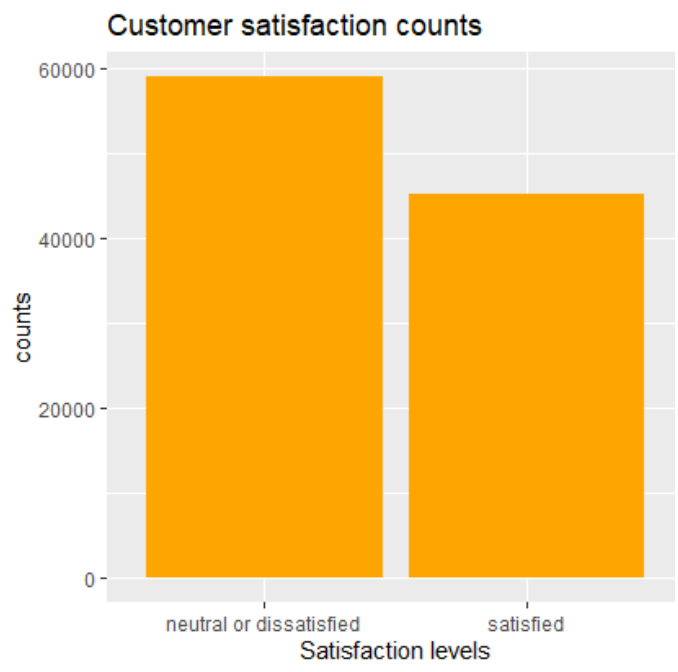


Fig. 14. Satisfaction rate in Hotel industry

RQ	ML Models	Accuracy	Kappa	Precision	Sensitivity	Specificity
Airline Satisfaction	Logistic Regression	87.5	74.4	87.8	90.3	83.6
	Random Forest	99.3	98.6	99.5	99.3	99.3
Hotel Booking Satisfaction	KNN	86.11	71.61	86.7	82.31	89.02

Fig. 15. Overall model performance on RQ1

total death of the people per 100k population based on gender and age parameter. This shows the people with less age will commit less number of suicide compare to the people with more age. Also it shows the male person will be involving in more suicide than female. Not only this we have analysed it with different parameters such as year and country to drive into deep insights.

The following graph 16 shows the suicide death rate per 100k population based on the age and gender using ggplot with bar function.

The following line graph 17 illustrates the suicide rate based on the year. This shows the how suicide rate flows next in future by analyzing the historical data.

The performance measure has been taken to all 3 regression models applied in this use cases. This has simplified R2 and RMSE values which is made after the prediction part of the data. The following table 18 shows the result of this model. In this We need to get more R2 (adjusted R Square) value and least RMSE (Root Mean Square Value) to conclude the best model.

With the measurement, we have selected decision tree as a best model with this approach. It gives around 85

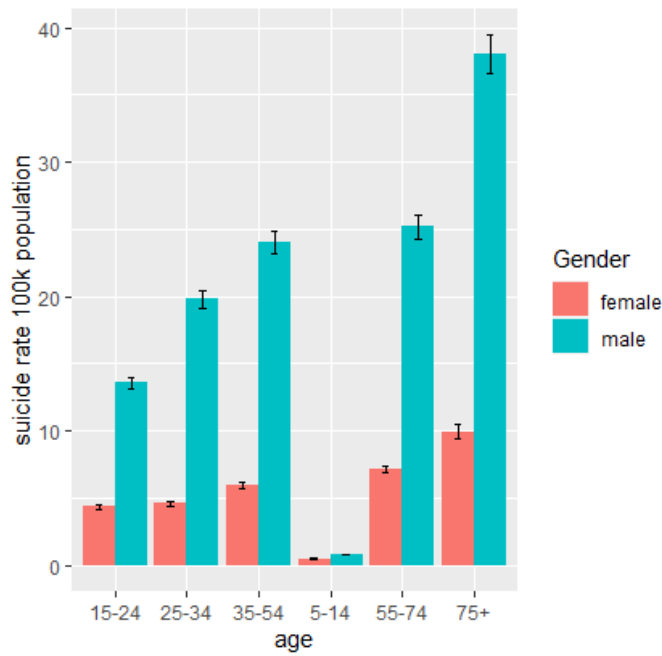


Fig. 16. Suicide Rate based on gender and age

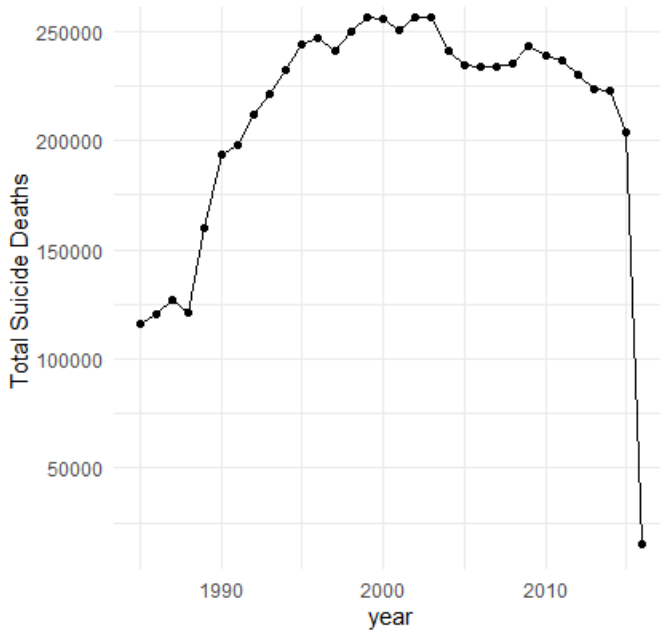


Fig. 17. Line Graph: Suicide Rate based on year

percent of R2 value and RMSE value as 7.

In this project, we have implemented 6 different machine learning algorithms to validate our assumptions which is opt model for our approaches. Although it has its own limitation it perform well in the approach like random forest takes more time for large data set to implement 500 decisions to make a good prediction. It is time consuming but overall it has more accuracy compared to all the machine learning model

RQ	ML Models	R2	MSE	RMSE
Suicide Rate Prediction	Linear Regression	54.6	160.5	12.6
	SVM	64	137.8	11.7
	Decision Tree	84.3	55.7	7.4

Fig. 18. Overall model performance on RQ2

we applied here.

V. CONCLUSIONS AND FUTURE WORK

We came to the end of the project, as we said earlier we have applied 3 classification model and 3 regression model to produce produce the proper comparison of the model. It shows the random forest is the better one out of others. Since it can be applied to both regression and classification as a decision tree, it plays a better role. In this project, we have gone through the supervised learning part, due to the time constrain and for the future we build a application with almost all the model to be covered. With this related dataset, we can estimate the participation of male and female in each and every movements. Thus the female is more involved to the community by saying that they have given more feedback as a customer and suicide rate also less compared to male. By this project, we can conclude that the female is more happiest person in the society.

REFERENCES

- [1] A. Eid, "Application of machine learning in the hotel industry: A critical review," in *Journal of Association of Arab Universities for Tourism and Hospitality(JAAUTH)*, vol. 18, no. 3, 2020, pp. 78–96.
- [2] L. Radlinski, "Predicting user satisfaction in software projects using machine learning techniques," in *15th International Conference on Evaluation of Novel Approaches to Software Engineering*, 2020.
- [3] C. Baydogan and B. Alatas, "Detection of customer satisfaction on unbalanced and multi-class data using machine learning algorithms," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 2019.
- [4] T. Tusar and T. Islam, "A comparative study of sentiment analysis using nlp and different machine learning techniques on us airline twitter data," in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, 2021.
- [5] V. varsha, A. Arti, A. Pooja, and S. L. B. Sanjay, "Iterative machine and deep learning approach for aviation delay prediction," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, 2017.
- [6] T. Parbat, R. Benhal, and H. Jain, "Understanding the customer perception using machine learning while booking flight tickets," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021.
- [7] H. Baskar, B. V. Vishwanath, and V.Rathod, "An exploratory analysis for predicting passenger satisfaction at global hub airports using logistic model trees," in *2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2016.
- [8] V. Natarajan, M. Swaminathan, B. Gowtham, and S. Shubham, "A novel approach: Airline delay prediction using machine learning," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018.
- [9] B. Agorakis, A. Ilias, A. Kanavos, C. Makris, and G. R. A. Savvopoulos, "A sentiment-based hotel review summarization using machine learning techniques," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 2020.
- [10] H. N. T. Thu, "Measuring guest satisfaction from online reviews: Evidence in vietnam," in *Cogent Social Sciences*, vol. 6, 2020, pp. 225–233.

- [11] J. Agustin., S. Medina, and S. Eleaer, "Using machine learning and big data for efficient forecasting of hotel booking cancellations," in *International Journal of Hospitality Management*, 2020.
- [12] A. Nuno, A. Ana, and L. Nunes, "Predicting hotel bookings cancellation with a machine learning classification model," in *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017.
- [13] S. Renjith., A. Annie, S. B. Jyothi, and J. Thomson, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," in *Journal of King Saud University - Computer and Information Sciences*, 2021.
- [14] B. I. Seraphim, S. Das, and A. Ranjan, "A machine learning approach to analyze and predict suicide attempts," in *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH and TECHNOLOGY (IJERT)*, vol. 10, 2021.
- [15] N. Nordin., Z. Zainol, M. Noor, and C. Fong, "A comparative study of machine learning techniques for suicide attempts predictive model," in *Health Informatics Journal*, 2021.
- [16] C. S. Gema., G. Marques, D. Enrique, R. R. Octavio, F. M. Manuel, and D. Isabel, "Suicide risk assessment using machine learning and social networks: a scoping review," in *Journal of Medical Systems*, vol. 44, no. 205, 2020.
- [17] T. Ananya., B. Deep, C. Arijit, and J. sanyal, "Machine learning based prediction of suicide probability," in *International Journal of Engineering and Advanced Technology*, 2020.
- [18] T. Agarwal, A. Dhawan, A. Jain, A. Jain, and S. Gupta, "Analysis and prediction of suicide attempts," in *2019 International Conference on Computing, Power and Communication Technologies (GUCON)*, 2019.