# Comparative Analysis on Airbnb, Online News Popularity and Bank Marketing using Data Mining Approach

Mohit Jain
*MSc in Data Analytics*
*National College of Ireland*
x18200991, Dublin, Ireland
x18200991@student.ncirl.ie

*Abstract*— **Data Science is the boom happening all over the world, and it's accelerating quickly, nowadays, consumption of Data is so massive that it required a whole new study to manipulate and analyze the data. Various techniques can be used for analysis like clustering, Dimensionality reduction, and Machine Learning. For this project, three completely different fields of datasets were considered, namely- Online News Popularity, which was published by Mashable, Airbnb of the Berlin region, and Bank marketing campaign, which was done by the Portuguese banking institution. Using these datasets, various machine learning models are implemented, and comparative analysis is done. According to the dataset, the model will be divided into classification and regression problems. For Online News Popularity, the machine learning model is KNN Classification, and Decision Tree is going to be executed. For Airbnb, multilinear regression and the random forest regression will be executed. And for the Bank Marketing, SVM and Logistics Regression will be used to implement.**

***Keywords—Data mining, Classification, Regression, Machine Learning, KNN, Decision Tree, SVM, Logistics, Linear, Random forest, Evaluation, KDD, Pre-processing, Dimensionality reduction, PCA, Feature Scaling, Outliers***

## I. INTRODUCTION (*HEADING 1*)

In today's world, there is not a single day where we didn't use social media to interact or to connect; we often consume our daily source of news and update through the internet. There are several factors on which the popularity of news depends. In this dataset, there are around 40000 articles with various independent variables such as the number of words in the content, is data channel lifestyle/ entertainment/ business/ social media/ tech/ world? When was the article published? And several others based on which we predict the number of shares that an article gets, through which we can classify whether the article is popular or not. The above analysis would be helpful for the news channel to study their news about the things they need to consider in their mind while publishing their article to reach out to the maximum people. In this paper, the methodology used is KDD. The model would be KNN classification and Decision tree algorithm, which can be compared to check which model works better on this dataset.

Airbnb is very famous for its short-term accommodation, facility to the traveler who is planning their journey and wants to feel like home. The aim for this project is to build a reliable price predictor model with the help of which traveler and the host of the property both can evaluate, it would be helpful for the host to analyze and upload the required information along with the price rate which is going in the neighborhood which is needed to activate their listing live on the website. For the traveler, it would be helpful to check out the price rate in the nearby region of providing according to which the traveler would manage the budget of the trip. In this, the dataset of berlin's Airbnb is going to use which we contain around 23000 data records, to run our model. It includes various factors like the neighborhood, room type, property name, availability in the last 365 days, total no of reviews, and price. Out of these independent variables, the variables with the high correlation will be selected to run our analysis. In this KDD methodology is going used, and the decision tree and KNN classification model will be implemented on this dataset, which would be compared using different evaluation techniques.

Advertisement and Marketing Industry is the ever-growing market which it penetrates every other sector of the market, like the mobile industry, banking sector, traveling industry, and various other industries. The dataset of the Portuguese Banking Institution, which was gathered through telemarketing, is going to use for the analysis. This dataset contains around 41000 records and various other factors, which include customer details like job, marital status, age, has a housing loan? number of times contacted to that customer during a particular campaign. Various others, considering these independent variables, the model can be built to predict whether a client has subscribed for a term deposit or not. After the execution of this model, it could be analyzed what type of audience is subscribing to the term deposit and where can the advertising company invest and what type of audience the company should target. On this dataset, the SVM and Logistics Regression is going to be executed using KDD methodology. This problem is converted into a classification problem, then compared using evaluation metrics.

## II. RELATED WORK

### *A. Airbnb Review*

[6]Pouya Rezazadeh Kalehbasti et al. has published a paper of Airbnb price prediction using Machine learning and Sentiment Analysis, in which they developed a reliable price prediction model using various machine learning, deep learning, and NLP technique, on the base model they proved that on choosing more independent variables leads towards high variance and low accuracy on the model. Out of these techniques, SVR gave the best results with 69% accuracy.

[7] Hujia Yu, Jiafu Wu has published Real Estate price prediction with Regression and Classification, and the aim is to compare and analyze the different regression and classification based on the price, they have also used PCA to improve the accuracy of the model. The most significant feature in this process is to come out to be area square foot, the material of the roof, and the neighborhood. In the regression problem, the best model is SVR with a Gaussian kernel with rmse 0.5271. In a classification problem, the best model us SVC with a linear kernel with an accuracy of 0.6740.

With the PCA, the accuracy of the model is increased to 0.6913.

[8] Yixuan Ma et al. has published a paper about warehouse rental price using machine learning, in which they used the data from the Beijing area and tried to conclude that the warehouse price is are highly dependent on the location and the land price of that area, also they concluded that tree model has better performance than the linear model. The best model on this dataset is the random forest with a correlation coefficient of 0.57 in the test set.

[9] Yang Li et al. has published a paper about the price recommendation on the vacation rental websites, in which they are predicting the price on rental properties which are typically used on a travel or vacation purpose, using various regression model, they use a framework that consists of multi-scale affinity propagation, which is the clusters which are formed to remove the outliers to remove the unwanted data from the dataset.

[10] Jian-Guo Liu et al., have published a paper of Real Estate price prediction using Fuzzy Neural Network with the available response variables according to the quality of data based on the correlation between factors that depend on the price. Also, in the Fuzzy neural network, they had created a price prediction model based on the hedonic price theory.

### B. *Online News Review*

[1] Alexandru Tatar et al. has published the popularity of online articles based on user comments paper to predict the popularity of articles published using the linear regression model using the number of comments an article has received after its publication. The authors of this paper only work on a single factor using the linear regression model. In contrast, they could have included several other factors to see the impact on the dependent variable.

[2] G. Szabo and B. A. Huberman have published the paper on predicting the popularity of online content based on user access. They consider two content- sharing portals Digg and YouTube on which they showed the shared that Digg gets in 30 days is achievable by YouTube in 10 days. The authors of the paper lack in implementing the statistical measure based on which we can further analyze or use our model.

[3] Kristina Lerman et al. has published the Model of Social Dynamics to Predict Popularity of News, in which the author predicted the popularity of new content based on early user reaction, which helps the companies to maximize the revenue and by choosing their content on ad placement. This paper lacks explicitly predicting what type of content will work and the right to invest.

[4] Guandan Chen et al. has published a neural popularity prediction model for social media content, in which they designed deep learning to build a reliable model using factors like text content, user and time series, to analyze the online content, which helps in social media analytics. This paper works on a data-driven approach that focuses more on the informative parts.

### C. *Bank Marketing Review*

[5] Sergio Moro et al. has published a paper on the success of bank telemarketing, in which they have used four data mining models which are – logistic regression, decision tree, neural network, and SVM, to predict the success of telemarketing calls for selling long-term deposits, out of these the accuracy of NN gave the best result.

## III. Methodology

During the implementation of this project, the KDD approach is used, which refers to be the knowledge Discovery in Databases. This approach work based on extracting the essential features of the extensive database. KDD is the five-step process that will be followed by this machine learning model.

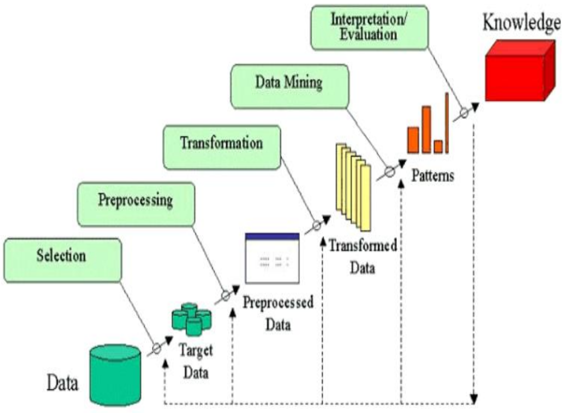


Fig 1

### A. *Data Selection*

Understanding the Domain: There are three different domain of datasets which has been selected to implement our model, which are Airbnb, Bank Telemarketing, and Online News popularity. Among them, Berlin’s Airbnb is selected in which the target variable is chosen the price. For the online news popularity, the data has been collected from the articles published by Mashable in the last two years, using the various feature to analyze and predict the number of shares of a particular article, and for the last dataset which is bank telemarketing taken from Portuguese banking institution on which the goal is to predict whether the client is subscribed a term deposit.

Target dataset: For all three datasets, the target variable will be selected based on the significance value with respect to the dependent variable; after that, the dataset is split into training and test data.

### B. *Data Preprocessing*

This is one of the most critical steps in any machine learning model, as this will decide how data is going to manipulate and the features required to perform the analysis.

In this, the first process is to remove the outlier or unwanted data from the dataset.

Online news popularity: three columns contain outliers. After that, the outlier is removed.

For Airbnb: there are outliers in the price column that have been removed.

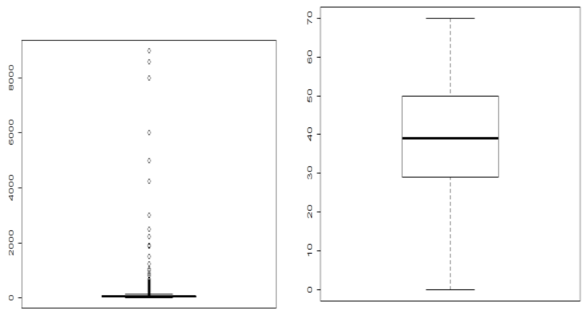


Fig 2

For Bank Marketing: in this dataset, there are no outliers, so this step is not required.

In the pre-processing, there are other processes like cleaning data and replacing the missing value with more suitable value. For Airbnb's price, the missing value of the price is replaced by the median, while for other datasets, there are no missing values.

Also, some of the features required to convert into categorical form and which helps in data manipulation, all of the three datasets have some columns on which has some categorical data.

## *C. Data Transformation*

This step is used to prepare an appropriate data to use for model evaluation. There are several techniques which include feature scaling and dimensionality reduction, in this project, the data is manipulated, and unnecessary columns of these datasets have been dropped and prepared to perform our analysis.

## *D. Data Mining Implementation*

This step is used to prepare an appropriate data to use for model evaluation. There are several techniques which include feature scaling and dimensionality reduction, in this project, the data is manipulated, and unnecessary columns of these datasets have been dropped and prepared to perform our analysis.

Online News Popularity Data: In this dataset, the target variable is numerical values, which are going to be converted into the classification problem, and it will be used to predict the popularity of the news. In this, the number of shares of news is going to be converted into two categories by the median of the number of shared news as a reference point, one of the categories would be popular news, and other categories would be unpopular news. In this classification problem, two models will be used- first is KNN classification, and the other is the Decision tree.

KNN classification is used to predict which type of articles are more popular based on several features, using this model, the classifier is created, which we the best result at k=9.

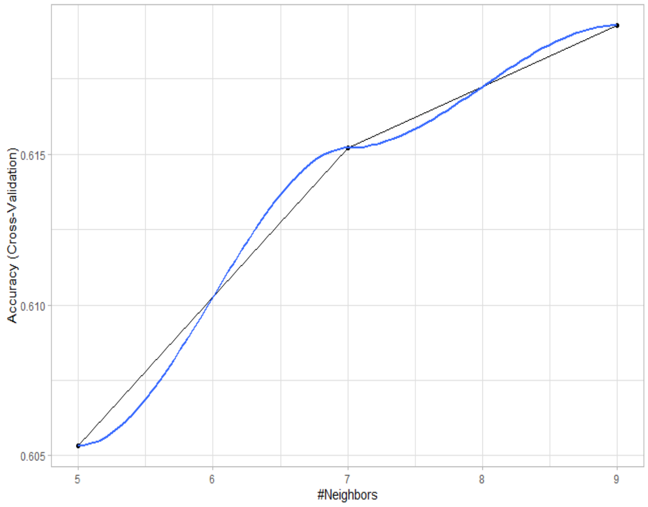
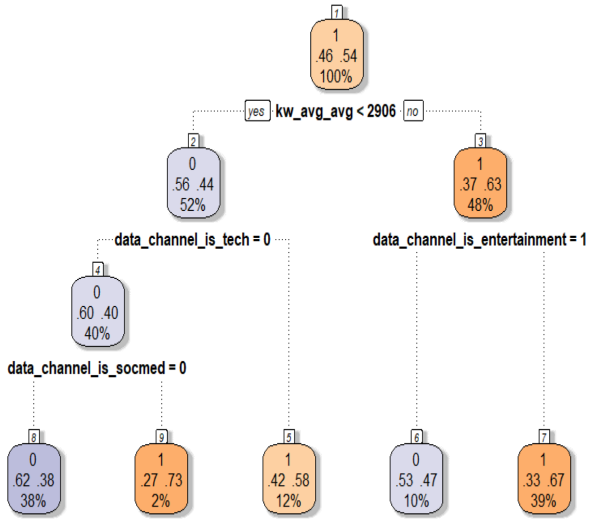


Fig 3

Decision Tree Classification is also implemented on the same dataset using the same set of features, and the classifier created by this is mentioned below

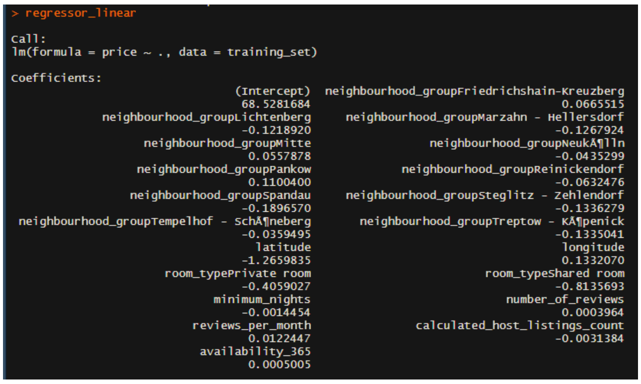


Rattle 2020-May-03 09:08:08 jnmht

Fig 4

Airbnb Data: In this dataset, the target variable is the price, and there are several featured which would be required for the prediction, on this regression model is the more suitable as predicted value is in numerical value, for regression problem Multilinear Regression and Random forest Regression are the two models which are going to implement on this dataset.

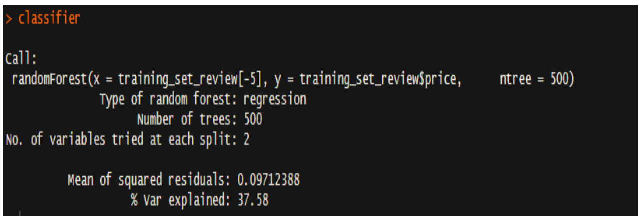
In Multi Linear Regression, there is more than one independent variable is going to use to depict the price of the Airbnb berlin, the independent variable with high correlation value is more significant on the dependent value with respect to the one with low correlation.



```
> regressor_linear

Call:
lm(formula = price ~ ., data = training_set)

Coefficients:
                            (Intercept)  neighbourhood_groupFriedrichshain-Kreuzberg
                             68.5281684                                    0.0665515
         neighbourhood_groupLichtenberg     neighbourhood_groupMarzahn - Hellersdorf
                             -0.1218920                                   -0.1267924
               neighbourhood_groupMitte             neighbourhood_groupNeukÃ¶lln
                              0.0557878                                   -0.0435299
              neighbourhood_groupPankow          neighbourhood_groupReinickendorf
                              0.1100400                                   -0.0632476
             neighbourhood_groupSpandau     neighbourhood_groupSteglitz - Zehlendorf
                             -0.1896570                                   -0.1336279
 neighbourhood_groupTempelhof - SchÃ¶neberg      neighbourhood_groupTreptow - KÃ¶penick
                             -0.0359495                                   -0.1335041
                               latitude                                    longitude
                             -1.2659835                                    0.1332070
                  room_typePrivate room                         room_typeShared room
                             -0.4059027                                   -0.8135693
                         minimum_nights                            number_of_reviews
                             -0.0014454                                    0.0003964
                      reviews_per_month               calculated_host_listings_count
                              0.0122447                                   -0.0031384
                       availability_365
                              0.0005005
```

Fig5

For the Random Forest, the tree used to create a classifier is 500, and rmse on the training dataset is 0.097, and as mentioned below.

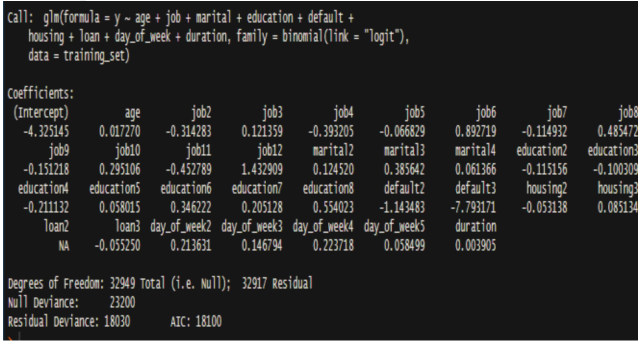


```
> classifier

Call:
 randomForest(x = training_set_review[-5], y = training_set_review$price,      ntree = 500)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 2

          Mean of squared residuals: 0.09712388
                    % Var explained: 37.58
```

Fig 6

Bank Marketing Data: It is a classification problem where the prediction is based on whether the client has subscribed for long term deposit or not considering various factors, in this two-classification model, will be implemented, which are SVM and Logistic Regression.

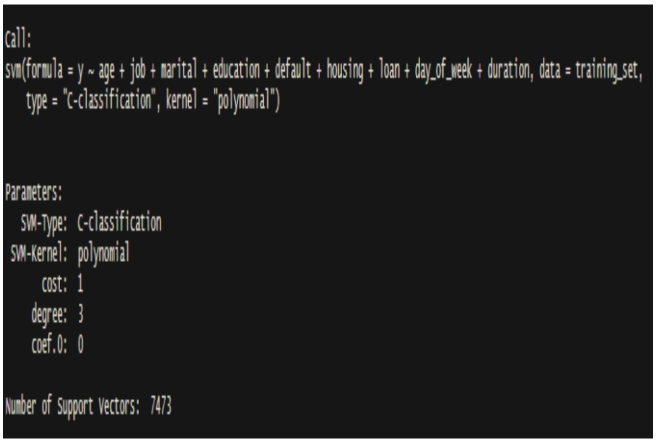
Logistic Regression uses mainly for the classification problem, and it uses logistics function to create a classifier of the training model and normalize the value between 0 to 1, in the below diagram, it is mentioned the correlation coefficient of the independent variable.



```
Call:  glm(formula = y ~ age + job + marital + education + default + 
    housing + loan + day_of_week + duration, family = binomial(link = "logit"), 
    data = training_set)

Coefficients:
 (Intercept)          age         job2         job3         job4         job5         job6         job7         job8
   -4.325145     0.017270    -0.314283     0.121359    -0.393205    -0.066829     0.892719    -0.114932     0.485472
        job9        job10        job11        job12     marital2     marital3     marital4   education2   education3
   -0.151218     0.295106    -0.452789     1.432909     0.124520     0.385642     0.061366    -0.115156    -0.100309
  education4   education5   education6   education7   education8     default2     default3     housing2     housing3
   -0.211132     0.058015     0.346222     0.205128     0.554023    -1.143483    -7.793171    -0.053138     0.085134
       loan2        loan3  day_of_week2  day_of_week3  day_of_week4  day_of_week5     duration
          NA    -0.055250     0.213631     0.146794     0.223718     0.058499     0.003905

Degrees of Freedom: 32949 Total (i.e. Null);  32917 Residual
Null Deviance:      23200 
Residual Deviance: 18030        AIC: 18100
```

Fig 7

Support Vector Machine, in this, the classifier first plots the data item of each independent variable on a hyperplane and then finds the best line to segregates into class and divide it into two parts. The classifier of SVM is shown below, in which it mentioned that the number of support vector in this classifier is 7473.



```
Call:
svm(formula = y ~ age + job + marital + education + default + housing + loan + day_of_week + duration, data = training_set, 
    type = "C-classification", kernel = "polynomial")


Parameters:
   SVM-Type:  C-classification 
 SVM-Kernel:  polynomial 
       cost:  1 
     degree:  3 
     coef.0:  0 

Number of Support Vectors:  7473
```

Fig 8

## E. *Evaluation*

For the evaluation of the machine learning models, there are two types of machine methods first is evaluation metrics, and the other is performance metrics.

Evaluation metrics are generally used in the case of a regression problem, in which how good the model is described on the basis of the Accuracy of training and test data, other factors that would be considered are R-square value, RMSE and MSE value which will be discussed later.

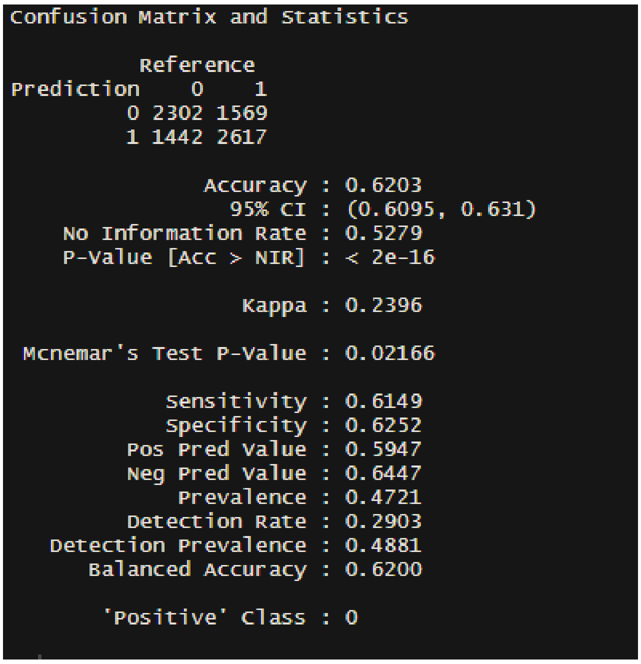
Performance metrics are generally used in case of classification problem, in which how good the model is fitted during the training of model and how good its performance is described using the factor like confusion matrix, accuracy, recall or sensitivity, precision, and specificity.

# IV. Evaluation Methods

For the three datasets mentioned in this paper, there are several machine learning models that are used on these datasets are being compared on the basis of evaluation metrics.

Online News Popularity: In this classification model, the evaluation will be on the basis of performance metrics for each model.

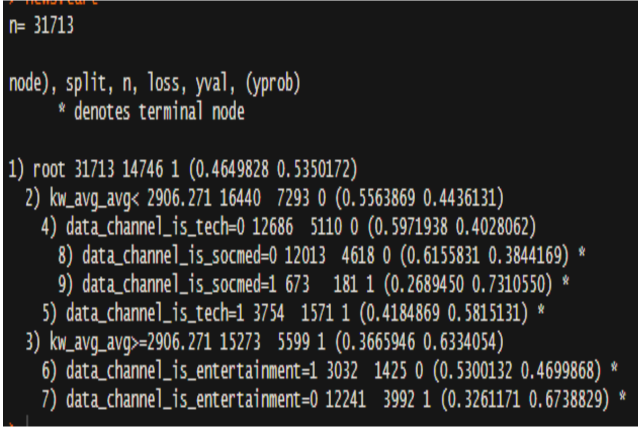
For KNN classification, after creating the classifier, the best training is described at k=9, which gives an accuracy of around 62% on a training set. Also, after the prediction and evaluation of the test set, the accuracy is turned out to be 62.03%. also, there are others mentioned based on that one can see how fit the model is.



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2302 1569
         1 1442 2617

               Accuracy : 0.6203
                 95% CI : (0.6095, 0.631)
    No Information Rate : 0.5279
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.2396

 Mcnemar's Test P-Value : 0.02166

            Sensitivity : 0.6149
            Specificity : 0.6252
         Pos Pred Value : 0.5947
         Neg Pred Value : 0.6447
             Prevalence : 0.4721
         Detection Rate : 0.2903
   Detection Prevalence : 0.4881
      Balanced Accuracy : 0.6200

       'Positive' Class : 0
```

Fig 9

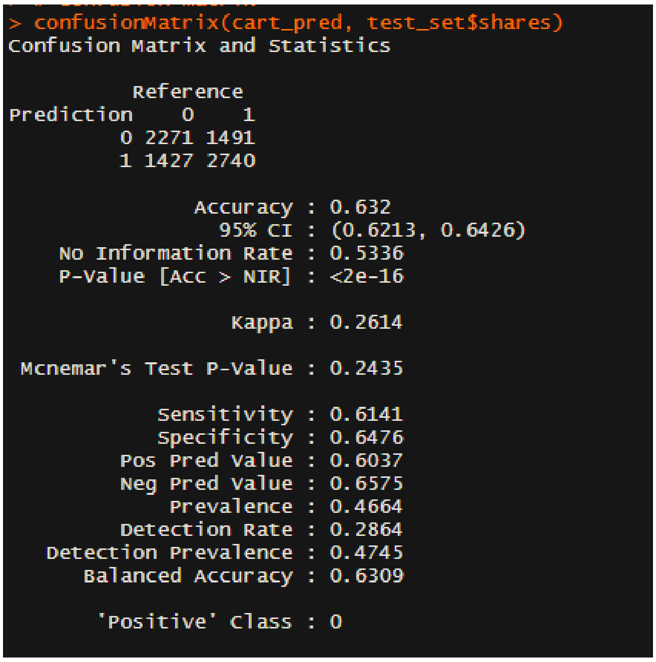
For Decision Tree, in the decision tree, the classifier is created with a root node of kw_avg_avg and divide into trees with the parent node mentioned below



```
n= 31713

node), split, n, loss, yval, (yprob)
      * denotes terminal node

1) root 31713 14746 1 (0.4649828 0.5350172)
  2) kw_avg_avg< 2906.271 16440  7293 0 (0.5563869 0.4436131)
    4) data_channel_is_tech=0 12686  5110 0 (0.5971938 0.4028062)
      8) data_channel_is_socmed=0 12013  4618 0 (0.6155831 0.3844169) *
      9) data_channel_is_socmed=1 673   181 1 (0.2689450 0.7310550) *
    5) data_channel_is_tech=1 3754  1571 1 (0.4184869 0.5815131) *
  3) kw_avg_avg>=2906.271 15273  5599 1 (0.3665946 0.6334054)
    6) data_channel_is_entertainment=1 3032  1425 0 (0.5300132 0.4699868) *
    7) data_channel_is_entertainment=0 12241  3992 1 (0.3261171 0.6738829) *
```

Fig 10

Also, using the confusion matrix, the accuracy on the test dataset is turned out to be 63.2%, and other values are also mentioned below, which shows the fitness of the model.



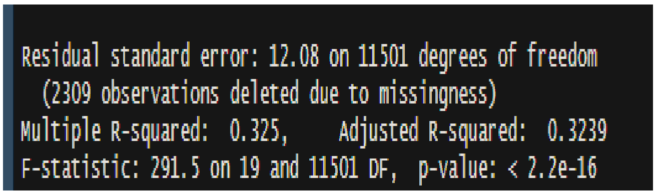
```
> confusionMatrix(cart_pred, test_set$shares)
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2271 1491
         1 1427 2740

               Accuracy : 0.632
                 95% CI : (0.6213, 0.6426)
    No Information Rate : 0.5336
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.2614

 Mcnemar's Test P-Value : 0.2435

            Sensitivity : 0.6141
            Specificity : 0.6476
         Pos Pred Value : 0.6037
         Neg Pred Value : 0.6575
             Prevalence : 0.4664
         Detection Rate : 0.2864
   Detection Prevalence : 0.4745
      Balanced Accuracy : 0.6309

       'Positive' Class : 0
```

Fig 11

Comparing both the classification model, it turned out to be a decision tree works better on the popularity of news than the KNN classification as their kappa, p-value, and specificity gave the better result.

Airbnb Data: In this regression model the evaluation is based on the evaluation metrics for each algorithm

For Multi Linear Regression, the classifier is created, and on the basis of more significant values, the model itself pick the features and run the model, and the r square value on the training set is turn out to be 0.321, while on the test dataset value is 0.325.

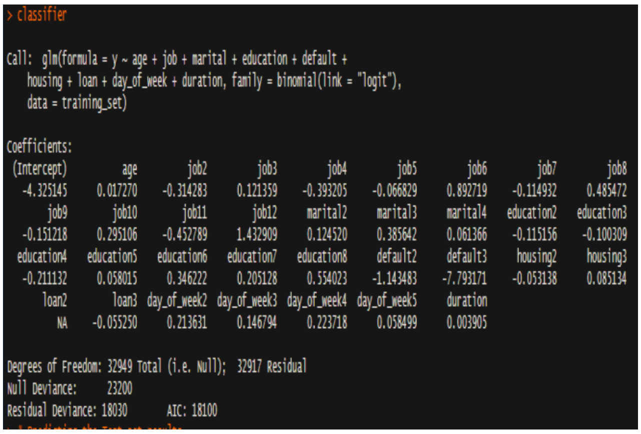


```
Residual standard error: 12.08 on 11501 degrees of freedom
  (2309 observations deleted due to missingness)
Multiple R-squared:  0.325,     Adjusted R-squared:  0.3239
F-statistic: 291.5 on 19 and 11501 DF,  p-value: < 2.2e-16
```

Fig 12

In this normalization, using a log function is also tried to improve the accuracy of the model, but no significant result found.

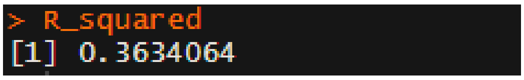
For the Random forest, in case of the random forest, the classifier created by the regression model mentioned below in which correlation of selective variable is defined:



```
> classifier

Call:  glm(formula = y ~ age + job + marital + education + default + 
    housing + loan + day_of_week + duration, family = binomial(link = "logit"), 
    data = training_set)

Coefficients:
 (Intercept)          age         job2         job3         job4         job5         job6         job7         job8  
   -4.325145     0.017270    -0.314283     0.121359    -0.393205    -0.066829     0.892719    -0.114932     0.485472  
        job9        job10        job11        job12     marital2     marital3     marital4   education2   education3  
   -0.151218     0.295106    -0.452789     1.432909     0.124520     0.385642     0.061366    -0.115156    -0.100309  
  education4   education5   education6   education7   education8     default2     default3     housing2     housing3  
   -0.211132     0.058015     0.346222     0.205128     0.554023    -1.143483    -7.793171    -0.053138     0.085134  
       loan2        loan3 day_of_week2 day_of_week3 day_of_week4 day_of_week5     duration  
          NA    -0.055250     0.213631     0.146794     0.223718     0.058499     0.003905  

Degrees of Freedom: 32949 Total (i.e. Null);  32917 Residual
Null Deviance:	    23200 
Residual Deviance: 18030 	AIC: 18100
```

Fig 13

The Accuracy of the random forest model on the test set is 36.3%.



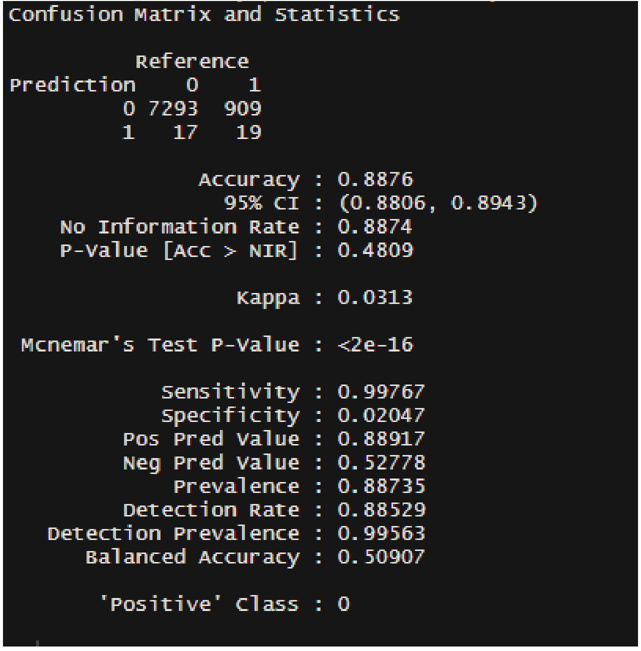
```
> R_squared
[1] 0.3634064
```

Fig 14

Comparing both regression models, it out that random forest works better in Airbnb dataset than the multilinear regression.

Bank Marketing Data: On this dataset, the classification model has been created, the model used is SVM and Logistic Regression, and to evaluate these model, the performance metrics will be used.

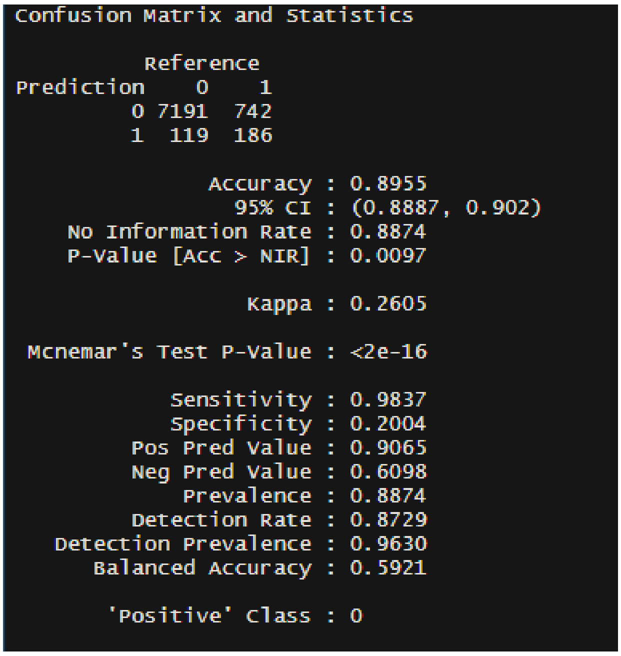
For SVM, in this model, the classifier is created, and the accuracy of the model is turned out to be 88.76%, and other factors of performance metrics are mentioned below



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 7293  909
         1   17   19
                                          
               Accuracy : 0.8876          
                 95% CI : (0.8806, 0.8943)
    No Information Rate : 0.8874          
    P-Value [Acc > NIR] : 0.4809          
                                          
                  Kappa : 0.0313          
                                          
 Mcnemar's Test P-Value : <2e-16          
                                          
            Sensitivity : 0.99767         
            Specificity : 0.02047         
         Pos Pred Value : 0.88917         
         Neg Pred Value : 0.52778         
             Prevalence : 0.88735         
         Detection Rate : 0.88529         
   Detection Prevalence : 0.99563         
      Balanced Accuracy : 0.50907         
                                          
       'Positive' Class : 0               
```

Fig 15

For Logistics Regression, this model has created a classifier, and the target value is predicted in a range of 0 to 1, which further convert into factors, and then evaluated using confusion matrix and the accuracy on the test set is 89.55% and other performance factors are also mentioned below.



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 7191  742
         1  119  186
                                         
               Accuracy : 0.8955         
                 95% CI : (0.8887, 0.902)
    No Information Rate : 0.8874         
    P-Value [Acc > NIR] : 0.0097         
                                         
                  Kappa : 0.2605         
                                         
 Mcnemar's Test P-Value : <2e-16         
                                         
            Sensitivity : 0.9837         
            Specificity : 0.2004         
         Pos Pred Value : 0.9065         
         Neg Pred Value : 0.6098         
             Prevalence : 0.8874         
         Detection Rate : 0.8729         
   Detection Prevalence : 0.9630         
      Balanced Accuracy : 0.5921         
                                         
       'Positive' Class : 0              
```

Fig 16

For the bank marketing dataset, it could be concluded that logistics regression is a slightly better model than the SVM.

## V. Conclusion

For all three datasets, it can be concluded that for online news popularity dataset decision tree works better to predict which articles could be more popular, for the Airbnb berlin dataset random forest work better to predict the price of the listing that has been published on the Airbnb website, and for the bank marketing dataset the logistics regression work better to analysis whether the client subscribed for long term deposit or not.

## VI. Future Work

For Airbnb, there could be more analysis run using this dataset, on room type, or the neighborhood group to give the idea about the realistic approach to invest on the property in the near future for the host.

For Online News Popularity, there are several other factors that could be considered individually to see the significance on the target variable. This will help the news channel to publish an article targeting the appropriate audience on which they want the news to be circulated.

Bank marketing dataset could use for the analysis on which type of campaigning works better on the people and also in the future it could be depicted, what type of client should be approached by the campaigning team, which eventually helps in selecting the target audience.

## References

[1] A. Tatar, P. Antoniadis, M. Amorim and S. Fdida, "From popularity prediction to ranking online news", *Social Network Analysis and Mining*, vol. 4, no. 1, 2014. Available: 10.1007/s13278-014-0174-8 [Accessed 2 May 2020].

[2]G. Szabo and B. Huberman, "Predicting the Popularity of Online Content", *SSRN Electronic Journal*, 2008. Available: 10.2139/ssrn.1295610 [Accessed 5 April 2020].

[3] K. Lerman and T. Hogg, “Using a model of social dynamics to predict popularity of news,” in Proceedings of the 19th international conference on World Wide Web. ACM, 2010, pp. 621–630

[4] G. Chen, Q. Kong, N. Xu and W. Mao, "NPP: A neural popularity prediction model for social media content", *Neurocomputing*, vol. 333, pp. 221-230, 2019. Available: 10.1016/j.neucom.2018.12.039 [Accessed 15 April 2020].

[5] S. Moro, P. Cortez and P. Rita, "A data-driven approach to predict the success of bank telemarketing", *Decision Support Systems*, vol. 62, pp. 22-31, 2014. Available: 10.1016/j.dss.2014.03.001 [Accessed 25 April 2020].

[6] Kalehbasti, Pouya Rezazadeh, Liubov Nikolenko, and Hoormazd Rezaei. "Airbnb Price Prediction Using Machine Learning and Sentiment Analysis." *arXiv preprint arXiv:1907.12665* (2019).

[7] Yu, Hujia, and Jiafu Wu. "Real estate price prediction with regression and classification." *CS229 (Machine Learning) Final Project Reports* (2016).

[8] Y. Ma, Z. Zhang, A. Ihler and B. Pan, "Estimating Warehouse Rental Price using Machine Learning Techniques", *International Journal of Computers Communications & Control*, vol. 13, no. 2, pp. 235-250, 2018. Available: 10.15837/ijccc.2018.2.3034 [Accessed 30 April 2020].

[9] Li, Yang, Quan Pan, Tao Yang, and Lantian Guo. "Reasonable price recommendation on Airbnb using Multi-Scale clustering." In *2016 35th Chinese Control Conference (CCC)*, pp. 7038-7041. IEEE, 2016.

[10] Pan, Rong. "Real Estate Price Prediction Model Based on Dynamic Neural Network." *Rev. Téc. Ing. Univ. Zulia*, 2016. Available: 10.21311/001.39.6.07 [Accessed 24 April 2020].