FoodCorp Churn Prediction Report

Part I. Executive Summary

As obtaining new customers induce high cost, understanding customer churn and retaining them are crucial for business. FoodCorp is a company operating four stores in London, Birmingham and Nottingham. As required by FoodCorp, a churn prediction system for listing people who may churn is developed and expected to run weekly. The ConsultingCorp is hired to analyse the current level of churn and provided a statistical description of various churn definition. In the first part, the report from the ConsultingCorp is interpreted, and churn is defined as16 days for balancing marketing budget and predictive modelling performance. The second part is the technical implementation for the churn prediction model. Ten features containing both temporal features and aggregate features are selected assisting by boruta wrapped with random forest. After parameter tuning for AdaBoost and Support Vector Machine, AdaBoost wins using recall and roc_auc score as evaluation methods. In the final part, the pen portraits for both churn and non-churn customers are described with the statistics summary supporting, and according to the pen portraits, marketing suggestions are such as giving special offers, promote loyalty program and build brand awareness.

Beyond the recommendations given in the final insight part, to improve the churn model further the cost of targeting one customer and bring them back is required. The benefits of a customer are also required to carry out the cost and benefits matrix for the better evaluating churn prediction model. The second suggestion is that although regularity is not taken into consideration at this stage due to the complexity. A better model can be developed in the future to take individual behaviour into consideration and label them according to their regularity. Also, when there is a large amount of customer labelled as churn, individual lifetime value based on personal buying behaviour can be used to indicating the benefit of targeting.

The report is complemented by the two Jupyter notebooks under the name ‘FoodCrop Churn Prediction_Evaluation Code.ipynb' which details the technical part of the analysis. Another ‘FoodCrop Churn Prediction_Final Model Code.ipynb' is the final model implementation which aims for FoodCrop to run on week base. There is a supporting document named ‘FoodCrop_Churn Prediction_Supporting Documentation.doc' illustrating the instruction for running the final model with data cleaning statement inside and the illustration for all figures used in this report.

Part II. Current levels of churn

Based on the report done by ConsultingCorp, global churn definition is selected rather than considering regularity as a result of its complexity. ConsultingCorp defines the churn as follows:

If date_today - date_of_last_purchace **>** β **then** churn
otherwise, not churn

To determine a proper β, ConsultingCrop examines two statistics which is 'distribution of times between visits' and 'percentage of customers who are considered churners' and plotting three figures. Churn definition with 16 days is recommended, and in the formula, it means if the churn prediction model is run on 10th April 2019 and the customer last purchase date is on 15th March 2019 which the gap between the dates is larger than 16. This customer will be identified as churned by definition and FoodCorp can take actions to target this customer until he or she carry out transaction again.

Under this definition, there will be 19.85% of active people with a perfect classifier to target. If β is set too high, i.e. 43 days, the churn prediction model is less reliable as the predicting period is relatively long. With the minimal churn day as one day, there are 84.76% of active people to target which may retain more customers but expensive to carry out. Due to the limitation of marketing budget and predictive model, '23 days' and '16 days' are recommended. The optimal choice of β should depend on the cost of customer retention and planned marketing budget. Since cost and budget are not pointed out here and there will be a percentage of customer losing even with preventing measures. As a result churn definition with 16 days is set with 59.9% of customers median days between visits is less than this and 19.85% of active people to target.

Part III. Churn prediction system.

- Data preparation and features

The data is import and analysed via python code connecting with the database in the document 'FodCorp Churn Prediction_Evaluation code.ipynb'. Data cleaning statement can be seen in the supporting document.

<u>Initial features</u>

The date of the transactional data is from 29th March 2017 to 27th November 2018 having 608 days in between. Temporal features are considered as they can be informative indicators enlighting the changes and patterns of customer behaviour improving the prediction model. The tumbling window is set as seven days since the weekly pattern is considered as informative. The output window size is the same as β. As shown in figure 1, for training model, parameter tuning and testing model performance, training set, validation set and testing set are set with reference day '2018-10-10', '2018-10-26' and '2018-11-11' respectively with ten input windows and one output window. Input features chunk data from the period one to period ten with window size seven days as backtracking most recent data can be much effective. The total spends, the total quantity of purchased items, the number of visits and the number of visited stores are aggregated from each period. Higher spend, the number of products engaged and frequency may suggest a higher commitment to FoodCorp and introducing a high shifting cost. Also, if customers visit multiple stores of FoodCrop, that can be a sign of their loyalty. In addition to all temporal features, aggregate features from period ten to one are also considered which are 'median_gap', 'total_spend', 'total_qty', 'total_count' and 'total_store'. After initial feature selection, the correlations among all features are examined which are moderate among temporal features. The correlation between 'total_qty' and 'total_spend' and 'total_store' with 'total_count' are relatively high. In this stage they are kept as Boruta is used for feature selection in the next stage. Also, as a result of moderate correlation among temporal features, machine learning techniques which are robust to high correlations among input features are utilised.

	Input periods (7 days in each period)												output (16 days)	now
Test Set			f10	f9	f8	f7	f6	f5	f4	f3	f2	f1	rf: 2018-11-11	2018-11-27
Valid Set		f10	f9	f8	f7	f6	f5	f4	f3	f2	f1	rf:2018-10-26		
Train Set	f10	f9	f8	f7	f6	f5	f4	f3	f2	f1	rf:2018-10-10			

Figure 1 illustration for train, valid and test sets

- Feature selection with Boruta

Boruta is a wrapper built with random forest creating shadow features by copying and shuffling the real one and comparing them with the real features using Z-score interactively. If the real features perform better than the copied shadow feature, it will be identified as important. Boruta can help to find out variables which have a significant impact on output without specifying how many input features to put in. After 44 iteration, 'f1_qty', 'meadian_gap', 'total_spend', 'total_qty' and 'total_count' are identified as the most important features. Predicting with all features have an accuracy score 0.8352 while it is 0.8352 with only five Boruta selected features which means the other 41 features are providing limited predictive power. However, only 'f1_qty' from all temporal features are selected which means temporal features are not included. A further consideration is 9658 records in the training set are less than the real world data points to put in the model. Training a model with only five features may result in underfitting. Hence, according to the rank of features by boruta, 'f1_spend','f5_spend','f4_qty','f2_spend','f5_qty' are added to selected features to put in the model.

- Prediction approach

As churn prediction is a classification problem and there are ten input features with one binary output feature, Adaboost and support vector machine are selected initially using the validation set to tune parameters.

<u>AdaBoost</u>

As an ensemble method, AdaBoost combines multiple classifiers to increase the accuracy; there is one important parameter for tuning which is 'n_estimators' representing the number of weak learners to train iteratively. 'for loop' in python are used to have a grid search on 'n_estimators' ranging in [5,10,15,20,25,50, 100, 150, 200, 250, 300,350,400,450,500,550,600]. Mean absolute error (MAE) is used to measure the performance of AdaBoost on training and validation set. As shown in figure 2, when n_estimator is 50 the errors of both sets are low and close to each other. With the increase of n_estimator, validation set and training set become diverging to each other suggesting the model may be overfitting. Hence, n_estimator is set as 50.

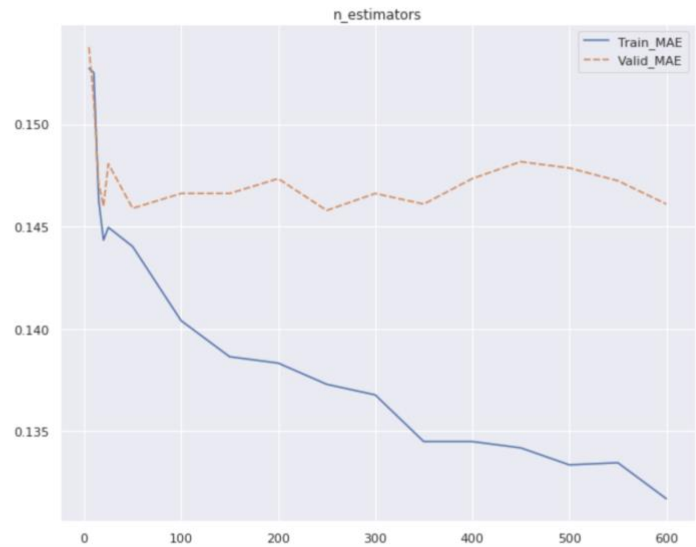

Figure 2 training and validation set MAE with different n_estimators for AdaBoost

<u>Support Vector Machine</u>

As the non-linearity of the problem, the RBF Kernel Support Vector Machine is selected. Initially, the standard scalar is used to standardizing the data, and it is embodied in a pipeline with SVM for coding efficiency. There are two main parameters are tuned which are ‘C' meaning the penalty and ‘gamma'. As figure 3 indicates, validation error and training error become diverging when C is 1000 and gamma is 0.00010. To avoid overfitting and have best model performance C is set as 1000 with gamma as 0.00010.

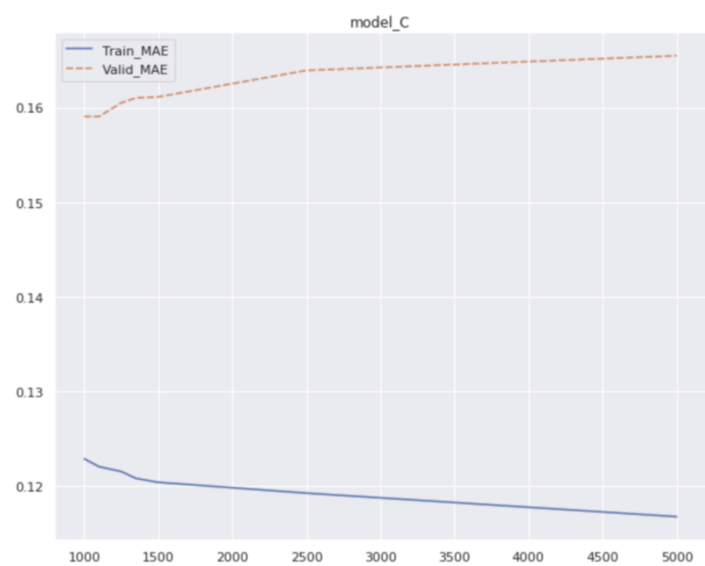

Figure 3 training and validation set MAE with different model_c for SVM

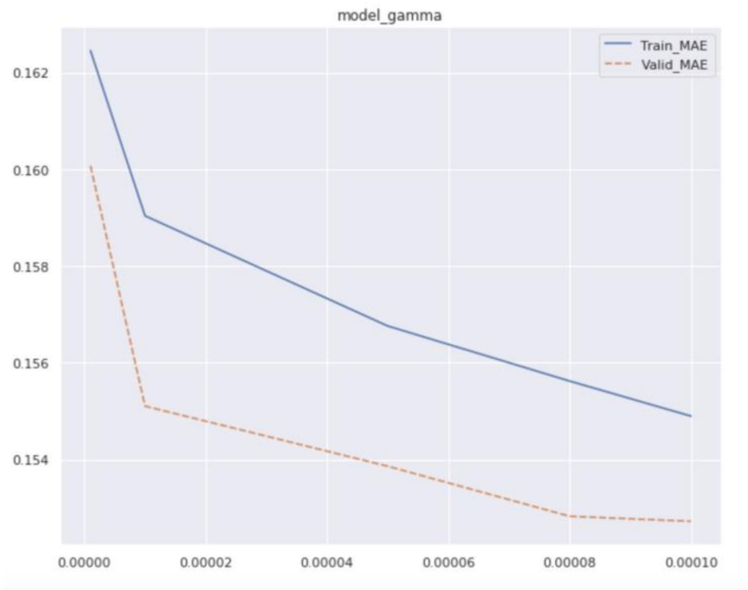

Figure 4 training and validation set MAE with different model_gamma for SVM

- Evaluation

There are many metrics that can be considered to evaluate model performance. Since FoodCrop is keen to identify people who may churn and take action to prevent that, the ability for the model to predict people who churn churned is most important with less missing. For measuring the ability of a classification model to identify all relevant instances, among all measures recall score is stressed. Due to the cost limitation, FoodCorp may not target all positive instances. Hence roc_auc score is considered as well.

DummyClassifier is used for understanding how much improvement AdaBoost and RBF SVM can bring. Since the input data are imbalanced, there is 70.5% of all training set are labelled as churn. Simply by predicting all data point as churn can gain a 70.5% accuracy. As shown in figure 5, AdaBoost and SVM both improve the two scores especially on the roc_auc score. Adaboost has a higher recall rate than SVM. In the confusion matrix (figure 6), SVM raises too many cases on the positive instance which is hard for FoodCrop to target all people. As a result of higher recall score and roc_auc score and proper number for churn targeting, AdaBoost is selected as the final winning model.

	Recall Score	Roc_auc score
Dummy Classifier	0.704	0.4948
AdaBoost	0.9199	0.7903
Support Vector Machine	0.8959	0.7907

Figure 5 recall and roc_auc Score for all models

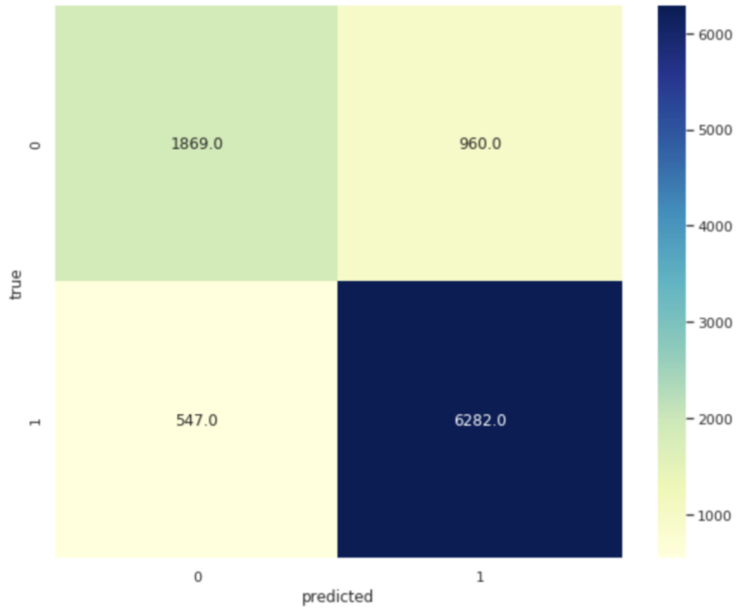

Figure 6 confusion matrix for AdaBoost

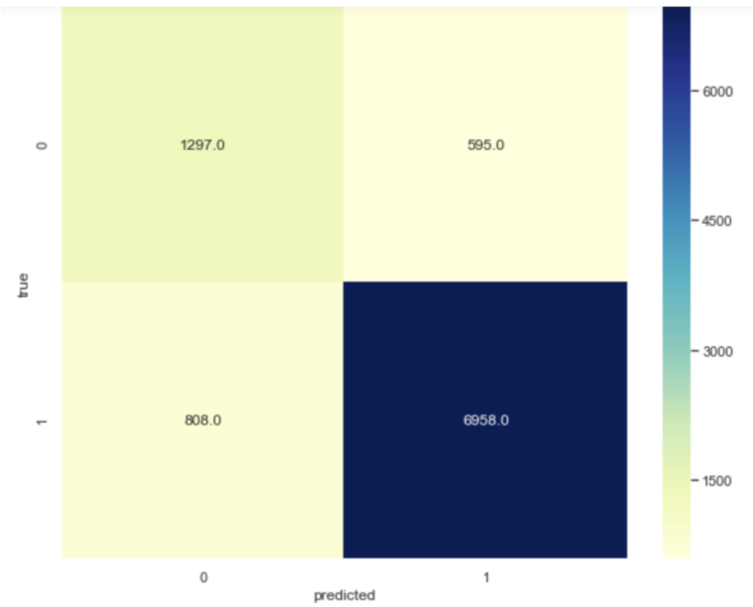

Figure 7 confusion matrix for SVM

- Summary

To summarise the methodology in building the churn prediction model. At the initial stage, fourty four features are generated with forty temporal features in ten weeks periods and four aggregate features with binary output feature under window size 16 days. By using Borut wrapped with random forest model, ‘f1_qty', ‘meadian_gap', ‘total_spend', ‘total_qty' and 'total_count' are selected and to avoid underfiiting and taking temporal features into consideration, ‘f1_spend','f5_spend','f4_qty','f2_spend','f5_qty' are added. RBF SVM and AdaBoost are trained and tuned. AdaBoost is selected as the final model scoring 0.9199 recall score and 0.7903 roc_auc scores with the n_estimator set as 50.

Part IV. Customer insights

- Technical support and pen portraits

According to the evaluation code in 'FoodCorp Churn Prediction_evaluation Code,ipynb', using the reference day 11th November 2018 as a reference point an csv table with forty-four features and churn or non-churn label is derived and import into tableau for evaluation as shown in figure 8 and 9.

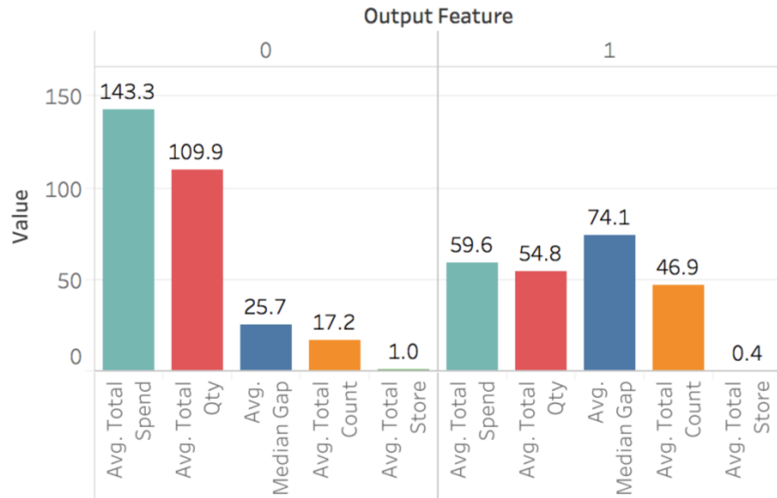

Figure 8 aggregate features differences between groups

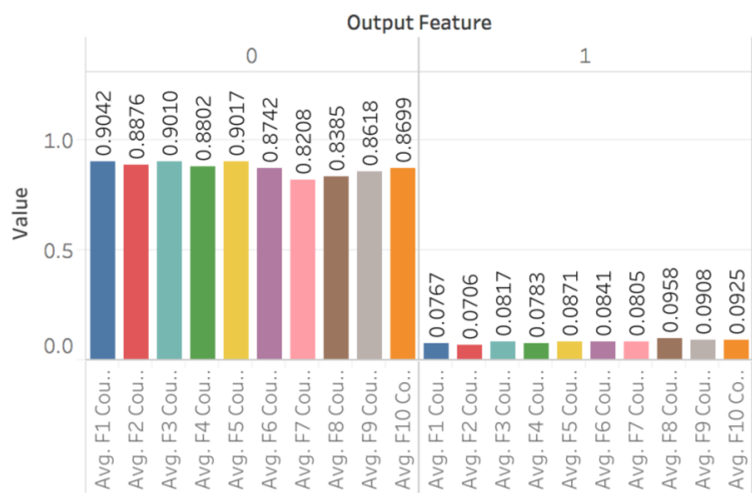

Figure 9 count differences over periods between groups

<u>Pen portraits for churn group</u>

Customers who may churn may just occasionally visit the store for shopping things they are not regular need. They may shop as only for immediate needs or one time shopping with high volume and find FoodCorp is the nearest store equipped with the product they want. As a result, they commit little brand loyalty. They only show up in a limited period even only once with the limited spend, product. As they have little brand awareness of FoodCorp, only one store is visited by them.

Pen portraits for a non-churn group

Customer who identified as loyal have a more stable pattern. They visit the store with more regularity and active in multiple periods. There is some weekly pattern exist suggesting their shopping mission is more likely to satisfy continuous requirements such as buying cooking material for the family on a weekly base. FoodCorp may be the nearest ideal location to satisfy these need since it near to their house or working location, or they may come to FoodCorp on purpose due to their loyalty. Loyal customers involve more products varieties and quantities each time they shop. They also visit FoodCorp in a different location as the recognition to brand.

- Summary and suggestions for marketing

Churn, and the non-churn group have different characteristics in many ways as loyal customers have higher spend, the number of products and frequency. They also shop in multiple stores of FoodCorp as the recognition to brand. For retaining customer and converting churn customer to non-churn, recommendations are given as following aiming increasing spend, quantity and frequency.

Special offers are recommended for people who may churn. FoodCorp could provide that customer with a discount to encourage them to consume again or providing vouchers with an expired date. Promotion bundle can be used to encourage the customer to consume more each time to develop their commitment to FoodCorp and educate the customer to develop their habits. Also, a recommender system can be developed by recommending products they may interest. inPeople may come back to the store. The key is to build up brand awareness and customer habit education. Since the feasibility of targeting, FoodCorp derives the rank according to customer life time value and probability of churn and set different levels of target strategy. Although churn is a problem to avoid, a customer with the lower value may not effective to target. After listing the churn customers by the churn prediction model, Food Corp should target them according to the rank. For a high-value customer, higher discount rate, more attractive offers and more frequent contacts are preferable. Beyond those different characteristics, many reasons may result in a customer churn such as the dissatisfaction about service or product quality, moving home resulting far distance from the store or require more various products. Ways to increase spend, quantity and frequency are posted strategies when the likely to churn already happening. Hence survey to investigate the reasons are recommended for further improvement and plan preventative strategies.