Telecommunications Industry: Analysis on Customer Attrition Prediction and Segmentation

Wanrou Zeng *

School of Continuing and Lifelong Education, University of Singapore, Singapore, Singapore

*Corresponding author: e0545952@u.nus.edu

Abstract. Churn prediction is essential for the survival of every business as it allows proactive action planning, whereas customer segmentation is an effective grouping approach commonly used for product marketing and customer relationship management. This paper analyzes the current development of customer attrition prediction and segmentation for the telecommunications (telco) industry, and explores the integration of both approaches that results in an actionable matrix framework for customer retention. This paper consists of four parts: review and analyses on the machine learning techniques used for customer attrition prediction, review and analyses on the machine learning techniques applied for customer segmentation, review on the integration framework, and finally provides discussions for future works. The background and business applications of these approaches are introduced in this paper so businesses can apply the techniques for their business needs. Most importantly, this paper lays out the directions of future works to be done for telco churn management based on the improvements needed form the current state of the art.

Keywords: Customer attrition; Customer segmentation; Telecommunications.

1. Introduction

Churn management in the telco industry is extremely important as the main business model is subscription-based. A small increment in churn can have a huge impact on enterprise value because losing a customer means the loss of its ongoing subscription revenue. The reason of churn can be categorized as voluntary or involuntary. The former happens when a customer initiates the termination of service due to problems such as price or customer service, whereas involuntary churn happens when a customer is removed by the company because of reasons such as fraud, non-payment, or underutilization [1]. Since the drivers contributing to voluntary churn are within control by service providing companies, churn prediction as a proactive analytics measure is adopted in the telco industry. Predicting customer churn helps businesses plan their resources and actions to prevent churn or mitigate the impact if it happens. This paper analyzes the current development of customer attrition prediction, specifically in the machine learning methods applied for the prediction procedure. The methods and authors reviewed are listed in Table 1.

Table 1. Churn prediction methods

Proposed methods	Researchers
Binary Logistic Regression	[2]
	[3]
Decision Tree	[3]
	[4]
Random Forests	[3]
	[4]
Naïve Bayes	[3]
AdaBoost	[3]
Multi-layer Perceptron	[3]
Logit Leaf Model (new hybrid method)	[6]
Multi-stacking ensemble model	[7]

However, simply predicting if a customer will churn is not enough to plan for any effective marketing activities because the reasons and situations can be vastly different for each individual. For businesses to better understand the characteristics and profiles of their customers, the customer segmentation technique is applied to group customers with common or similar traits into one cluster, such as a cluster that has a major weightage in senior people, or a cluster that is consisted of people with monthly subscription. This paper analyzes the current development of customer attrition prediction machine learning methods applied for customer segmentation across research. The methods and authors reviewed are listed in Table 2.

Table 2. Customer segmentation methods

Proposed methods	Researchers
Hierarchical Agglomerative Clustering	[8]
RFM Model Using K-Means	[9]

After reviewing methods used for churn prediction and customer segmentation, this paper looks at a particular work [5] that seamlessly combines churn prediction and customer segmentation techniques together to form a robust integrated framework. The framework first implements churn prediction on customers, then extracts the predicted churners, and carries our customer segmentation on the predicted churners. This approach brings further insight into the data by understanding the characteristics of the predicted churners in groups. For example, group one is likely to churn because people in this group share the common characteristics of a, b, and c, whereas group two is likely to churn because people in this group share the common characteristics of c, d, and e. With this information, companies can manage these groups separately with the different marketing activities according to their unique needs, so that the likelihood of churning of these predicted churners can be lower.

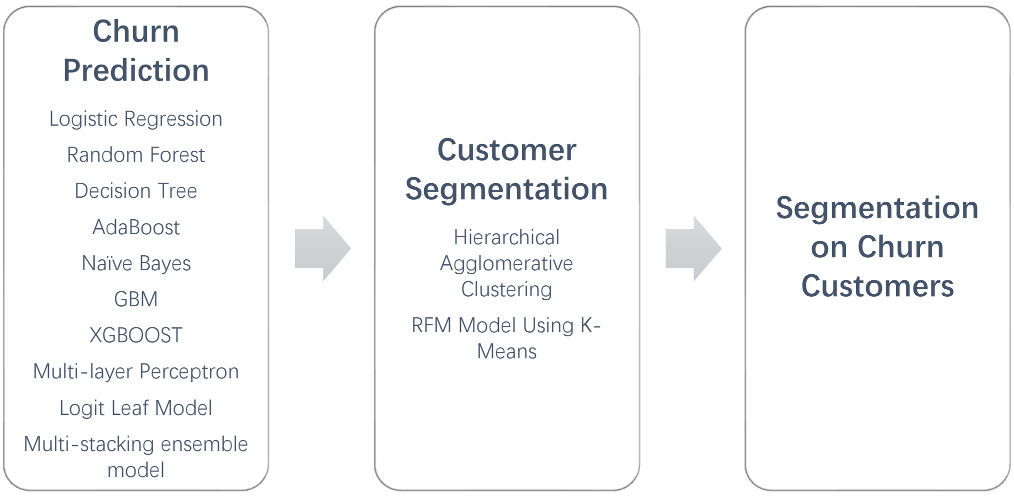

Fig. 1 Flowchart of the Review Process

2. Churn Prediction

This section describes the machine learning methods applied across research for churn prediction. The definitions and results of these methods are included. Cross validation is utilized to evaluate model performance with Precision, Accuracy, Recall, F1-score, and Receiver Operating Characteristic Curve (ROC) as the measure metrics. Precision demonstrates the ratio of true positive over the total number of positively predicted samples, whereas Accuracy indicates the percentage of the number of correct outputs over the entire outputs. Recall represents the number of positive and negative correctly predicted samples in the actual sample. Finally, F1-score brings together the result of recall and precision, whereas ROC calculates the false positive and true positive rates as presented in the formulas as below [3]:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = \frac{2*Recall*Precision}{Recall+Precision} \tag{4}$$

$$FPR = \frac{FP}{FP+TN} \tag{5}$$

$$TPR = \frac{TP}{TP+FN} \tag{6}$$

2.1 Binary Logistic Regression

It is a specialized form of regression for predicting the probability of binary responses. The logistic function outputs values between 0 and 1 for all inputs X, with values closer to 1 indicating a higher probability of event to occur (1) given X, as presented in Formula 7. Coefficients α and β indicate how much the dependent variables will change given a one-unit increase on a predictor [10].

$$Y = \frac{\exp(\alpha+\beta_1 X_1+\ \beta_2 X_2+\cdots+\ \beta_k X_k)}{1+\exp(\alpha+\ \beta_1 X_1+\ \ \beta_2 X_2+\cdots+\ \ \beta_k X_k)} + e. \tag{7}$$

The study on China Mobile, China Telecom, and China Unicom [2] applied binary logistic regression to predict telco client churn based on 4,126 client data. The results show that call information factors, SMS, and expense are the influential drivers that contribute to churn. The prediction accuracy is measured to be 93.94%, and the paper concluded to adopt logistic regression model for future prediction. However, such a high accuracy might be the result of overfitting, and there was no further validation in the study to feed in new or a more substantial dataset to testify the model.

Regardless of the high accuracy score, logistic regression is also observed to be the most effective model in another churn prediction study [3] with the accuracy of 80.19%. The overall performance is shown in Table 3.

Table 3. Logistic regression performance without SMOTE [3]

Methods (without vs with SMOTE)	Acc		Pre		Rec		F1		AUC	
Logistic Regression	80.19	74.82	65.17	51.74	54.57	78.76	59.37	62.43	84.36	84.39
AdaBoost	80.08	77.19	65.39	55.44	53.24	73.35	58.61	63.11	84.51	84.52
Random Forest	79.55	76.99	66.10	55.14	47.51	73.25	55.25	62.86	83.79	83.80
Decision Tree	78.33	76.74	61.50	54.97	50.72	72.07	55.28	62.26	82.37	82.83
Naïve Bayes	75.14	73.76	52.32	50.42	73.51	78.01	61.12	61.24	82.30	82.24
Multiple Layer Perceptron	80.12	75.60	65.46	52.88	53.40	76.30	58.76	62.45	84.06	84.05

Interestingly, applying Synthetic Minority Oversampling Technique (SMOTE) had increased the recall score for logistic regression but decreased the accuracy, as seen in Table 3. This is because SMOTE put more weight to the minority class, so logistic regression can predict the minority with higher accuracy, but the overall accuracy will decrease. In this case, it is more suitable to look at the F1-score, which is a combination of recall and precision.

2.2 Decision Trees

Decision Trees have a flow-chart-like structure that represents a top-down approach with each node being a condition check. The decision tree algorithm utilizes Attribute Selection Measures (ASM) to select the best attributes, and makes the attribute become a decision node [11]. This process is carried out recursively until there is no remaining attribute.

There is study that compares Decision Tree and Logistic Regression on predicting customer churn, where Decision Tree outperformed Logistic Regression with an AC of 99.67% [4].

2.3 Random Forest

Random Forest is a classifier combining a collection of tree-structured classifiers where each tree from T_1 to T_k is a single decision tree model and acts as an input variable to vote for the best classification result [12] (see in Figure 2).

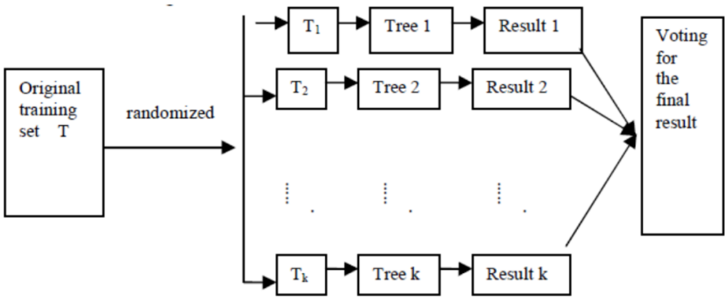

Fig. 2 Random Forests Schema

2.4 Naïve Bayes

Naïve Bayes algorithm is a classification method that works based on the assumption of conditional independence between variable pairs, given class variable y and dependent feature vector x [13]. This assumption is made so that the multi-variate probabilities can be estimated from training data. Although Naïve Bayes has shown remarkable accuracy in the past, it is unlikely that the conditional independence assumption is true in reality, resulting in one of the drawbacks of this algorithm [14]:

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1,\dots,x_n|y)}{P(x_1,\dots,x_n)}. \tag{8}$$

2.5 AdaBoost

AdaBoost is a meta-estimator which recursively fits predicting function classifiers on the same dataset with increasingly adjusted weights for incorrectly classified instances. Each predicting function gives weight with the effective learner having a higher weight. Combining adjusted instances, the final predicting function is formed after recursive times of circulation, with a weighted majority voting method [15].

2.6 Multi-layer Perceptron

Multi-layer Perceptron (MLP)can be represented by a finite acyclic graph that is trained to learn a non-linear function approximator f(.) = R^m→R^o where m is the input dimensions count and o is the output dimensions count. It can learn one or more non-linear layers, also known as hidden layers [16] (see in Figure 3).

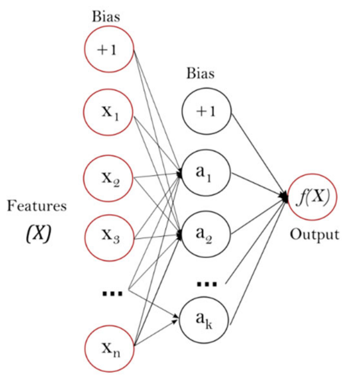

Fig. 3 MLP Finite Acyclic Graph

2.7 Gradient Boosted Machine Tree (GBM)

GBM is an iterative ensemble tree-based model that creates a strong predictive learning model by combining many weak learning models for supervised learning. It is commonly used due to its flexibility which supports heterogeneous data, but this iterative approach also has a drawback of lacking rigorous finite-time convergence guarantees [17].

2.8 Logit Leaf Model (new hybrid method)

LLM is a new hybrid classification algorithm developed based on logistic regression and decision trees [6]. LLM uses forward selection to fit logistic regression only at the terminal nodes. LLM is designed to address the problem of decision trees having difficulties handling linear relations between variables, and logistic regression have difficulties due to the interaction effects between variables (see in Figure 4). LLM was found to be more efficient and accurate with better performance and comprehensibility.

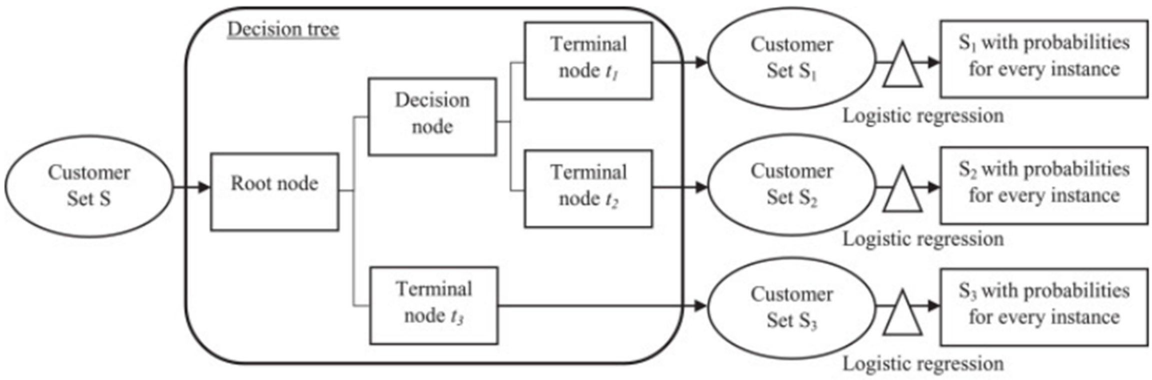

Fig. 4 LLM Conceptual Representation

2.9 Multi-stacking ensemble model

Multi-stacking ensemble model is a final prediction model that combines results of multiple machine learning models, which are, in this case, Logistic Regression, Random Forests, Naïve Bayes, and Extreme Gradient Boosting [7]. This study achieved a better result with the stacking model compared to a standalone model, obtaining the AUC of 81.24%.

To have a holistic and comprehensive comparison across all the machine learning models above and more, a study was done [18] in which over 100 classifiers were applied for telco churn prediction. The results show that Regularized Random Forest and Bagging Random Forest were the most effective classifiers with the AUC of 73.04% and 67.20% respectively.

Overall, it is fair to conclude that machine learning models perform differently within various circumstances based on the training and testing data. It is also a common process to apply Logistic Regression as a baseline model. To select the most suitable model for telco churn prediction, the most essential step is to define the objective of a problem. For example, if business planning only requires true and false churn rates, then Linear Classification can be sufficient. On the other hand, if marketers require the probability of churn, then Logistic Regression may be applied.

3. Customer Segmentation

This section reviews the methods for customer segmentation and compares their performance results.

3.1 Hierarchical Agglomerative Clustering

Hierarchical Agglomerative Clustering (HAC), also known as the bottom-up approach, merges single data points in succession close to one another until data points with similar characteristics are grouped together [19]. The number of clusters k does not need to be specified, as the tree-like hierarchy can be visualized with all values of k in a dendrogram [8] (see in Figure 5). As an example, HAC technique was applied on telco fraud detection, to test if fraud cases from the same customer profile category (behavior representation) have similar characteristics and under which condition will this profile become a fraud [8].

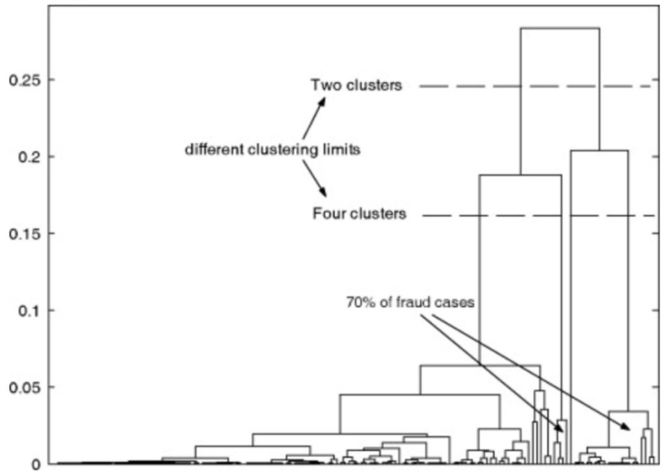

Fig. 5 Hierarchical Agglomerative Clustering of a Customer's Profile Category

3.2 RFM Analysis and K-Means

RFM analysis is a commonly used marketing technique that determines customer's recency, frequency, and monetary values using quantitative metrics [20]. The RFM values can then be ranked or grouped by clustering methods. In this case, it is K Means separates customers into groups with similar characteristics. K-Means algorithm using Euclidean distance metric (can be changed) is applied to partition RFM values amongst customers [9]. The clustering results can be validated by Silhouette Coefficient (see in Figure 6), where a value closer to 1 indicates a better clustering outcome.

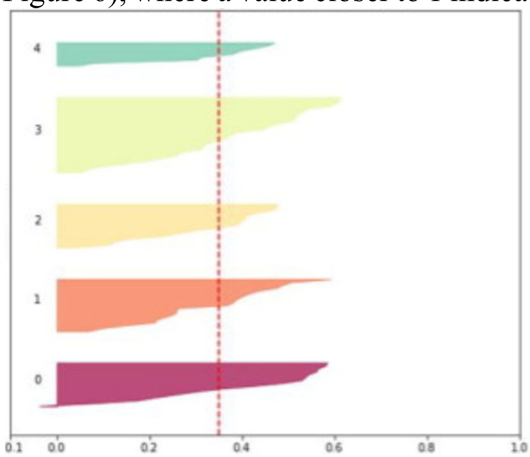

Fig. 6 Visualization of Cluster = 5

4. Integration of Customer Attrition Prediction and Segmentation

Few studies were carries out on both customer attrition prediction and segmentation. A standalone churn prediction only allows companies to know which customers are likely to leave, without knowing the characteristics of these churners. Conducting customer segmentation on predicted churners will provide a more rigorous group-based retention measures according to the similar patterns within each cluster. Following customer segmentation, essential churn factors can also be identified to determine the root causes of churn. Being well-informed of the significance of the churn factors can tremendously improve marketing activities of the company. Two studies were found [3][5] that focus only on churn class data, and identify the similar characteristics of customers by segmentation. Appropriate marketing policies can then be developed for churners only. To construct a more rigorous churn-segmentation matrix, the predicted churn customers can be further partitioned into three groups: low, medium, and high likelihood to churn, forming a two-dimensional churn-segmentation matrix [5]. The table below illustrates an example of the matrix as a result. Actions taken for each group can be evaluated and planned by marketers.

Table 4. Churn and Segmentation Framework

Segment	Low	Medium	High
High total calls and high online revenue	Example: keep engaged	Example: keep engaged	Example: manage closely
Low total calls and low online revenue	...	...	...
High total calls and low online revenue	...	...	...
Low total calls and high online revenue	...	...	...

Furthermore, a recommender system can be derived from the churn-segmentation framework using a similarity measure [5] based on similar past customer behavior and preferences. For example, customers who had high total calls and high revenue in the past tend to buy more cellphones and have a 500 GB add-on, which means that having cellphones and data related offers would interest this type of churners.

5. Discussion

This section discusses some future works and areas that can be explored on the subject of business analytics in the telco industry.

5.1 Matrix of Customer Value Measurement and Churn Prediction

Similar to the churn-segmentation framework mentioned in section 4, a framework of customer value and churn can be one of the areas to explore (see in Table 5). Customer value is computed by factors that what customer attributes companies think are valuable to their businesses. Attributes such as margin, recency, frequency, and cancelation rates are the common contributing factors for customer value. These customer values can be represented by numerical values, with bigger numbers indicating higher customer values. Finally, to group the customer values, businesses can set their own threshold to define which threshold can be regarded as high, medium, and low values of customers. The advantage of such framework is so that companies can prioritize focusing more resource allocation on the customers with high customer values.

Table 5. Customer Value and Churn Prediction Framework

Segment	Low churn probability	Medium churn probability	High churn probability
High customer value	Example: keep engaged	Example: keep engaged	Example: manage closely
Medium customer value	...	...	...
Medium customer value	...	...	...

5.2 Churn Prediction on Product Level

Telco businesses have been expanding their scope throughout the years, as the competition is getting more rampant due to the advent of new technologies. With an expanded business scope and product lines, customers might have high churn rate in one product but low in another. For example, telco businesses have subscription-based products such as phone plan, sim card plan, broadband, tv content, etc. These products are fundamentally different catering to unique customer needs, and the churn rate for tv content might be naturally higher than sim card plan, as tv entertainment is more akin to an optional luxury than an essential necessity. Breaking down the customer dataset by product line helps improve the accuracy of churn prediction and the effectiveness of product marketing.

5.3 Building an Enterprise Datawarehouse (EDW)

Across all the research reviewed in this paper, some datasets were obtained across companies, whereas some studies were done on multiple datasets containing different attributes for the same group of customers. Data collection and pre-processing can be a painful process if data are incomplete, non-comprehensive, and small-sized. Information silos and data silos are common problems that companies face when the infrastructure development throughout different systems are not on the same pace. It is also a result of lack of communication across departments and systems. Solving the data silo issue can result in more data being obtained and fully utilized for insight discovery. To achieve this, an EDW can be built as a central repository with full data integration from all systems, and limited access for people with appropriate access credentials. EDW is a core component for business intelligence as it brings a large variety of parameters, making data mining, feature selection, and feature engineering much more effective.

6. Conclusion

In general, this paper reviews the state of the art in churn prediction and customer segmentation for the telco industry, demonstrates the integration of both subjects, and discusses the benefits of adopting such an integration framework. Machine learning methods including Binary Logistic Regression, Decision Tree, Random Forests, Random Forests, AdaBoost, Multi-layer Perceptron, GBM, Logit Leaf Model, and Multi-stacking Ensemble Model for churn prediction were reviewed in this paper. It is concluded that models performed to different degrees of effectiveness across different experiments and datasets, as the circumstances are different. However, the rule of thumb is to take Logistic Regression as the baseline model, as it has low complexity and computation time. Logit Leaf Model and Multi-stacking Ensemble Model combine multiple models together to address the drawback of a standalone model. They usually yield a better result regarding accuracy with the compromise of computation time. Measurements such as F1-score and ROC were utilized for performance evaluation. Moreover, Hierarchical Agglomerative Clustering and K-Means methods for customer segmentation were reviewed, where the former did not require a specification of K, and the latter required the number of clusters as an input. Silhouette Coefficient were utilized for performance evaluation. For further observation, an integration of churn prediction and customer segmentation was reviewed, which was able to provide insightful analytics for businesses to take actions accordingly. Finally, three recommendations were given for future works, including constructing a framework of customer value and segmentation, conducting churn prediction from a product level, and building an EDW for a more robust business intelligence and analytics.

References

[1] Mattison, R. The Telco Churn Management Handbook. Xit Press, 2006.

[2] Zhang, T., Moro, S., & Ramos, R. F. A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation. Future Internet, 2022, 14(3), 94.

[3] Wu, S., Yau, W. C., Ong, T. S., & Chong, S. C. Integrated churn prediction and customer segmentation framework for telco business. IEEE Access, 2021, 9, 62118-62136.

[4] Dahiya, K., & Bhatia, S. Customer churn analysis in telecom industry. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE, 2015.

[5] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE access, 2018, 7, 60134-60149.

[6] De Caigny, A., Coussement, K., & De Bock, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research, 2018, 269(2), 760-772.

[7] Vo, N. N., Liu, S., Li, X., & Xu, G. Leveraging unstructured call log data for customer churn prediction. Knowledge-Based Systems, 2020, 212, 106586.

[8] Hilas, C. S., & Mastorocostas, P. A. An application of supervised and unsupervised learning approaches to telecommunications fraud detection. Knowledge-Based Systems, 2008, 21(7), 721-726.

[9] Anitha, P., & Patil, M. M. RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University-Computer and Information Sciences, 2019.

[10] Osborne, J. W. (Ed.). Best practices in quantitative methods. Sage, 2007.

[11] Decision Tree Classification in Python Tutorial. 2018. Datacamp. https://www.datacamp.com/tutorial/decision-tree-classification-python.

[12] Liu, Y., Wang, Y., & Zhang, J. New machine learning algorithm: Random forest. In International Conference on Information Computing and Applications (pp. 246-252). Springer, Berlin, Heidelberg, 2012.

[13] Naive Bayes. 2022. Scikit-Learn. https://scikit-learn.org/stable/modules/naive_bayes.html.

[14] Chen, S., Webb, G. I., Liu, L., & Ma, X. A novel selective naïve Bayes algorithm. Knowledge-Based Systems, 2019, 192, 105361.

[15] AdaBoostClassifier. 2022. Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html#:~:text=An%20AdaBoost%20%5B1%5D%20classifier%20is,focus%20more%20on%20difficult%20cases.

[16] Neural network models (supervised). 2022. Scikit-Learn. https://scikit-learn.org/stable/modules/neural_networks_supervised.html.

[17] GradientBoostingClassifier. 2022. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html.

[18] Adhikary, D. D., & Gupta, D. Applying over 100 classifiers for churn prediction in telecom companies. Multimedia Tools and Applications, 2020, 80(28), 35123-35144.

[19] Lone, H., & Warale, P. Cluster Analysis: Application of K-Means and Agglomerative Clustering for Customer Segmentation. Journal of Positive School Psychology, 2022, 7798-7804.

[20] Wright, G. What is RFM (recency, Frequency, monetary) analysis and how does it work? SearchDataManagement, 2022.