Classification of risk attitudes from customer behavior with machine learning

Teeranai Sriparkdee, Prabhas Chongstitvatana

Department of Computer Engineering

Chulalongkorn University

Bangkok, Thailand

6070927021@student.chula.ac.th ; prabhas.c@chula.ac.th

***Abstract*—Every product and service in the market has its characteristic which has an impact on a consumer's decision to buy or use them. The risk is a distinctive characteristic of financial products, so in financial product and service design must use risk as a key factor. On the other hand, the consumer has different attitudes to the risk which can distinguish in 3 categories: risk aversion, risk neutral and risk seeking. Therefore, knowing risk attitudes of consumer who is the target market is an important key to define marketing strategy such as designing service and product, campaign, and promotion which going to be offered to them. Using the customer historical data, machine learning can be used to classify risk attitudes of each consumer. In this paper, we compare three machine learning methods to classify consumer's risk attitudes from their behaviors and identify important features. The results of the experiment show that the ensemble method, XGBoost, when used with resampling method ADASYN shows the best accuracy,**

Keywords—*Risk Attitudes, customer behavior, Ensemble Model, Random Forest, Gradient Boosting, XGBoost*

I.INTRODUCTION

Financial products are products and services which are offered by financial firms to their consumers such as deposit products, debt instruments, funds, equities, stocks, and derivatives, etc. Each financial product has a different level of risk, such as fund which can categorize type by its risk into eight levels.

With the different risks of each financial product, the attitude toward risk of each customer affects the decision to use the firm's financial products. Therefore, if the firm can know the attitude towards the likelihood of each customer, the firm will be able to develop the financial service and products, including campaigns and effective promotion to each customer by the attitude towards risk.

Generally, in finance and economics, there are three types of attitudes towards risk : risk aversion, risk neutral and risk seeking. The firms can evaluate their consumers' risk attitudes when consumers are buying investment products from the firm. Therefore, it is possible to assess the attitude towards risk only when they purchase the investment product from the firm. While in the case that the consumer never have a profile of buying investment products from the firm will not be able to evaluate the attitude towards the risk of such customers at all. The firm can specify their consumers risk attitude by the questionnaire which uses for the consumer to self-assess with this method result may affect consumers bias. Using consumer behaviors can reflect their risk attitude from their real-life activities that make this method free from consumers bias and more effective to specify consumers risk attitude than questionnaire method.

In this work machine learning is used to classify risk attitudes and specify features or consumer behavior that affects the attitude towards risk. It can classify customer into each risk attitude due to their behavior information, and at the same time, the important factor that effect to the classification of customer can be discovered.

II. RELATED WORK

Machine learning has been used extensively to analyse customer behavior. The work in [9] used machine learning to predict the customer cancellation of the service. They use the gradient boosting decision method to identify and rank the importance of the factors in the decision among the raw 898 features of the data set. The nature of the data is such that there are more positive class than negative class as the number of customer who canceled the services is much less than who are still using the service. To tackle this imbalance data, resampling is used. The work in [15] aims to use machine learning to predict values throughout the life span of a customer lifetime and factors affecting the value of passengers using airline services. There are more than 300 features. XGBoost, a gradient boosting decision tree algorithm, is used. With the data of 240,000 passengers, it was able to identify important features.

Data is used in this paper include 20,000 records and about 500 features. The data set contains customers' data who have purchase fund profile from the firm. We found that data is imbalanced. The number of records are very different between different classes in the data set. Learning on data like this leads to reduce accuracy. The classifier classifies all majority example correctly but misclassifies in minority example. Classifier's accuracy is high from the number of majority example more than minority example, but it cannot reflect the performance of classifier to classify minority samples.

There is a lot of techniques for resampling to resolve the imbalance problem. The simplest techniques are random over-sampling and random under-sampling, but both techniques have a drawback. The random over-sampling randomly generates more minority examples. This leads to classifier that is more specific and overfit to the data. On the other hand, random undersampling tries to eliminate some majority example this leads to some useful information being eliminated. Nitesh V. Chawla introduced synthetic minority over-sampling technique [4,6,8] to create synthetic data point on linear interpolation between minority samples and their nearest neighbor. Borderline-SMOTE [7] identifies minority samples and the borderline between majority and minority samples. It synthesises minority class member along the borderline. Adaptive synthetic sampling (ADASYN) [10] is proposed by Haibo He, E.A. Garcia. It uses synthetic sample generation process with weighted resampling density of minority samples around the location where that minority sample located. With this method, the synthetic sample generation process will generate synthetic data around the minority sample which has a larger minority sample's density.

Gustavo E. A. P. A. Batista, Ronaldo C. Prati, Maria Carolina Monard introduced a combination of over-sampling and under-sampling is SMOTE with Tomek Link Removal (SMOTE-Tomek) and SMOTE with Edited Nearest Neighbors (SMOTE-ENN) [11]. Both methods begin with over-sampling by SMOTE, and then SMOTE-Tomek uses Tomek Link Removal (Tomek) [12] as the under-sampling method on the other side SMOTE-ENN (ENN) use Edited Nearest Neighbors for this process. Tomek [13] tries to find a pair of samples which is called a Tomek link, and if they are nearest neighbors and have different class, this pair is removed from datasets or only remove the sample which is majority class. The Edited Nearest Neighbors method (ENN) [14] removes the sample which has the class that differs from a majority class of k nearest neighbors.

III. EXPERIMENT

This work uses a resampling technique to solve imbalance data in the data preparation. We explore three methods: adaptive synthetic sampling (ADASYN), synthetic minority oversampling technique with Edited Nearest Neighbors (SMOTE-ENN) and synthetic minority oversampling technique with TOMEK (SMOTE-Tomek). These techniques are chosen because of they can reduce overfitting problem and data overlapping between classes after resampling.

In the data preparation process, after the irrelevant features are eliminated, the remaining features are divided into two categories: numerical and categorical. The categorical features such as gender are encoded to be discrete value. The features which have numerical value are transformed to the same scale.

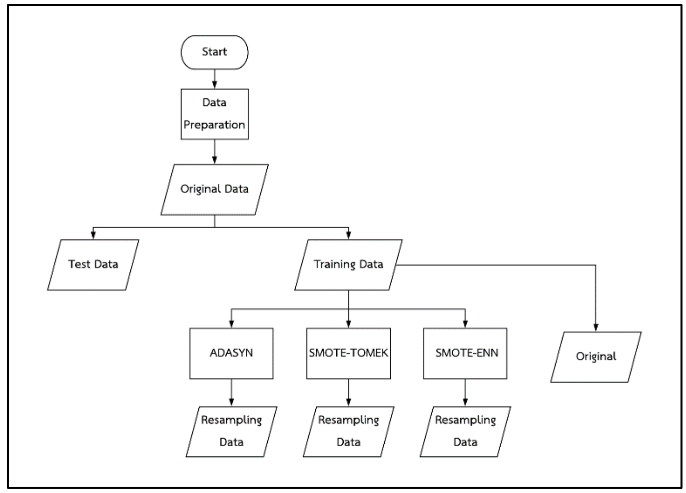

Fig. 1. Data preparation process

After the previous process, the data is imbalance. It is designated as the original data. The original data is divided into the test set and training set in 70:30 ratio. Then the package from imbalanced-learn is used to resamples data in training set. Three techniques are used for resampling data ADASYN, SMOTE-ENN, and SMOTE-Tomek. They generate new data set which are more balance than the original. This data is called resampling data. After this step, there are five data sets: the test set, training set (from original data), data from resampling with SMOTE-ENN, data from resampling with SMOTE-Tomek and data from resampling with ADASYN.

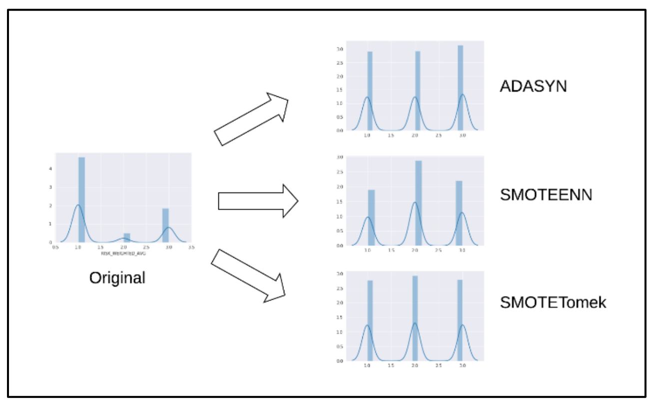

Fig. 2. Class density before and after resampling

We use four data sets (training set and resampling data) from the last step as the source data for training and use the test set for testing. Thee machine learning methods are compared: a decision tree classifier and ensemble methods, i.e. a random forest classifier, Gradient Boosting, and XGBoost [1,3]. Ensemble methods have a particular characteristic that they can provide specific features importance which will be used to define business strategy such as product designing or determine marketing promotion. Because of this reason in this experiment we use ensemble methods in place of neural network or deep learning which may give more accuracy than ensemble methods but difficult to specific features importance.

To find the best parameter for each model, we use a grid search process for hyperparameter tuning. We try to find hyperparameters which make the model has the best area under curve (AUC) score. [2]

Since data is imbalanced, using accuracy only is not appropriate for the evaluation of model's performance. Two performance measures precision and recall take an important role for the case of imbalance data. The goal is to try to improve the model's recall without impact on precision. When trying to improve recall by increasing the true positive of minority class that may be a cause to increase false positive which effect to decrease precision. F-score is harmonic mean of precision and recall, so it is the number which shows the balance between precision and recall [5] (Fig. 3).

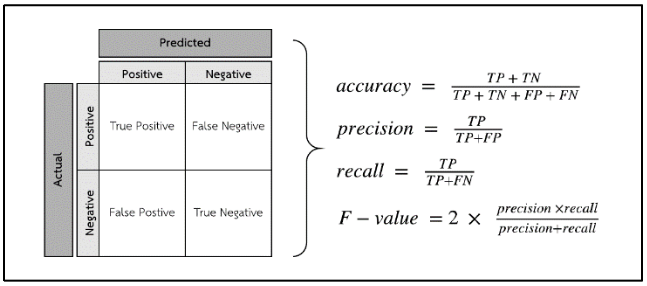

Fig. 3.The performance measure's formula

In addition to accuracy, precision, recall, and F-score, the receiver operating characteristic curve (ROC) is used as another measure model's performance on each class. ROC curve shows the relation of the probability of a true positive rate on Y-axis and a false positive rate (Fig. 4) of each class. The ideal point on the ROC curve is (0,1) which mean the model can classify correctly 100% and the line x = y represented classify of the model has a performance like randomly guessing.

$$true\ positive\ rate = \frac{TP}{TP + FN}$$

$$false\ positive\ rate = \frac{FP}{TN + FP}$$

Fig. 4. The formula of true positive rate and false positive rate

ROC shows the balance of true positive rate and false positive rate that gives information of the tradeoff between them. The area under curve (AUC) is the total area under the ROC curve which is a single number to define and compare the model's performance. The larger AUC means more effective model.

There is a cost from type I errors and type II errors. Type I errors, or false positive is mean that the model classifies that a customer has risk attitude in this class but it is not true. In the other way, Type II errors, or false negative means that the model classifies that a customer does not have a risk attitude in a specified class but actually it is in that class. This error creates opportunity loss from offering an inappropriate product to customer's risk attitude hence make it highly likely that the customer will reject that offering.

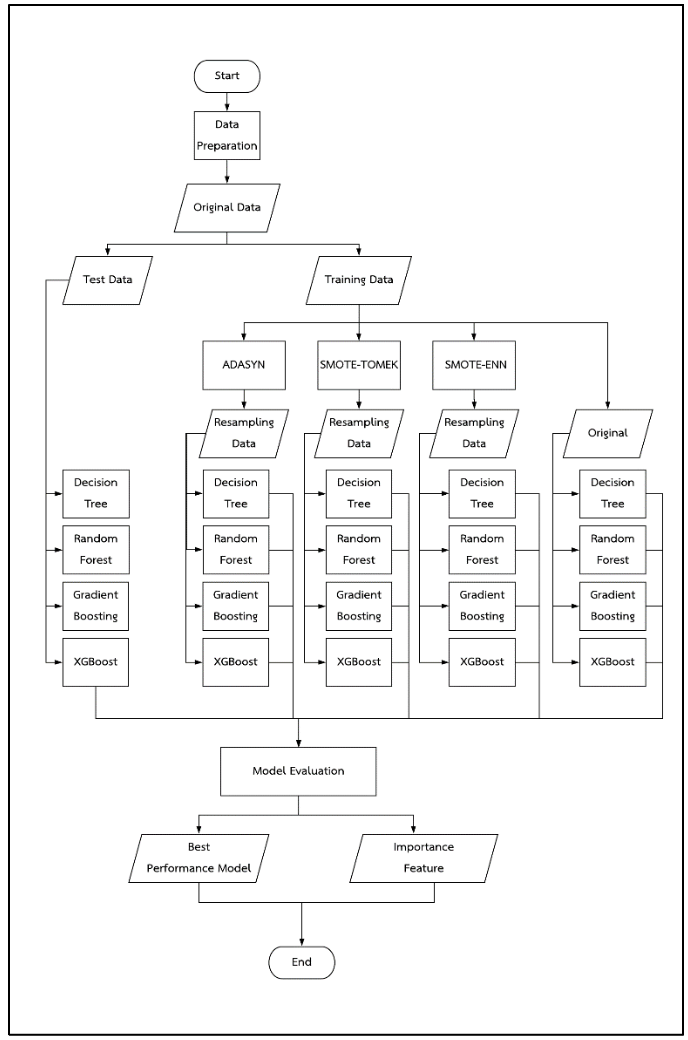

Fig. 5. Overall process

The overall process of the experiment is shown in Fig. 5. The list of the hyperparameters are shown in the Table I. The models are implemented with these hyperparameters to each data from training set and resampling techniques. The results are shown in Table II.

TABLE I CLASSIFIERS AND PARAMETERS FROM GRIDSEARCH

Classifier	**Parameter**
Decision Tree	class_weight=None,criterion='gini',max_depth=10, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None,min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, presort=False, random_state=None, splitter='best'
Random Forest	bootstrap=True, class_weight=None, criterion='entropy', max_depth=None, max_features=170, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=400, n_jobs=None, oob_score=False, random_state=None, verbose=0, warm_start=False
Gradient Boosting	criterion='friedman_mse', init=None, learning_rate=0.5, loss='deviance', max_depth=200, max_features=150, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=150, n_iter_no_change=None, presort='auto', random_state=None, subsample=1.0, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False
XGBoost	base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bytree=0.7, eta=0.05, eval_metric='auc', gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=15, min_child_weight=1, missing=nan, n_estimators=1000, n_jobs=1, nthread=5, objective='multi:softprob', random_state=0, reg_alpha=0, reg_lambda=0.1, scale_pos_weight=1, seed=None, silent=True, subsample=1

IV. RESULT

The result from experiment shows that XGBoost has the best F-score, and AUC. XGBoost with data from SMOTE-TOMEK resampling techniques has highest F1 score (54.19%) but in case of AUC, XGBoost with ADASYN gives the highest AUC.

TABLE II EXPERIMENT'S RESULT

	Decision tree				
	Accuracy	Precision	Recall	F1	AUC
Original	55.42%	49.32%	47.64%	47.85%	60.41%
ADASYN	53.02%	47.82%	49.11%	48.22%	60.74%
SMOTE-ENN	40.35%	45.58%	50.17%	40.20%	61.62%
SMOTE-TOMEK	53.07%	48.10%	49.25%	48.52%	61.39%
	Random Forest				
	Accuracy	Precision	Recall	F1	AUC
Original	63.05%	62.11%	50.53%	49.17%	61.88%
ADASYN	62.35%	58.58%	52.72%	51.93%	62.93%
SMOTE-ENN	40.48%	50.80%	54.25%	39.08%	63.64%
SMOTE-TOMEK	62.30%	58.27%	53.17%	52.27%	63.23%
	Gradient Boosting				
	Accuracy	Precision	Recall	F1	AUC
Original	61.60%	57.25%	51.94%	52.19%	72.88%
ADASYN	60.50%	55.06%	51.98%	52.12%	73.09%
SMOTE-ENN	43.15%	50.26%	54.43%	42.64%	71.53%
SMOTE-TOMEK	61.15%	55.56%	52.99%	52.83%	72.33%
	XGBoost				
	Accuracy	Precision	Recall	F1	AUC
Original	64.32%	61.53%	54.10%	54.16%	76.71%
ADASYN	63.70%	60.11%	53.58%	53.53%	76.97%
SMOTE-ENN	47.17%	52.11%	57.02%	47.02%	74.64%
SMOTE-TOMEK	64.03%	60.38%	54.33%	54.19%	76.95%

According to figures of ROC of XGBoost with various data from different resampling technique (Fig.6-8), it is found that XGBoost with data from ADASYN (Fig.7) can classify class 3 (risk seeker) most correctly and it has AUC 0.88. This means that the model has 88% chance to distinguish between class 3 (positive class) and not class 3 (negative class). From this result, it can be concluded that XGBoost which is trained with data from ADASYN can distinguish the customer who has risk attitude as risk seeker.

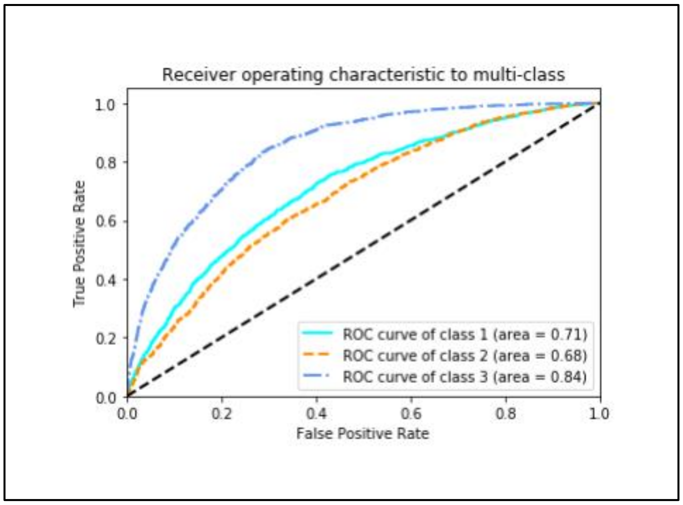

Fig. 6. ROC curve of XGBoost and SMOTE-ENN resampling technique

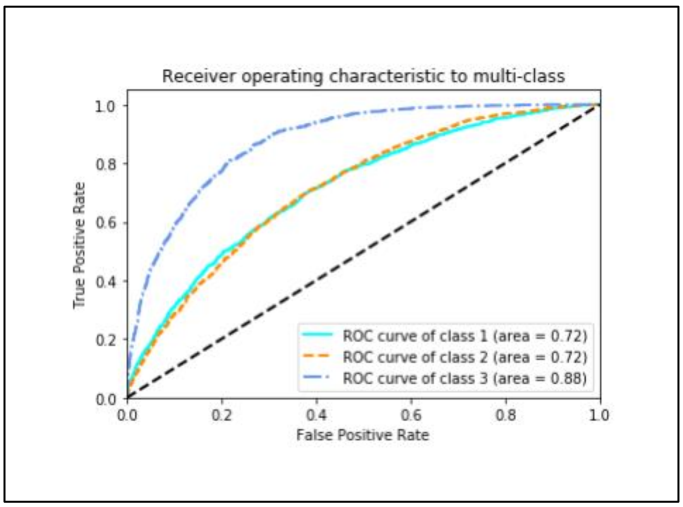

Fig. 7. ROC curve of XGBoost and ADASYN resampling technique

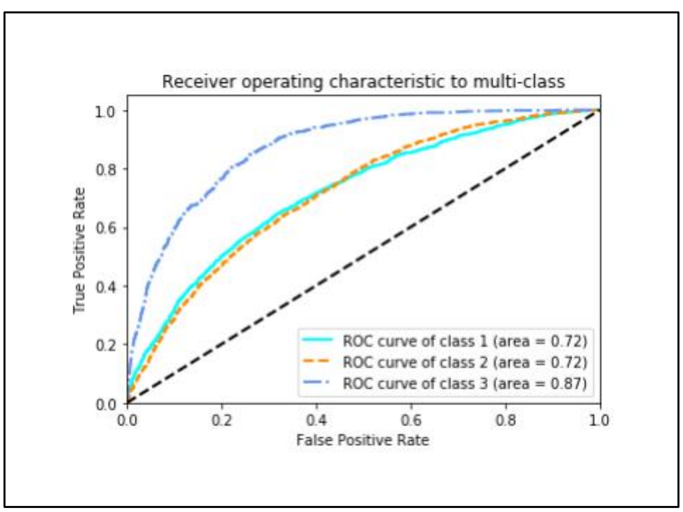

Fig. 8. ROC curve of XGBoost and SMOTE-Tomek resampling technique

The model which uses data from ADASYN can distinguish class 1 (risk aversion), 2 (risk neutral) better than the model which uses data from SMOTE-ENN (Fig. 6) and it is equal to the model which uses data from SMOTE-TOMEK (Fig. 8) . It can be concluded that in case the firm wants to classify customer's risk attitudes with imbalance data ADASYN is the better choice.

To find out the important features, the best model – XGBoost with ADASYN is used. From the Pareto principle or 80/20 rule, only 20 percents of features affects to 80 percents performance of classification. The best 95 features or 20 percents of all features in the best model are used to redo experiment again and the results are shown in Table III.

Table III EXPERIMENT'S RESULT FROM USING 20% OF ALL FEATURES

	XGBoost				
	Accuracy	Precision	Recall	F1	AUC
ADASYN	56.65%	50.19%	48.43%	48.21%	69.28%

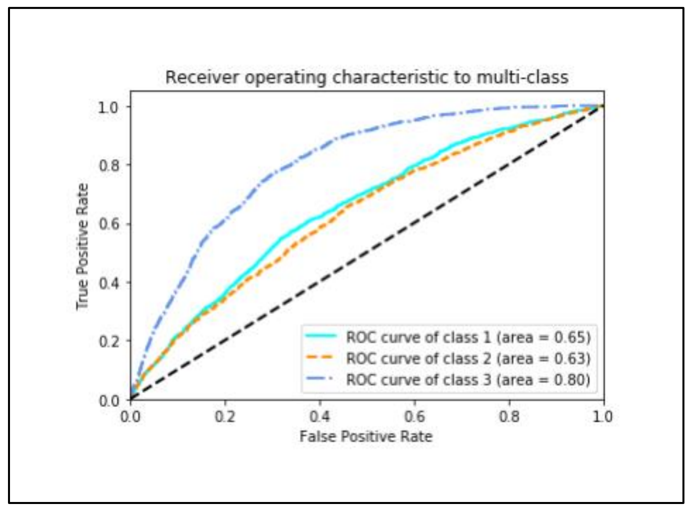

Fig. 9 ROC curve of XGBoost and SMOTE-ENN resampling technique with 20% of all features.

Compare to the performance of all features, the 20% features classifier has the performance drop in all measurements. However, the aim is to identify the important feature. Table III shows that the classification result is acceptable. The result shows that these features are factors which affect to risk attitudes of the customer. The firm should use these factors for planning marketing strategy and designing the financial product to serve their needs and get more customers satisfaction.

V. CONCLUSION

This work proposes a machine learning method to classify customer risk attitude from their historical profile. A gradient boosting decision tree learning method, XGBoost has the best performance when compared to other three methods: decision tree, random forest and gradient boosting. We preprocess the data by resampling to reduce the effect of imbalance data. Using XGBoost with the data processes with ADASYN shows the best performance is all meassurements. Moreover, we are able to identify the important factors that affect customer risk attitude by selecting the 20% highest weight features of the classifier. This information can help the firm to make strategic decision.

ACKNOWLEDGMENT

This work would not have been possible without support data from the customer relationship management department of the commercial bank and our colleagues in that department.

VI. REFERENCE

[1] Aurelien Geron, Hands-On Machine Learning with Scikit-Learn & TensorFlow. CA:O'Reilly Media , 2017.

[2] Ian H. Witten, Eibe Frank and Mark A. Hall, Data Mining Learning Tools and Techniques. MA:Elsevier Inc , 2011.

[3] Peter Bruce and Andrew Bruce, "Chapter5 Classification", in Practical Statistics for Data Scientise, CA:O'Reilly Media, 2017.

[4] Chawala, N., "Chapter 40 Data Mining For Imbalanced Datasets: An Overview", in Data Mining and Knowledge Discovery Handbook, NY: Springer Science+Business Media, Inc, 2005.

[5] Maryuri S. Shelke, Dr. Prashant R. Deshmukh, Prof. Vijaya K. Shandilya, "A Review on Imbalance Data Handling Using Undersampling and Oversampling Technique", International Journal of Recent rends in Engineering & Research,2017.

[6] Chawala, N. V., Bowyer,K.W., Hall, L.O., Kegelmeyer W.P.,"SMOTE: Synthetic Minority Over-Sampling Technique", Journal of Artificial Intelligence Research Society, pp.49-65, 2018.

[7] Hui Han, Wen-Yuan Wang and Bing-Huan Mao, "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning", ICIC 2005, LNCS 3644 ,pp. 878-887,2005.

[8] Chawala, N. V., Bowyer,K.W., Hall, L.O., Kegelmeyer W.P.,"SMOTE: Synthetic Minority Over-Sampling Technique", Journal of Artificial Intelligence Research, vol 16 , pp. 321-357 , 2002.

[9] Wang, Qiu-Feng, Mirror Xu and Amir Hussain "Large-scale Ensemble Model for Customer Churn Prediction in Search Ads." Cognitive Computation, vol.11 , pp.262-270, 2018.

[10] He, H., Bai, Y., Garcia, E.A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322-1328.

[11] Batista, G.E., Prati, R.C., & Monard, M.C. (2004). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explorations, 6*, 20-29.

[12] Boardman, J., & Biron, K. (2018). MITIGATING THE EFFECTS OF CLASS IMBALANCE USING SMOTE AND TOMEK LINK UNDERSAMPLING IN SAS ®.

[13] Tomek, I. (1976). Two Modifications of CNN.

[14] More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. *ArXiv, abs/1608.06048*.

[15] Chen, S. (2018). Estimating Customer Lifetime Value Using Machine Learning Techniques.