Capstone Project

Final Presentation

Done by: Daoud Tebbakh

30.09.2020

Problem statement

- What made Falcon airlines customer satisfied in the past?
- How to keep satisfied customer from churning ?
- Who will stay satisfied in the future?

SOLUTION IN 3 STEPS

#1 Improve the quality of the service

#2 will generate more satisfied customer

#3 will generate more profit

Data process

- Import the dataset
- Renaming variable
- Transformation variable
- Missing values
- Outlier treatment
- Correlation Variable
- Important variables
- Splitting the dataset
- Visualizations

Setting general parameter for the optimal best model

- **Random:** repeated random sub-sampling validation
- **Cross-validation:** high-quality training for model to use all data of training
- **Accuracy:** the highest accuracy
- **Sensitivity :** percentage of all 1's were correctly predicted. The highest
- **Specificity :** percentage of all 0's were correctly predicted. the highest
- **Sensitivity and Specificity :** must be very close and very high
- **Concordance:** the higher Concordance the better the model on cutoff value and easy for observation to be classified with very high prediction.
- **ROC curve:** can be used to know what cutoff gives the best sensitivity, specificity or both. the highest
- **Gini:** Coefficient is an indicator of how well the model outperforms random predictions. the highest
- **KS:** used to make decisions like: How many customers to target for a marketing campaign? or How many customers should we pay for to show ads, also helps to understand?, **KS statistic** is the perfect portion of the population should be targeted to get the highest response rate. The highest
- **Interpretability:** easy or possible

Our Optimal Model target

A Good Fit is at the sweet spot between Under-Fitting and Over-Fitting by applying general parameter to multiple "7" models:

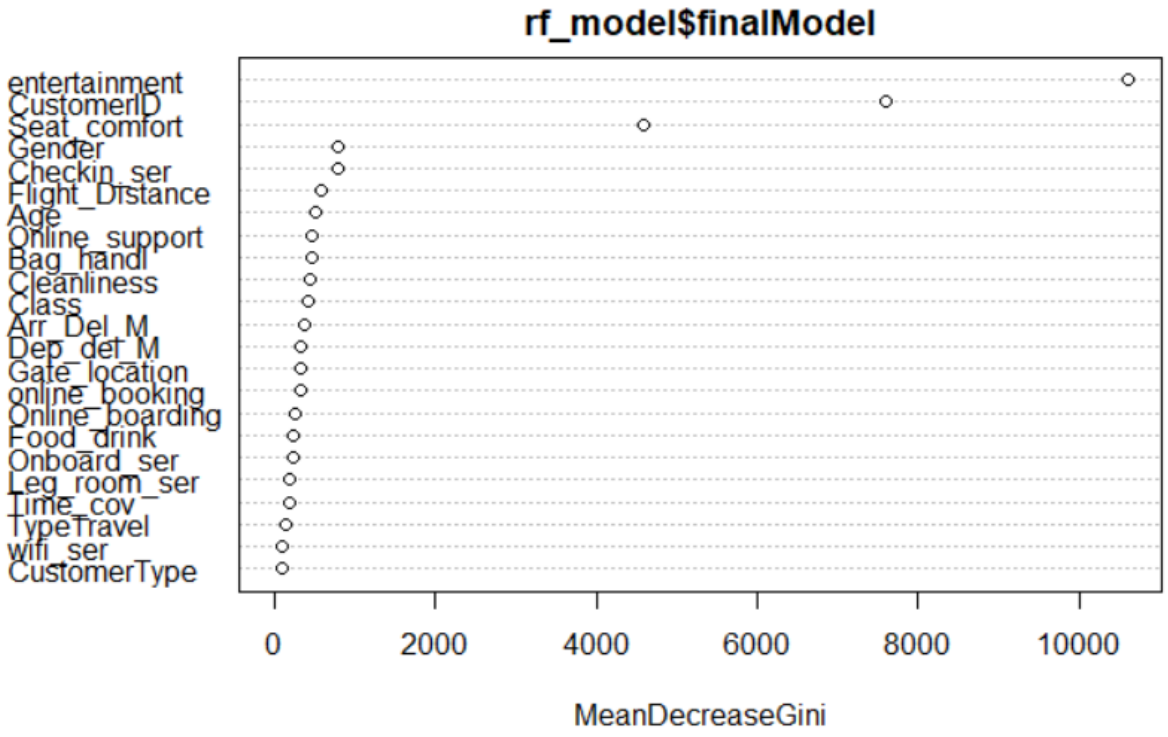
1. Single Decision Tree
2. Random Forest
3. K-Nearest Neighbors
4. Naïve Bayes
5. Logistic Regression
6. Bagging
7. Extreme Gradient boosting

FINAL MODEL SELECTION : RANDOM FOREST

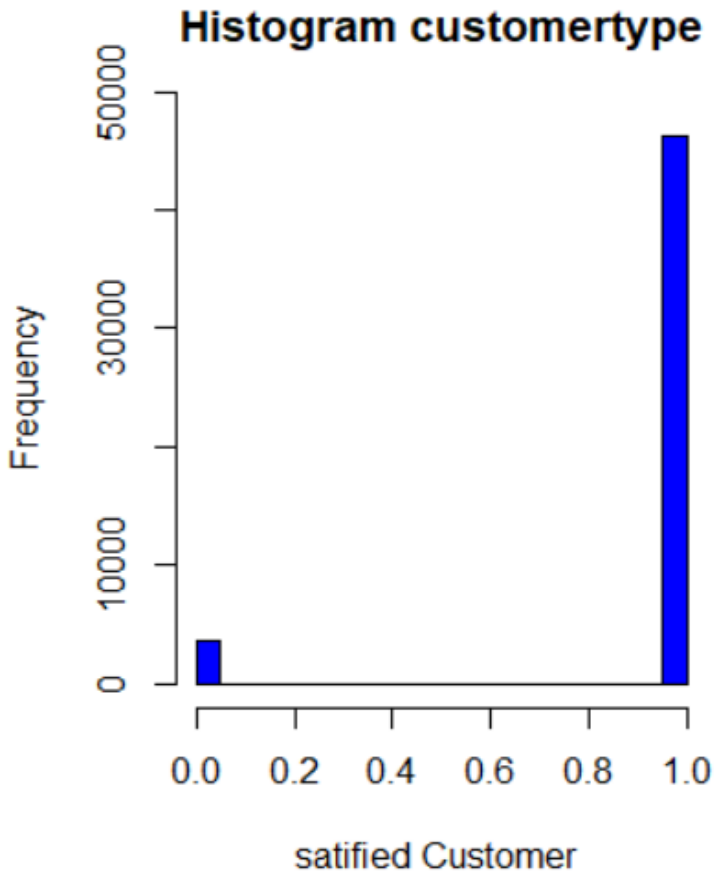
- **Random:** repeated and randomly selected
- **Cross-validation:** k=10
- **Accuracy:** 'random forest' 96.31% the highest
- **Sensitivity:** 'random forest' 95.86% the highest
- **Specificity:** 'random forest' 96.85% the highest
- **Sensitivity and Specificity small different :** 96.85%- 95.86%= 0.99% very small
- **Concordance:** 'random forest' 99.18% the highest
- **ROC :** 'random forest' 99.34% the highest
- **Gini** : 'random forest' 44.5% very high
- **KS statistic:** 'random forest' 90.78% very high
- **Interpretability:** possible

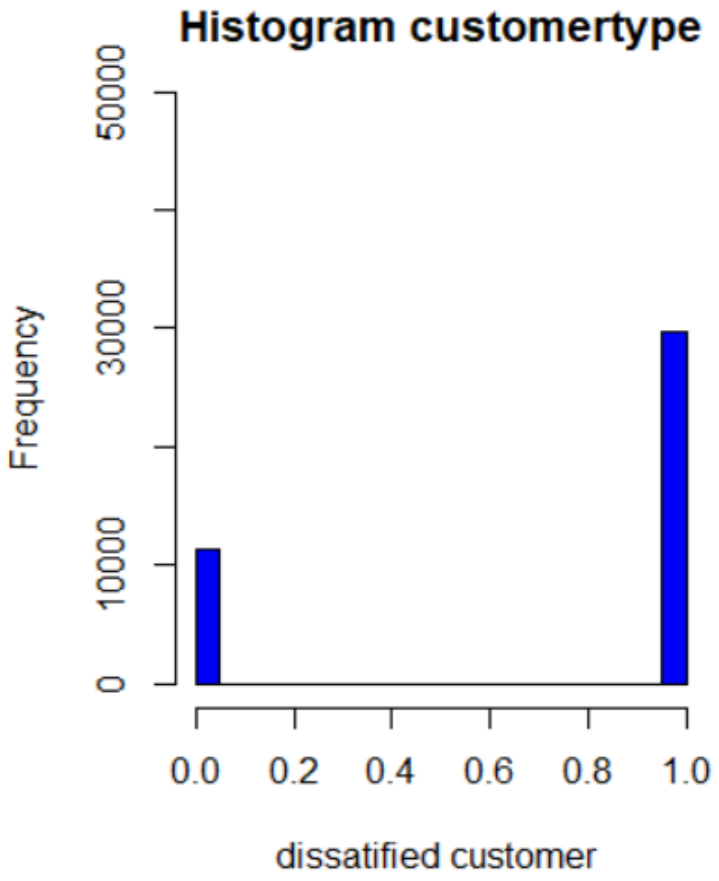
1 Understand your Audience

- Biased on Random Forest model the most important variables that make the customer satisfied are: “entertainment” “CustomerID” “Seat_comfort” upgraded those will help us to keep satisfied customer away from churning.



- 50% of Loyal customer are satisfied, 32% of Loyal customer are dissatisfied, we can churn them satisfied by using important variables.





2 IMPROVE THE SERVICE

- KS =90.78% is portion of the population should be targeted to get the highest response rate of 2973 responders out 3031 using random forest with 90.78% customer satisfaction.
- The KS Chart is particularly useful in marketing campaigns and ads click predictions where you want to know the right population size to target to get the maximum response rate.

rank <int>	total_pop <dbl>	non_responders <int>	responders <int>	expected_responders_by_random <dbl>	perc_responders <dbl>	perc_non_responders <dbl>	cum_perc_responders <dbl>	cum_perc_non_responders <dbl>	difference <dbl>
1	3031	0	3031	1654.118	0.1832638007	0.000000e+00	0.1832638	0.0000000000	0.1832638
2	3031	2	3029	1654.118	0.1831428744	1.452749e-04	0.3664067	0.0001452749	0.3662614
3	3031	3	3028	1654.118	0.1830824113	2.179124e-04	0.5494891	0.0003631873	0.5491259
4	3031	1	3030	1654.118	0.1832033376	7.263747e-05	0.7326924	0.0004358248	0.7322566
5	3031	58	2973	1654.118	0.1797569381	4.212973e-03	0.9124494	0.0046487978	0.9078006
6	3031	1771	1260	1654.118	0.0761835661	1.286410e-01	0.9886329	0.1332897509	0.8553432
7	3031	2887	144	1654.118	0.0087066933	2.097044e-01	0.9973396	0.3429941164	0.6543455
8	3031	3031	0	1654.118	0.0000000000	2.201642e-01	0.9973396	0.5631582770	0.4341813
9	3031	3015	16	1654.118	0.0009674104	2.190020e-01	0.9983070	0.7821602383	0.2161468
10	3027	2999	28	1651.935	0.0016929681	2.178398e-01	1.0000000	1.0000000000	0.0000000

10 rows | 1-6 of 10 columns

3 Generate More Profit

- The KS Chart is particularly useful in marketing campaigns and ads click predictions where you want to know the right population size to target to get the maximum response rate.
- By targeting the top 60% of the population (point it touches the X-axis), the Random forest model is able to cover 98.86% of responders as satisfied customers.