IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 5, 2021, accepted April 6, 2021, date of publication April 16, 2021, date of current version April 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073776

Integrated Churn Prediction and Customer Segmentation Framework for Telco Business

SHULI WU[1], WEI-CHUEN YAU[1], (Member, IEEE), THIAN-SONG ONG[2], (Senior Member, IEEE), AND SIEW-CHIN CHONG[2], (Senior Member, IEEE)

[1]School of Electrical and Computer Engineering, Xiamen University Malaysia, Sepang 43900, Malaysia
[2]Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia

Corresponding authors: Wei-Chuen Yau (wcyau@xmu.edu.my) and Thian-Song Ong (tsong@mmu.edu.my)

This work was supported in part by the Xiamen University Malaysia Research Fund under Grant XMUMRF/2019-C4/IECE/0011, and in part by the Multimedia University Mini Fund under Grant PRJMMUI/180251.

ABSTRACT In the telco industry, attracting new customers is no longer a good strategy since the cost of retaining existing customers is much lower. Churn management becomes instrumental in the telco industry. As there is limited study combining churn prediction and customer segmentation, this paper aims to propose an integrated customer analytics framework for churn management. There are six components in the framework, including data pre-processing, exploratory data analysis (EDA), churn prediction, factor analysis, customer segmentation, and customer behaviour analytics. This framework integrates churn prediction and customer segmentation process to provide telco operators with a complete churn analysis to better manage customer churn. Three datasets are used in the experiments with six machine learning classifiers. First, the churn status of the customers is predicted using multiple machine learning classifiers. Synthetic Minority Oversampling Technique (SMOTE) is applied to the training set to deal with the problems with imbalanced datasets. The 10-fold cross-validation is used to assess the models. Accuracy and F1-score are used for model evaluation. F1-score is considered to be an important metric to measure the models for imbalanced datasets since the premise of churn management is to be able to identify customers who will churn. Experimental analysis indicates that AdaBoost performed the best in Dataset 1, with accuracy of 77.19% and F1-score of 63.11%. Random Forest performed the best in Dataset 2, with accuracy of 93.6% and F1-score of 77.20%. Random Forest performed the best in Dataset 3 in terms of accuracy, at 63.09%, while Multi-layer Perceptron performed the best in terms of F1-score, at 42.84%. After implementing churn prediction, Bayesian Logistic Regression is used to conduct the factor analysis and to figure out some important features for churn customer segmentation. Churn customer segmentation is then carried out using K-means clustering. Customers are segmented into different groups, which allows marketers and decision makers to adopt retention strategies more precisely.

INDEX TERMS Telco business, churn prediction, Bayesian analysis, customer segmentation.

I. INTRODUCTION

The advent of 5G technology and the shifting customer preferences have created a lot of opportunities for telecommunication (telco) companies. Such immense business opportunities have also led to fierce competition in the telco market, accompanied by a high customer churn rate. It is pressing for telco operators to come up with effective marketing strategies based on extensive customer analytics to prevent customer turnover and bolster company revenue.

Customer analytics in the telco industry consists of two key components, namely churn prediction and customer segmentation. As the telco market tends to be saturated, too much focus on attracting new subscribers is no longer applicable to the telco industry. The cost of attracting new customers by investing considerable resources is confirmed to be significantly higher than the cost of retaining existing customers [1]. In this context, churn management becomes instrumental in the telco industry. Churn management includes identifying

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu.

customers who are likely to churn and making relevant recommendations based on their characteristics. There are two types of methods for managing customer churn: reactive and proactive [2]. Most current studies, for example [3], considered the proactive method to be a better strategy. Operators are recommended to identify those customers who are about to churn before they actually churn. Customer churn prediction allows operators to have a period of time to remediate and implement a series of tactical retention measures before existing customers migrate to other operators. On the other hand, customer segmentation is an important mean to perform customer analytics, which targets customers into several different groups according to different schemes. The most common segmentation method in the telco industry is segmentation based on customer value and behaviour [4]. Through effective customer segmentation, telco operators are able to provide differentiated products and personalised services and carry out precision marketing based on customer needs and consumption characteristics in different customer segments.

To help telco marketers make better decisions, churn prediction is supposed to be combined with customer segmentation. Additionally, telco operators often need more than just predictions about whether customers will churn. They also need a more detailed analysis of the factors that may cause churn as well as the overall probability of customer churn. Among all the customers who are about to churn, in practice, telco operators are not supposed to take the same measures for every single customer. To be more specific, not all customers are of high-value, telco operators should spend more retention resources on those high-value customers. And for some customers who are not very helpful to the company's revenue, telco operators do not have to pay too much attention to them.

This paper proposes an integrated customer analytics framework for churn management in telco industry, aiming to achieve efficient company resource allocation and improve customer retention. Specifically, this work provides insights on customer churn prediction and customer segmentation while combining Bayesian Analysis to seamlessly connect these two parts. After predicting whether the customer will churn, Bayesian Analysis will be used to select important factors and feed them into customer segmentation, which makes the segmentation process more effective. And the characteristics of each cluster will be analysed to provide recommendations to the operator and lay the foundation for further countermeasures in the future.

II. LITERATURE REVIEW

A. CHURN PREDICTION

In the telco business, most of the earlier researches related to churn prediction focused on a few machine learning classifiers. To make comparisons and get the best predictive model, [5] performed three experiments on different feature sets with 7 predication methods, which are Logistic Regression, Linear Classification, Naïve Bayes, C4.5 Decision Tree (C4.5), Multilayer Perceptron, Support Vector Machine (SVM), and Data Mining by Evolutionary Learning (DMEL). C4.5 and SVM were considered more effective in their research. DMEL method was found impractical on a large dataset context. They indicated that different methods can be used depending on different marketing objectives since the outputs of them are different. If strategists only interested in churn rates, SVM or Linear Classification can be used. If probabilities of churn are required, then Naïve Bayes, Logistic Regression can be more effective. In the research of [6], Induja and Eswaramurthy proposed Kernelized Extreme Learning Machine (KELM) algorithm to predict the customer churn patterns. Their methods achieved AUC of 83%.

Verbeke *et al.* [7] also explored different ways on multiple datasets to predict customer churn, with twenty-one methods, including Naïve Bayes, Random Forest, Support Vector Machine, Gradient Boosting, Decision Tree and so on. They conducted experiments both with and without oversampling and both with and without input selection. They obtained the highest AUC of 97.2% on Dataset of Operator in East Asia with ID O5, using Alternating Decision Tree.

In the research of [8], Logistic Regression and Decision Tree were used to predict customer churn. Their results showed that Decision Tree outperformed Logistic Regression in their datasets. And they obtained the highest accuracy of 99.67% on their large dataset using Decision Tree. KNN, Random Forest and XGBoost were used by [9] to predict customer churn. In their research, XGBoost obtained the highest accuracy of 79.8% and the highest AUC of 58.2%.

KNN, Naïve Bayes, C4.5, Random Forest, AdaBoost, and ANN were used by [1] to predict customer turnover. They used Synthetic Minority Over-sampling Technique (SMOTE) to balance the instances in their dataset. Different parameter combinations of different algorithms were explored to obtain the most adequate model. In their research, KNN, C4.5, and Random Forest classifiers performed well in AUC value. The results also showed that Random Forest ranked top, at 91.10%. They indicated that older data contained in the datasets may negatively affect their experimental results. Naïve Bayes, Random Forest were also used by [10]. Their results indicated that Random Forest performed better than Naïve Bayes, with the accuracy of 71.99%.

In the research of [11], they compared the performance of more than 100 classifiers in the churn prediction problem in the telecom industry. Their results showed that Regularized Random Forest obtained the highest accuracy of 73.04% and Bagging Random Forest outperformed all other classifiers in terms of AUC, standing at 67.20%.

Idris *et al.* [12] studied undersampling methods based on particle swarm optimisation (PSO) with different feature reduction methods and used random forest and KNN classifiers. Experimental results showed that the method based on PSO, mRMR and RF (Chr-PmRF) performed the best in terms of AUC, at 75.11%.

In the research of [13], Ahmed and Maheswari adopted meta-heuristic algorithms to predict the loss of telecom customers. Their results show that the Firefly algorithm is suitable while the hybrid Firefly algorithm can provide effective and faster results. In the research of [14], Vijaya and Sivasankar used particle swarm optimisation (PSO) and combined feature selection and simulated annealing (SA) to predict customer churn. Their method was compared with traditional machine learning algorithms, and the results showed that PSO-FSSA performed the best, with the accuracy of 94.08% and F1 score of 96.06%.

In the research of [15], Ahmed *et al.* used deep learning method and proposed a method called ''TL-DeepE'', which first conducted Transfer Learning (TL) by fine-tuning multiple pre-trained deep convolutional neural networks (CNN). They converted the telecom dataset into 2D image format. Then, they used these CNNs as base classifiers, and the Genetic Programming (GP) and AdaBoost as the meta-classifier. Their method obtained accuracy of 75.4% and 68.2%, and AUC of 83% and 74%, for the Orange and Cell2Cell datasets, respectively.

In the research of [16], Amin *et al.* proposed an intelligent rule-based decision-making method based on rough set theory. They used four algorithms, which are Exhaustive Algorithm, Genetic Algorithm, Covering Algorithm and LEM2 Algorithm. Their results showed that the Genetic Algorithm based on rough set theory is the most effective method, with accuracy of 98.1% and F1 score of 92.5%. In the research of [17], Amin *et al.* proposed a novel method of customer churn prediction. They divided the dataset into different regions based on the distance factor and then divide it into data with higher certainty and data with lower certainty to predict customer churn. Their results show that the distance factor has a great relationship with the certainty of the classifier, and the greater the distance factor, the greater the accuracy of the classifier.

Due to the possibility of missing historical data, [18] used cross-company dataset to predict customer churn. In order to explore the performance of data conversion methods in cross-company churn prediction, they made extensive comparisons and adopted the methods of Naïve Bayes, K-Nearest Neighbour, Gradient Boosted Tree, Single Rule Induction and Deep learner Neural net. Experimental results show that Naïve Bayes achieved the highest AUC values 0.51, 0.51, 0.513 in raw, log, and Box-Cox transformation. In the research of [19], Amin *et al.* proposed a just-in-time churn prediction method. The cross-company dataset was used for training, and the results show that the heterogeneous ensemble-based just-in-time model is more suitable for predicting customer churn, with accuracy of 77.27% and F1 score of 71.42%.

Six oversampling methods were explored by [20] for handling the problems with imbalanced datasets in the telco industry. These six methods are Adaptive Synthetic Sampling Approach (ADASYN), Couples Top-N Reverse k-Nearest Neighbour (TRkNN), Immune Centroids Oversampling Technique (ICOTE), Mega-trend Diffusion Function (MTDF), SMOTE, and Weighted Minority Oversampling Technique (MWMOTE). They applied these oversampling methods on 4 public telco datasets to predict customer churn, with 4 rules-generation algorithms (LEM2, Covering, Exhaustive, and Genetic algorithms). Their experiments showed that MTDF with the Genetic algorithm performed the best.

B. CUSTOMER SEGMENTATION

Telco customer segmentation was introduced in four schemes in [4] and they pointed out that different segmentation methods can be applied to different specific business purposes. They gave us some insights that combined value segmentation with behavioural segmentation, the customer base can be sub-divided by several levels of value and behavioural clusters. And each of these sub-segments can be further characterised by attributes such as their retention likelihood, their stickiness, their promotion score, and their average value.

A novel customer segmentation method based on the customer life cycle was proposed by [21] to identify high-value customers more effectively. Customer value was divided into direct value, including historical value, long-term value, current value, and indirect value. The calculation of these five parts constitutes five models. These five components were assigned weights by experts and the high-value customers were identified through ranking.

A customer segmentation method was proposed by [22] for Jiangsu Changzhou Telco by using K-means clustering and commercial automated tool KXEN. The customer segmentation was done within small business customers in two dimensions, namely values and behaviours. Customers are divided into six groups based on their value and five groups based on their behaviour, and a crossing matrix was illustrated. By analysing the customer characteristics of each segment, marketing analysts were able to understand customers' needs and recommend their favourite business or packages. Practical results showed that this method was successful and effective in Jiangsu Changzhou Telco.

Namvar *et al.* [23] also proposed a 2-dimensional segmentation to segment telco users in both behavioural and beneficial phases using K-means clustering. Usage-based features were applied to the behavioural segmentation, while revenue-based features were applied by the beneficial segmentation. Both studies have shown that customer segmentation with a two-dimensional method has better results than considering all features in one dimension.

C. CHURN PREDICTION AND CUSTOMER SEGMENTATION

Few researches have explored both churn prediction and customer segmentation. In the research of [24], exploratory data analysis (EDA) was first done to explore those possible variables in the database that contributed to customer churn. Two methods were used to build predictive models. One was to assess the customer churn behaviours in different 'value-loyalty' segments. First K-means clustering method was used

to segment the customers into five groups. Then a decision tree (C5.0) model was built in each cluster to predict customer churn. In the other method, the entire dataset was used, and the customer churn was predicted, using Back Propagation Neural Network (BPN) followed by the Decision Tree. However, Hung *et al.* focused only on predicting clients' churn. Predictive models can only let operators know which customers are leaving without providing effective retention measures to be proposed to meet the needs of those customers.

To combined churn prediction and customer segmentation, [25] proposed a framework to predict customer churn and conduct customer profiling. They first used Random Forest to predict customer churn. Then, in order to get a deeper insight of important factors that cause customer churn, factor identification was conducted using Attribute Selected Classifier. After that, they extracted all churn data that Random Forest predict correctly, and conducted a customer profiling to see similarity of these churn customers. Finally, based on the results of customer profiling, some retention strategies and recommendations were proposed.

III. RESEARCH MOTIVATION AND CONTRIBUTIONS

The cost of acquiring new customers is 5 to 10 times higher than the cost of retaining existing customers [26], and the churn rate of new customers is often higher than existing customers [27]. For telco operators, the adoption of a retention strategy can obtain greater profits by regaining trust from customers who have already enjoyed their services and products. However, there are more and more telco operators in the same areas nowadays, and every operator is constantly updating its own services and products. In this context, telco customers have too many choices. As the shifting cost of telco customers reduces, it is increasingly difficult for telco operators to retain existing customers. The high service expectations of telco customers also lead to the high marketing costs of telco operators. The considerable customer base and daily generated information lead to rich telco databases. How to make use of data mining methods to effectively utilise these data and maintain existing customer resources has become one of the hot spots in this industry.

Churn management is supposed to be implemented promptly and effectively. However, a critical problem for churn prediction is that customers in telco operators will not easily turnover in a short period of time [28]. Most of them will choose to stay with their operators and continue to enjoy the services rather than shift to another operator. This situation causes the datasets in the telco industry to be imbalanced and thereby affects a lot when doing churn prediction. It is interesting to explore some advanced re-sampling methods and looking for the most suitable metrics to evaluate the models. In our research, SMOTE is applied to the training datasets, and different metrics are compared and discussed.

Most researchers only focused on one element of customer analytics, that is, either churn prediction or customer segmentation. And the majority of current researches applied segmentation to the whole customer dataset [21], [22] [23]. However, If only churn prediction is conducted, it is not able to understand the reasons behind it well, since the operator can only know which customers are likely to churn. If only customer segmentation is conducted, the operator can only know the customer characteristics of different clusters, and it is not able to focus on the churn customer and responds differently to clusters with different characteristics. To develop a more effective retention program and reduce the management cost, customer segmentation is supposed to be focused on churn customers only. Therefore, this research aims to fill the gaps as mentioned above. Combining churn prediction and customer segmentation, an integrated telco customer analytics framework for churn management is proposed. In research of [25], Ullah *et al.* conducted both churn prediction and customer segmentation. However, in order to connect them seamlessly, our proposed framework adds an intermediate process, that is, the Bayesian Analysis. At present, there is only very limited research that applying the Bayesian method to the churn analysis. And so far, there is no research that combines Bayesian Analysis with churn prediction and customer segmentation seamlessly. The research on telco customers cannot be limited to identifying which customers are more likely to churn. It is more important to find out the reasons behind customer churn [29]. And the models are supposed to provide the churn factors for marketers to better understand the reason behind churn [30]. The Bayesian Analysis is implemented to conduct the factor analysis in this research. And by utilising it as an intermediate process, the key factors contributing to customer churn can be selected before segmentation to group the customers more effectively. Particularly, in addition to factor analysis, through Bayesian Analysis, the overall probability of churning for 3 datasets can be calculated in the research. In this way, factors that contribute to the result of churning can be well analysed, and then different retention strategies can be proposed to different customer groups, so as to achieve effective churn management and precision marketing.

IV. INTEGRATED CUSTOMER ANALYTICS FRAMEWORK

In this research, an integrated customer analytics framework is proposed as shown in Figure 1. First, data cleaning, data transformation, and data normalisation are carried out in the pre-processing process. This is followed by the exploratory data analysis (EDA) step which consists of uni-variate and bi-variate analysis. The purpose of conducting EDA is to make sense of the data and help us better understand each feature before feeding them to machine learning models, thereby, making the modelling more efficient.

Secondly, six machine learning classifiers are used to predict customer churn, that is, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, AdaBoost, and Multi-layer Perceptron. Justified from the literature review [1], [5] [7], these classifiers are common and performed well in different researches of churn prediction, and therefore, they are used in the research for performance comparison purpose. Besides, the oversampling method SMOTE

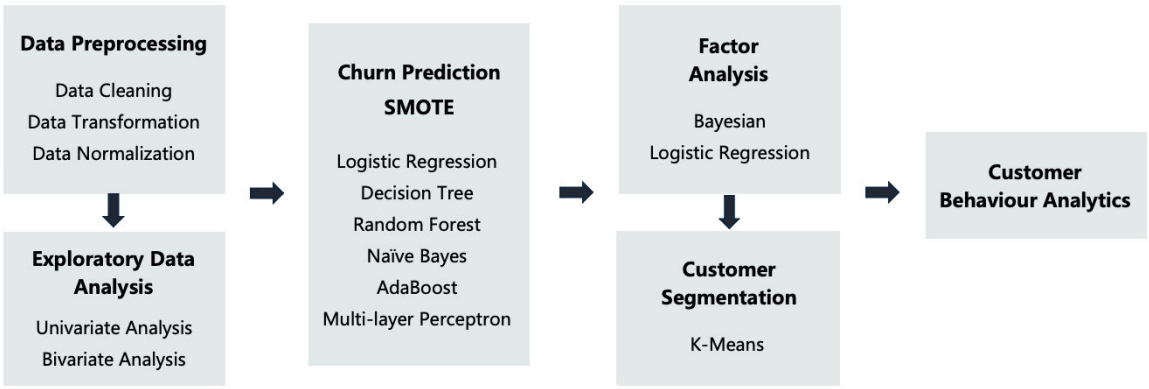


FIGURE 1. Integrated telco customer analytics framework.

is applied to the training set to improve the performance of the classifier. For each sample x in the minority class, SMOTE first utilises the Euclidean distance to calculate its distance to all samples in the minority class in the sample space, and then its k-Nearest Neighbour samples are obtained. The default value of k is 5. For each randomly selected neighbour $\widetilde{x}$, a new sample is constructed using (1), where x stands for the original sample, $\widetilde{x}$ represents the neighbour sample and x_{new} is the synthetic sample. Variants of SMOTE, such as SMOTE-ENN and SMOTE Tomek, have an additional cleaning mechanism, which clean up overlapping synthetic samples that are difficult to distinguish from majority class samples. And also some of other hybrid sampling methods mentioned in [20] outweighed SMOTE in their research. However, different methods behave differently in different datasets. In our research, we mainly focus on SMOTE and its variants, and among them SMOTE performs the best in our datasets. Thus, SMOTE is selected as to balance the instances of two classes.

$$x_{new} = x + random(0, 1) \times (\widetilde{x} - x) \tag{1}$$

Thirdly, factor analysis is conducted using Bayesian Logistic Regression to figure out some important factors behind the churning. This process is called Bayesian Logistic Regression modelling, which is used to prepare useful features for more precise customer segmentation. The customer segmentation is conducted only on the churn data, since the goal of this research is to perform churn management. K-means clustering is a popular method for segmentation, so it is used in this research to segment the churn customers. The characteristics of each cluster can then be obtained.

Finally, based on the results of churn prediction and customer segmentation, customer behaviour analytics are conducted. The uncertainty of churn is visualised. Different machine learning classifiers are compared. And the characteristics of each churn segment are summarised, which can be useful for marketers or strategists to come up with different retention measures on different segments.

This paper utilises Bayesian Analysis as the intermediate process to connect churn prediction and customer segmentation seamlessly. After performing the churn prediction, Bayesian Logistic Regression is used to conduct the factor analysis, aiming to find the reason behind churn, and provide some important factors for churn customer segmentation. Other than traditional probability theory, Bayesian believes that probability is a subjective concept of individuals, indicating how much individuals believe in the occurrence of something. Bayesian Analysis refers to a process of obtaining the probability of an event by accumulating evidence. It points out that when predicting something, the first thing is to infer a prior probability based on existing experience and knowledge, and then adjust this probability as new evidence accumulates. Bayesian Logistic Regression combines the Bayesian theorem with Logistic Regression model. Bayesian methods use probability to quantify the uncertainty [31]. To conduct Bayesian Analysis, custom priors are chosen first. Normal distribution with mean and standard deviation is defined for prior distribution for each parameter θ. Then, different parameters are drawn from the prior distribution and put into the logistic regression model to obtain simulated data, which are compared with actual data. And those parameters result in inconsistent data with actual data are filtered away. Finally, the posterior can be obtained [32], [33] [34]. In the posterior probability distribution, the parameters with the highest frequency can be observed, and at the same time the uncertainty can be expressed as a probability. The posterior probability can be expressed as (2).

$$P(\theta|D) = \frac{P(\theta) \cdot P(D|\theta)}{\sum P(\theta) \cdot P(D|\theta)} \tag{2}$$

After the modelling, the probability of churn can be visualised for each attribute, in order to find the reason behind churning. The most important factors that contribute to churning are selected by odds ratio for further segmenting the churn customers into different groups. The characteristics of each cluster are analysed and the recommendations are made for the operator to better manage the churn customers. And based on the result of Bayesian Analysis and customer segmentation,

the overall probability of each cluster are calculated using (3), where n is the number of customers in a cluster, and z is the linear combination of a series of features represented by (4). Here x_0 represents the offset, $x_0, \cdots, x_m$ represents each feature of customers, and $w_0, \cdots, w_m$ represents the weight of each feature.

$$P = \frac{g(z)}{n} = \frac{\frac{1}{1+e^{-z}}}{n} \tag{3}$$

$$z = x_0 + w_1x_1 + w_2x_2 + \ldots + w_mx_m \tag{4}$$

V. EXPERIMENTAL RESULT ANALYSIS

A. SETUP

This experiment is carried out using Jupyter Notebook with Python 3 on a 2.7 GHz Intel Core i5 processor, RAM 8GB. The features and the corresponding data formats and descriptions are shown in following tables.

Three datasets are used for experiments. Dataset 1 is a sample dataset from IBM used to predict customer churn status to develop customer retention programs [35]. Dataset 2 was released from the latest Kaggle telco customer churn prediction competition 2020 [36]. Only the train document is used for Dataset 2, since the test document does not contain the churn label of customers. Dataset 3 is called Cell2Cell provided by Teradata centre for customer relationship management at Duke University [37]. The features without pre-processing for three datasets are shown in Table 1-4.

The number of samples in each experiment for three datasets are shown in Table 5. The hyper-parameters are tuned for all experiments. And Table 6 shows the optimal hyper-parameters obtained from empirical experiments, while other hyper-parameters are set to default values.

B. PERFORMANCE METRICS

Cross-validation is used to assess the performance of models, which groups the original data into different parts. The training set is used to train the model and the test set is used to evaluate the performance of models. Cross-validation is beneficial to obtain a reliable and stable model since the out-of-sample data are validated as well. In this research, 10-fold cross-validation is used. The dataset is divided into 10 segments, and 9 of them are used as training data, and the remaining 1 segment is used as a test set. Each test will give the corresponding performance. And the overall performance is the average of all tests.

Accuracy, Precision, Recall, F1-score and AUC are measured to assess the performance of models. Accuracy refers to the ratio of the number of samples that the model predicts correctly to the total number of samples. Precision indicates the ratio of truly positive samples among the samples predicted to be positive. Recall represents how many positive examples in the actual sample are predicted correctly, while specificity represents how many negative examples in the actual sample are predicted correctly. Expanding on the basis of these four metrics, F1-score is generated. It combines the results of Precision and Recall. The equations are shown as (5)-(8), where TP (True Positive) indicates that the sample is actually a positive sample and is predicted to be a positive sample. FN (False Negative) indicates that the sample is actually a positive sample and is predicted to be a negative sample. FP (False Positive) indicates that the sample is actually a negative sample and is predicted to be a positive sample. TN (True Negative) indicates that the sample is actually a negative sample and is predicted to be a negative sample.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

TABLE 1. Features in Dataset 1 without pre-processing.

#	Features	Data Format	Description
1	customerID	*string*	Customer ID
2	gender	*(Female /Male)*	Whether the customer is a male or a female
3	SeniorCitizen	*(0/1)*	Whether the customer is a senior citizen or not
4	Partner	*(Yes/No)*	Whether the customer has a partner or not
5	Dependents	*(Yes/No)*	Whether the customer has dependents or not
6	tenure	*numerical*	Number of months the customer has stayed with the company
7	PhoneService	*(Yes/No)*	Whether the customer subscribes a phone service or not
8	MultipleLines	*(Yes/No /No phone service)*	Whether the customer subscribes multiple lines or not
9	InternetService	*(DSL/Fiber optic /No)*	Customer's internet service provider
10	OnlineSecurity	*(Yes/No /No internet service)*	Whether the customer subscribes online security or not
11	OnlineBackup	*(Yes/No /No internet service)*	Whether the customer subscribes online backup or not
12	DeviceProtection	*(Yes/No /No internet service)*	Whether the customer subscribes device protection or not
13	TechSupport	*(Yes/No /No internet service)*	Whether the customer subscribes tech support or not
14	StreamingTV	*(Yes/No /No internet service)*	Whether the customer subscribes streaming TV or not
15	StreamingMovies	*(Yes/No /No internet service)*	Whether the customer subscribes streaming movies or not
16	Contract	*(Month-to-Month /One Year /Two Year)*	The contract term of the customer
17	PaperlessBilling	*(Yes/No)*	Whether the customer uses paperless billing or not
18	PaymentMethod	*(Bank Transfer /Credit Card /Electronic Check / Mailed Check)*	The customer's payment method
19	MonthlyCharges	*numerical*	The amount charged to the customer monthly
20	TotalCharges	*numerical*	The total amount charged to the customer
21	Churn	*(Yes/No)*	Whether the customer churned or not

TABLE 2. Features in Dataset 2 without pre-processing.

#	Features	Data Format	Description
1	state	*string*	2-letter code of the US state of customer residence
2	account_length	*numerical*	Number of months the customer has been with the current telco provider
3	area_code	*string*	"area_code_AAA" where AAA = 3-digit area code
4	international_plan	*(yes/no)*	Whether the customer has international plan
5	voice_mail_plan	*(yes/no)*	Whether the customer has voice mail plan
6	number_vmail _messages	*numerical*	Number of months the customer has stayed with the company
7	total_day_minutes	*numerical*	Total minutes of day calls
8	total_day_calls	*numerical*	Total number of day calls
9	total_day_charge	*numerical*	Total charge of day calls
10	total_eve_minutes	*numerical*	Total minutes of evening calls
11	total_eve_calls	*numerical*	Total number of evening calls
12	total_eve_charge	*numerical*	Total charge of evening calls
13	total_night_minutes	*numerical*	Total minutes of night calls
14	total_night_calls	*numerical*	Total number of night calls
15	total_night _charge	*numerical*	Total charge of night calls
16	total_intl_minutes	*numerical*	Total minutes of international calls
17	total_ intl _calls	*numerical*	Total number of international calls
18	total_ intl _charge	*numerical*	Total charge of international calls
19	number_customer _service_calls	*numerical*	Number of calls to customer service
20	churn	*(yes/no)*	Whether the customer churned or not

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \tag{8}$$

The ROC curve of the models can be drawn by calculating the values of the False Positive Rate (FPR) and True Positive Rate (TPR) of the classification model and taking them as the horizontal and vertical axes, respectively. FPR and TPR are calculated as (9)-(10). Area Under the Curve (AUC) is the area under the ROC curve. Similarly, a larger value of AUC is better.

$$FPR = \frac{FP}{FP + TN} \tag{9}$$

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

C. RESULTS OF DATASET 1

There is a total of 19 features in Dataset 1 used for churn prediction as shown in Table 7.

1) RESULTS WITHOUT SMOTE

It is observed from Table 8 that Logistic Regression got the highest accuracy at 80.19%. Random Forest outweighs other classifiers in precision with 66.10%. Naïve Bayes got the highest recall, and F1-score which is 73.51%, and 61.12%,

TABLE 3. Features in Dataset 3 without pre-processing.

#	Features	Data Format	Description
1	CustomerID	*int*	Customer ID
2	Churn	*(Yes/No)*	Whether the customer churned or not
3	MonthlyRevenue	*float*	Mean monthly revenue
4	MonthlyMinutes	*float*	Mean monthly minutes of use
5	TotalRecurringCharge	*float*	Mean total recurring charge
6	DirectorAssistedCalls	*float*	Mean number of director assisted calls
7	OverageMinutes	*float*	Mean overage minutes of use
8	RoamingCalls	*float*	Mean number of roaming calls
9	PercChangeMinutes	*float*	% Change in minutes of use
10	PercChangeRevenues	*float*	% Change in revenues
11	DroppedCalls	*float*	Mean number of dropped voice calls
12	BlockedCalls	*float*	Mean number of blocked voice calls
13	UnansweredCalls	*float*	Mean number of unanswered voice calls
14	CustomerCareCalls	*float*	Mean number of customer care calls
15	ThreewayCalls	*float*	Mean number of threeway calls
16	ReceivedCalls	*float*	Mean unrounded mou received voice calls
17	OutboundCalls	*float*	Mean number of outbound voice calls
18	InboundCalls	*float*	Mean number of inbound voice calls
19	PeakCallsInOut	*float*	Mean number of in and out peak voice calls
20	OffPeakCallsInOut	*float*	Mean number of in and out off-peak voice calls
21	DroppedBlockedCalls	*float*	Mean number of dropped or blocked calls
22	CallForwardingCalls	*float*	Mean number of call forwarding calls
23	CallWaitingCalls	*float*	Mean number of call waiting calls
24	MonthsInService	*int*	Months in service
25	UniqueSubs	*int*	Number of unique subscription
26	ActiveSubs	*int*	Number of active subscription
27	ServiceArea	*string*	Communications service area
28	Handsets	*float*	Handsets issued
29	HandsetModels	*float*	Handset models issued
30	CurrentEquipmentDays	*float*	Number of days of the current equipment
31	AgeHH1	*float*	Age of first HH member
32	AgeHH2	*float*	Age of second HH member
33	ChildrenInHH	*(Yes/No)*	Presence of children in HH
34	HandsetRefurbished	*(Yes/No)*	Whether the handset is refurbished or not
35	HandsetWebCapable	*(Yes/No)*	Whether the hanset is web capable or not
36	TruckOwner	*(Yes/No)*	Whether the subscriber owns a truck or not
37	RVOwner	*(Yes/No)*	Whether the subscriber owns a recreational vehicle or not
38	Homeownership	*(Known Unknown)*	Whether the home ownership is missing or not
39	BuysViaMailOrder	*(Yes/No)*	Whether the subscriber buys via mail order or not

TABLE 4. Features in Dataset 3 without pre-processing (Continued).

#	Features	Data Format	Description
40	RespondsToMailOffers	*(Yes/No)*	Whether the subscriber responds to mail order or not
41	OptOutMailings	*(Yes/No)*	Whether the subscriber has chosen not to be solicited by mail or not
42	NonUSTravel	*(Yes/No)*	Whether the subscriber has traveled to non-US country or not
43	OwnsComputer	*(Yes/No)*	Whether the subscriber owns a personal computer or not
44	HasCreditCard	*(Yes/No)*	Whether the subscriber possesses a credit card or not
45	RetentionCalls	*int*	Number of calls previously made to retention team
46	RetentionOffersAccepted	*int*	Number of previous retention offers accepted
47	NewCellphoneUser	*(Yes/No)*	Known to be a new cell phone user
48	NotNewCellphoneUser	*(Yes/No)*	Known not to be a new cell phone user
49	ReferralsMadeBySubscriber	*int*	Number of referrals made by subscriber
50	IncomeGroup	*int*	Income group
51	OwnsMotorcycle	*(Yes/No)*	Whether the subscriber owns a motorcycle or not
52	AdjustmentsToCreditRating	*int*	Number of adjustments made to customer credit rating
53	HandsetPrice	*string*	Handset price
54	MadeCallToRetentionTeam	*(Yes/No)*	Whether the subscriber has made call to retention team
55	CreditRating	*string*	The credit rating of the subscriber
56	PrizmCode	*(Rural /Suburban/ Town / Other)*	The prizm code of the subscriber
57	Occupation	*(Clerical /Crafts /Homemaker /Professional /Self /Student /Retired /Other)*	The occupation of the subscriber
58	MaritalStatus	*(Yes/No /Unknown)*	The marital status of the subscriber

TABLE 5. Dataset configuration.

Dataset	Training Set		Test Set		Total
	Non-Churn	Churn	Non-Churn	Churn	
Original Dataset 1	4647	1628	516	187	7032
Dataset 1 with SMOTE	4647	4647	516	187	9997
Original Dataset 2	3138	490	349	54	4031
Dataset 2 with SMOTE	3138	3138	349	54	6679
Original Dataset 3	32702	13240	3634	1471	51047
Dataset 3 with SMOTE	32702	32702	3634	1471	6679

respectively. It is noticable that the recall value is not very good, which means that it is not able to find many customers who actually churn, leading to the churn management

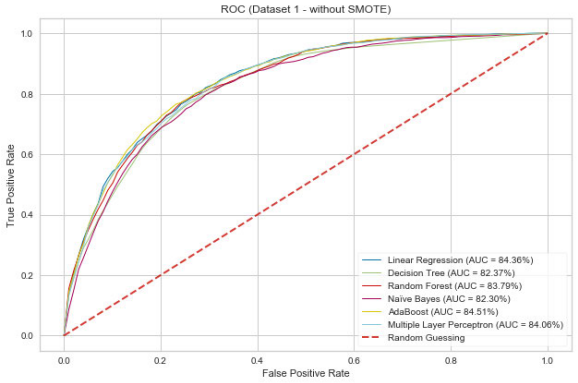


FIGURE 2. ROC curves for Dataset 1 without SMOTE.

ineffective. The ROC curves are shown in Figure 2. It can be observed that the curves of Logistic Regression and AdaBoost are close to each other. They achieved the similar performance in terms of AUC, while AdaBoost performed slightly better, with 84.51%.

2) RESULTS WITH SMOTE FOR DATASET 1

It is found from Table 9 that generally after oversampling, the recall, F1-score and AUC are improved. Logistic Regression obtains the highest recall, with 78.76%. AdaBoost obtains the highest accuracy, precision, and F1-score, with 77.19%, 55.44%, and 63.11%, respectively. This time, more customers who actually turnover are identified, which is considered better in churn management, compared to previous results without SMOTE. The ROC curves are shown in Figure 3. Similarly, the curves of Logistic Regression and AdaBoost are found close to each other, while AdaBoost performed slightly better as well, with AUC of 84.52%.

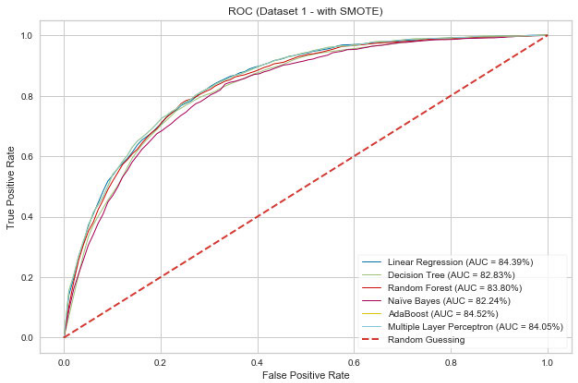


FIGURE 3. ROC curves for Dataset 1 with SMOTE.

The best results are compared with previous works of [9], [17] using the same telco operator, as shown in Figure 4. Our method shows good performance in terms of accuracy and F1-score.

TABLE 6. Hyper-parameters configurations for all experiments.

Classifiers	Hyperparameters	Dataset 1	Dataset 1 - SMOTE	Dataset 2	Dataset 2 - SMOTE	Dataset 3	Dataset 3 - SMOTE
Logistic Regression	penalty	l1	l2	l2	l2	l1	l1
	solver	liblinear	liblinear	sag	sag	liblinear	liblinear
Decision Tree	max_features	10	15	15	15	15	10
	max_leaf_nodes	20	20	20	20	25	35
Random Forest	n_estimators	10	10	10	35	10	25
	max_features	10	15	15	15	10	20
	max_leaf_nodes	15	25	30	45	35	45
AdaBoost	n_estimators	25	60	15	25	17	50
	learning_rate	1.0	1.0	0.5	0.6	0.4	1.0
Multi-layer Perceptron	solver	adam	adam	lbfgs	lbfgs	adam	adam
	hidden_layer_sizes	(10,)	(10,)	(50,)	(25,)	(50,)	(30,)

TABLE 7. Features of Dataset 1 for churn prediction.

#	Features	#	Features
1	Gender	**11**	Tech Support
2	Senior Citizen	**12**	Streaming TV
3	Partner	**13**	Streaming Movies
4	Dependents	**14**	Paperless Billing
5	Phone Service	**15**	Contract
6	Multiple Lines	**16**	Payment Method
7	Internet Service	**17**	Tenure
8	Online Security	**18**	Monthly Charges
9	Online Backup	**19**	Total Charges
10	Device Protection		

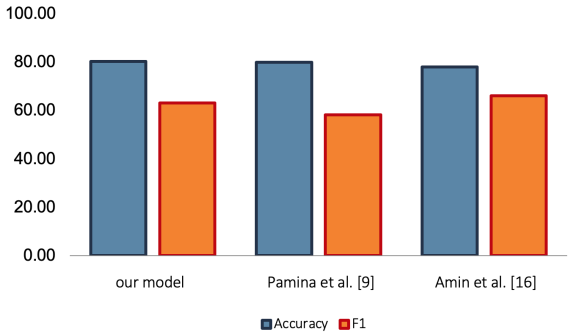


FIGURE 4. Comparisons for Dataset 1.

TABLE 8. Results without SMOTE for Dataset 1.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	**80.19**	65.17	54.57	59.37	84.36
Decision Tree	78.33	61.50	50.72	55.28	82.37
Random Forest	79.55	**66.10**	47.51	55.25	83.79
Naïve Bayes	75.14	52.32	**73.51**	**61.12**	82.30
AdaBoost	80.08	65.39	53.24	58.61	**84.51**
Multiple Layer Perceptron	80.12	65.46	53.40	58.76	84.06

TABLE 9. Results with SMOTE for Dataset 1.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	74.82	51.74	**78.76**	62.43	84.39
Decision Tree	76.74	54.97	72.07	62.26	82.83
Random Forest	76.99	55.14	73.25	62.86	83.80
Naïve Bayes	73.76	50.42	78.01	61.24	82.24
AdaBoost	**77.19**	**55.44**	73.35	**63.11**	**84.52**
Multiple Layer Perceptron	75.60	52.88	76.30	62.45	84.05

TABLE 10. Features of Dataset 2 for churn prediction.

#	Features	#	Features
1	Account Length	**10**	Total Eve Calls
2	Area Code	**11**	Total Eve Charge
3	International Plan	**12**	Total Night Minutes
4	Voice Mail Plan	**13**	Total Night Calls
5	Number Vmail Messages	**14**	Total Night Charge
6	Total Day Minutes	**15**	Total Intl Minutes
7	Total Day Calls	**16**	Total Intl Calls
8	Total Day Charge	**17**	Total Intl Charge
9	Total Eve Minutes	**18**	Number Customer Service Call

TABLE 11. Results without SMOTE for Dataset 2.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	87.42	63.84	19.32	29.21	82.05
Decision Tree	94.59	88.68	68.55	77.07	87.91
Random Forest	**95.34**	**91.71**	71.90	80.44	**91.34**
Naïve Bayes	88.14	56.30	55.71	55.94	84.04
AdaBoost	88.27	71.00	21.87	33.12	87.45
Multiple Layer Perceptron	95.26	89.85	**73.16**	**80.51**	91.21

D. RESULTS OF DATASET 2

There is a total of 18 features in Dataset 2 used for churn prediction as shown in Table 10.

1) RESULTS WITHOUT SMOTE

It is observed from Table 11 that Random Forest outperformed all other classifiers in terms of accuracy and

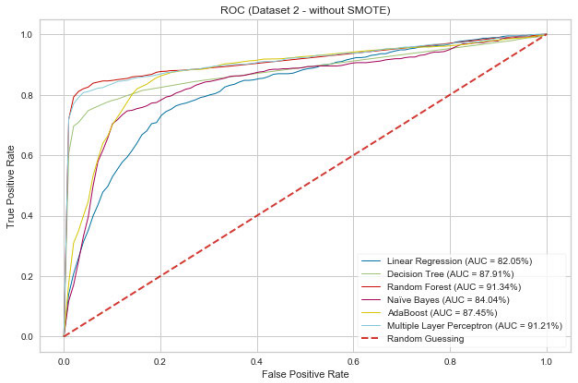


FIGURE 5. ROC curves for Dataset 2 without SMOTE.

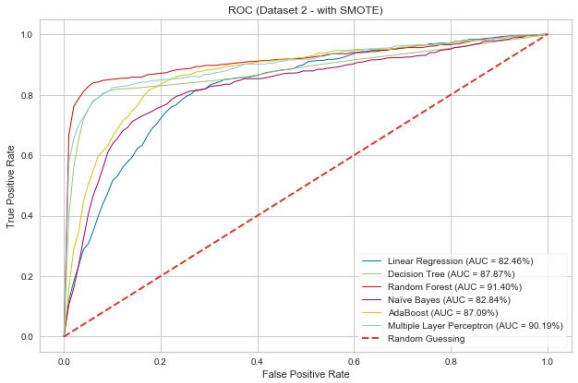


FIGURE 6. ROC curves for Dataset 2 with SMOTE.

precision, which are 95.34% and 91.71%, respectively. While Multi-layer Perceptron got the highest recall and f1-score, at 73.16% and 80.51%, respectively. The ROC curves are shown in Figure 5. It is found that the curves of Random Forest and Multi-layer Perceptron are both close to left upper corner, with some intersections. However, Random Forest performed the best for Dataset 2 without SMOTE, with AUC of 91.34%.

TABLE 12. Results with SMOTE for Dataset 2.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	77.18	34.42	75.37	47.21	82.46
Decision Tree	92.58	70.67	79.04	74.22	87.87
Random Forest	**93.60**	**74.63**	**80.71**	**77.20**	**91.40**
Naïve Bayes	78.64	36.62	77.04	49.54	82.84
AdaBoost	86.21	49.51	66.53	55.61	86.20
Multiple Layer Perceptron	90.37	61.03	79.95	69.17	90.19

2) RESULTS WITH SMOTE

Random Forest performed the best in all metrics with accuracy of 93.60%, precision of 74.63%, recall of 80.71%, and F1-score of 77.20%, as shown in Table 12. The ROC curves are shown in Figure 6. It can be observed that Random Forest and Decision Tree perform better in term of accuracy, while Random Forest obtained the highest AUC of 91.40%.

Particularly, it is obvious that the accuracy drops as compared with previous model without SMOTE, but this does not mean that the model is worse. On the contrary, this model is more practical. It is observed that its recall value and AUC are improved.

At the same time, some useful insights can be obtained from these results. In the extremely imbalanced test set, accuracy should not be considered as the best suitable metric for comparison. The research of [38] indicated that in an imbalanced dataset, the selection of metrics must also consider the preferences of the user. If all samples are classified as majority class, accuracy may be significantly higher. However, when the minority class is of more interest, the result of this kind of higher accuracy is worthless. For example, if a classifier predicts all samples into non-churn groups, and it still gets 85% accuracy. In practice, it does not identify any single customer who is about to leave, which is contrary to the purpose since the goal of this research is to retain customers and develop an effective churn management program.

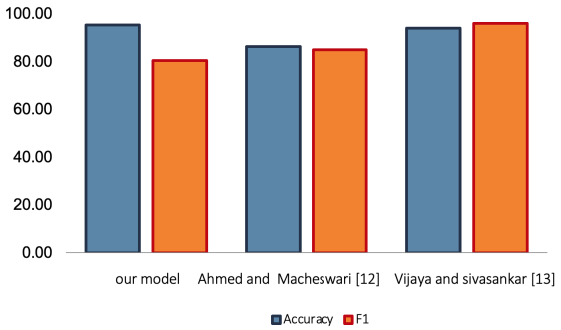


FIGURE 7. Comparisons for Dataset 2.

The best results are also compared with previous works using the same telco operator [13], [14], as shown in Figure 7. It is found that our model shows outstanding accuracy.

E. RESULTS OF DATASET 3

Since there are too many attributes in Dataset 3, chi-square test is first conducted to reduce the dimensionality of the dataset. There is a total of 20 attributes selected for churn prediction. The features with corresponding scores tested by chi-square are shown in Table 13.

1) RESULTS WITHOUT SMOTE

It can be found from Table 14 that Logistic Regression performed the best in accuracy, at 71.05%, while Naïve Bayes obtained the highest recall and F1-score, with 16.16% and 21.39%, respectively. AdaBoost got the highest precision of 57.46%. The ROC curves are shown in Figure 8. It is observed that the curve of AdaBoost is closer to the upper

TABLE 13. Features with Chi^2 score of Dataset 3 for churn prediction.

#	Features	Scores	#	Features	Scores
1	Made Call To Retention Team	223.92	**11**	Handset Models	8.59
2	Retention Calls	62.37	**12**	Off Peak Calls In Out	7.32
3	Current Equipment Days	50.47	**13**	AgeHH1	6.86
4	Handset Refurbished	39.43	**14**	Received Calls	6.42
5	Retention Offers Accepted	23.08	**15**	Homeownership	5.89
6	Credit Rating	22.64	**16**	Handsets	5.26
7	Handset Web Capable	19.28	**17**	Peak Calls In Out	4.76
8	Responds To Mail Offers	16.70	**18**	AgeHH2	4.65
9	Buys Via Mail Order	15.38	**19**	Total Recurring Charge	4.56
10	Monthly Minutes	9.26	**20**	Outbound Calls	4.03

TABLE 14. Results without SMOTE for Dataset 3.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	**71.05**	46.76	2.01	3.75	59.02
Decision Tree	68.36	50.19	5.59	8.75	58.08
Random Forest	68.85	48.99	4.96	8.01	59.27
Naïve Bayes	68.84	42.67	**16.16**	**21.39**	57.13
AdaBoost	70.01	**57.46**	3.36	5.69	**60.31**
Multiple Layer Perceptron	69.56	43.71	4.79	7.80	58.66

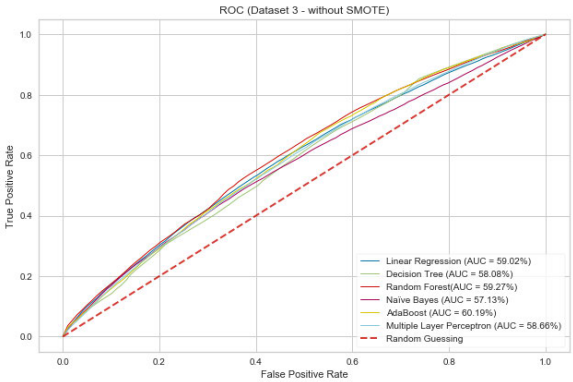


FIGURE 8. ROC curves for Dataset 3 without SMOTE.

left corner, and thus, it performed the best with the highest AUC of 60.31%.

From the results, we can also deduce that, without SMOTE, though Logistic Regression got higher accuracy than other classifiers, the recall and F1-score are much lower. This means Logistic Regression wrongly classify a lot of churn samples to non-churn samples, which may be caused by the insufficient learning of churn characteristics in the training process.

2) RESULTS WITH SMOTE

From Table 15, it is observed that Random Forest got the highest accuracy of 63.09% and AdaBoost obtained the highest precision of 36.53%. Logistic Regression performed the best in recall, standing at 53.67%. Multi-layer Perceptron

TABLE 15. Results with SMOTE for Dataset 3.

Methods	Acc	Pre	Rec	F1	AUC
Logistic Regression	57.60	35.16	**53.67**	41.09	**58.66**
Decision Tree	59.38	34.69	40.84	35.90	56.50
Random Forest	**63.09**	34.67	27.05	29.34	56.97
Naïve Bayes	59.29	34.74	46.25	39.26	56.67
AdaBoost	58.63	**36.53**	49.32	40.53	57.23
Multiple Layer Perceptron	53.47	34.68	61.17	**42.84**	58.39

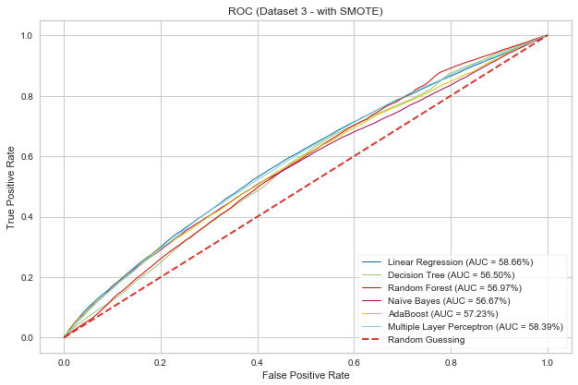


FIGURE 9. ROC curves for Dataset 3 with SMOTE.

achieved the highest F1-score of 42.84%. The ROC curves are shown in Figure 9. It is observed that the curves of Logistic Regression and Multi-layer Perceptron are close, while Logistic Regression got higher AUC of 58.66%.

The results are not as good as the previous 2 datasets due to the nature of dataset. Dataset 3 contains too many attributes which makes the feature selection process much more complicated. The data in Dataset 3 is much noisier than Dataset 1 and Dataset 2, which makes the data pre-processing much more difficult. Different pre-processing combinations will be further explored in the future to obtain better results.

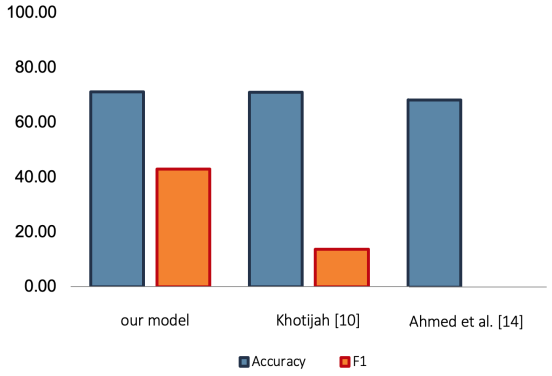


FIGURE 10. Comparisons for Dataset 3.

The best results are also compared with previous works using the same telco operator [10], [15], as shown in Figure 10. It is found that the accuracy of our model is

similar to other studies, but in comparison, our F1-score is much higher. This suggests that our model can figure out more churn customers, which is more important in the context of churn management. Note that F1-score is not provided in [15] and therefore it is not plotted in Figure 10.

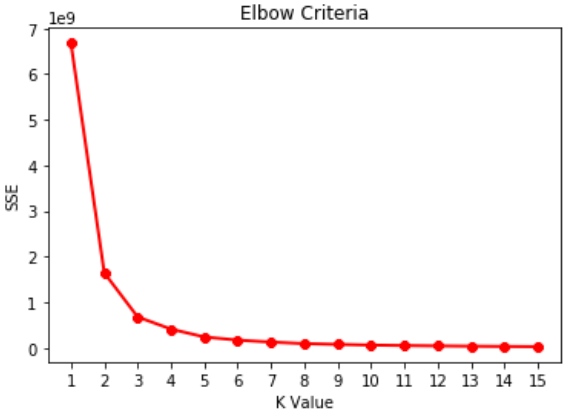


FIGURE 11. Elbow criteria for Dataset 1.

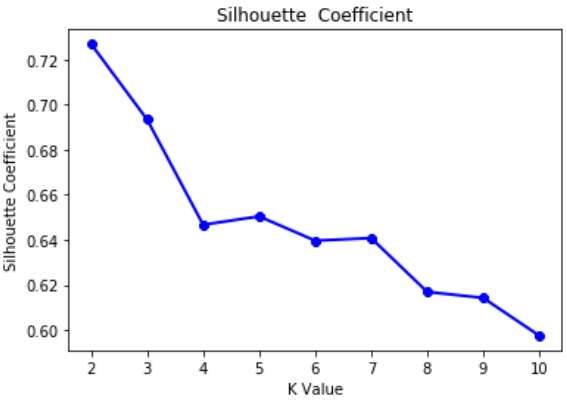


FIGURE 12. Silhouette coefficient for Dataset 1.

VI. BAYESIAN ANALYSIS AND SEGMENTATION

Before segmentation, factor analysis was carried out to better understand the impact of each factor on churning, using Bayesian Logistic Regression. Churn customer segmentation is conducted to further analyse the customers and provide different groups of customers characteristics for telco marketers to propose retention strategies. For the Dataset 1, the optimal value of K is found to be 3 using the elbow method. Figure 11-12 shows the highest silhouette score when $K = 2$. However, the corresponding SSE value when $K = 2$ in elbow method is higher as compared to $K = 3$. When looking at the silhouette coefficient, it is also observed that the silhouette score at $K = 3$ is just slightly smaller than the score at $K = 2$. Therefore, the optimal value of K for Dataset 1 to be selected is 3. Dataset 2 and Dataset 3 have similar patterns as shown in Figure 13-16, and thus, the optimal K for Dataset 2 and Dataset 3 are also considered as 3.

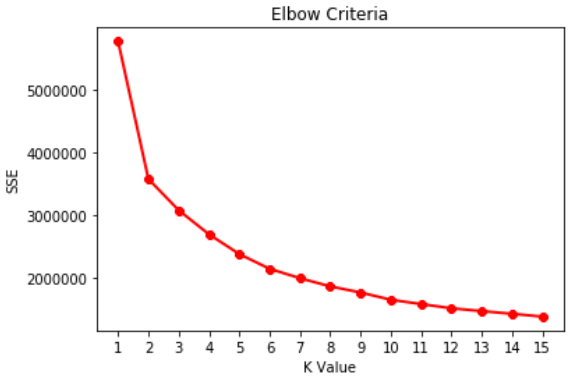


FIGURE 13. Elbow criteria for Dataset 2.

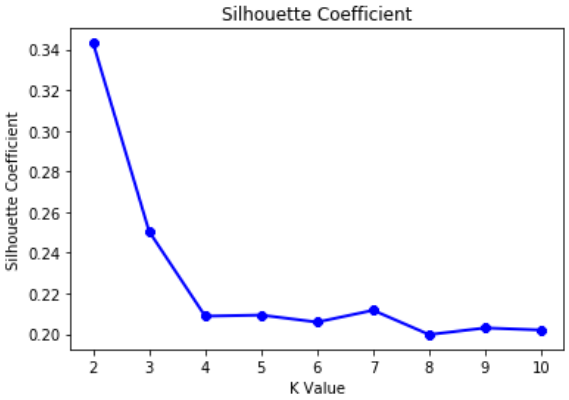


FIGURE 14. Silhouette coefficient for Dataset 2.

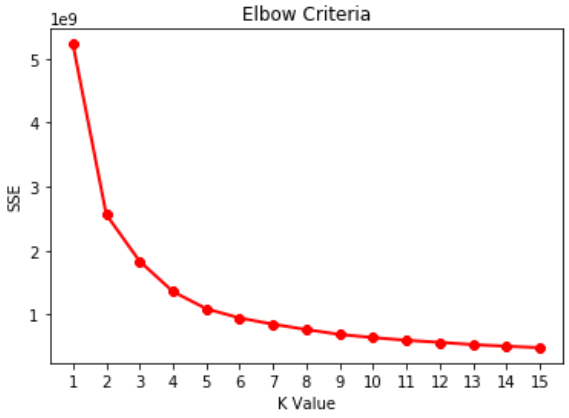


FIGURE 15. Elbow criteria for Dataset 3.

A. RESULTS OF DATASET 1

Based on Bayesian Logistic Regression modelling, how a feature of a person affects his or her turnover can be explored. Different operators may focus on different feature factors. In this research, we show how the contract factor affect one's churn as example. From Figure 17, it can be observed that customers with short period contract and high monthly charges have higher risks to churn. The probability to churn

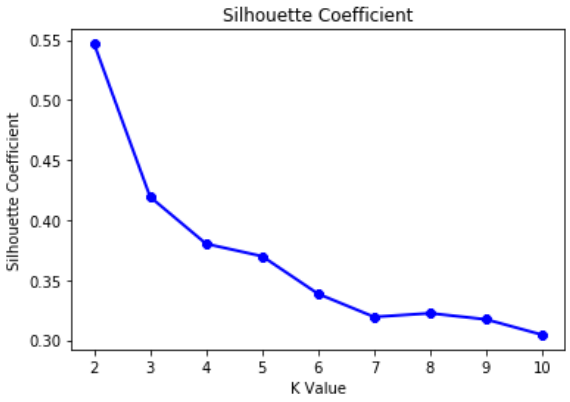


FIGURE 16. Silhouette coefficient for Dataset 3.

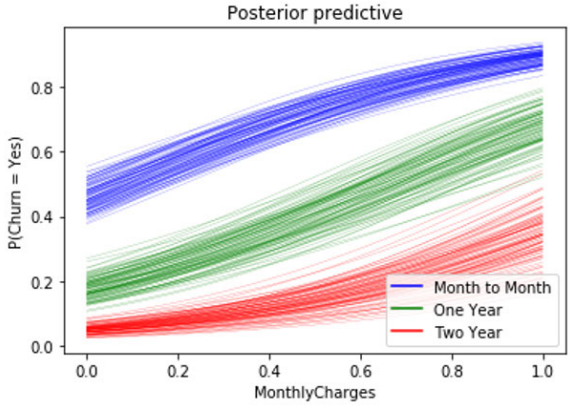


FIGURE 17. Probability of churn (contract and monthly charges) for Dataset 1.

is more than around 80%. This group of customers should be focused more on by operator to see whether they are using packages that are not suitable for them.

Odds ratios are used to measure the relative odds of the occurrence of the outcome, given a factor of interest [39]. In our case, we want to compare the enrichment of different features to churn. Based on the results, the odds ratio is used to determine whether a particular attribute is a risk factor or protective factor for a particular class (churn/non-churn) and the magnitude of percentage effect is used to compare the various risk factors for that class. The positive percentage effect means that the factor is positively correlated with churn and vice versa. If the effect is positive, the greater the factor, the likely that the customer will churn. And these factors are considered as risk factors. While if the effect is negative, the greater the factor, the greater the possibility that the customer will not churn (i.e., the less likely the customer will churn). These factors are considered as protective factors. The odds ratio and percentage effect of each feature are estimated as (11)-(12), where θ is the value of weight of each feature in Logistic Regression model.

$$OddsRatio = e^{\theta} \tag{11}$$

TABLE 16. Odds ratio and percentage effect of each feature for Dataset 1.

Attributes	Weight	OR	Effect %
Gender	-0.1438	0.9734	**-2.66**
Senior Citizen	-0.0270	1.2749	27.49
Partner	0.2429	1.0143	**1.43**
Dependents	0.0142	0.8493	-15.07
Phone Service	-0.1633	0.3568	-64.32
Multiple Lines	-1.0307	1.1574	15.74
Internet Service	0.1462	1.6018	60.18
Online Security	0.4711	0.5805	-41.95
Online Backup	-0.5439	0.7508	-24.92
Device Protection	-0.2866	0.8397	-16.03
Tech Support	-0.1747	0.5905	-40.95
Streaming TV	-0.5268	0.9912	**-0.88**
Streaming Movies	-0.0088	1.0021	**0.21**
Paperless Billing	0.0021	1.4457	44.57
Contract	0.3686	0.2437	-75.63
Payment Method	-1.4117	1.1575	15.75
Tenure	0.1463	0.0147	-98.53
Monthly Charges	-4.2216	10.0543	905.43
Total Charges	2.3080	14.5051	1350.51

$$Effect(\%) = 100 \cdot (OddsRatio - 1) \tag{12}$$

The odds ratio and percentage effect of each feature are calculated as shown in Table 16. Gender, partner, streaming TV, and streaming movies are found to have little impact on churn. Therefore, they are excluded when doing the customer segmentation. It is observed that dependents, phone service, online security, online backup, device protection, technical support, contracts, and tenure are protective factors to churn. It can be seen that for every additional unit, monthly charges and total charges are very risky to churn.

TABLE 17. Features of Dataset 1 for segmentation.

#	Features	#	Features
1	SeniorCitizen	**9**	TechSupport
2	Dependents	**10**	PaperlessBilling
3	PhoneService	**11**	Contract
4	MultipleLines	**12**	PaymentMethod
5	InternetService	**13**	Tenure
6	OnlineSecurity	**14**	MonthlyCharges
7	OnlineBackup	**15**	TotalCharges
8	DeviceProtection		

The features selected for segmentation are shown in Table 17.Then, customers are divided into 3 clusters by K-means clustering. From Table 18, it is found that Cluster 1 has 220 samples, with ratio 11.77%. Cluster 2 has the largest number of samples (1259), with ratio 67.36%. Cluster 3 has 390 samples, with ratio 20.87%.

TABLE 18. Number of samples in each cluster for Dataset 1.

	Cluster 1	Cluster 2	Cluster 3	Total
Dataset 1	220	1259	390	1869

The summary of clustering is shown in Table 19. The character H means that, among 3 clusters, this cluster has most customers have the specific attribute. Similarly, L means this cluster has less customers have the specific attribute.

TABLE 19. Summary of clusters for Dataset 1.

Attributes	C1	C2	C3
Senior Citizen	M	L	H
Dependents	H	L	M
Phone Service	H	L	M
Multiple Line	H	L	M
Internet Service - DSL	L	H	M
Internet Service - FO	H	L	M
Internet Service - No	L	H	L
Additional Services	H	L	M
Paperless Billing	M	L	H
Contract	H	L	M
Payment Methods - Automatic	H	L	M
Payment Methods - Check	L	M	H
Tenure	H	L	M
Bill Amount	H	L	M

M is the middle group. For senior citizen and dependents, H represents that the more customers in this cluster are found senior citizens, or have dependents among 3 clusters. For service-related features, H represents more customers in this cluster have signed up this service among 3 clusters. For paperless billing, payment methods, H means more customers in this cluster prefer these methods among 3 clusters. For contract and tenure, H means longer period. Bill amount includes monthly charges and total charges. And H stands for group with the highest amount.

The characteristics of 3 clusters for Dataset 1 are concluded. Most customers in this Cluster 1 have great demands for services. Among the customers who have subscribed to extended services, this cluster accounts for the largest proportion. For internet service, this cluster prefers fiber optic. Fiber optic is much faster and more expensive. Among churn customers, this cluster has more customers that prefer automatic payment methods. This cluster like a longer contract and they have stayed a long period with this telco company. They also spend more money and have contributed a lot to company revenue, but they still turnover. This cluster should be paid more attention on, since they seem to have high loyalty and contribution to the operator. The reason behind their churn needs to be further figured out.

The customers in Cluster 2 are relatively young and few of them have dependents. They occupy a relatively small proportion for basic services. For internet service, this cluster prefers DSL. DSL is a relatively economical and low-speed package. And the demands for additional services in this cluster are very low. However, compared to other clusters, this cluster has more customers not using paperless billing. They hardly use automatic payment methods. They have been stayed with this operator for a short time, and their contracts are short periods. This cluster has a low contribution to company revenue since the bill amount is low. They may be new customers who just join the operator and tend to choose more economical packages.

More senior citizens are assigned to Cluster 3. This cluster has also large demands for basic services, but the demand for extended services is not very high. Same as customers in Cluster 1, they generally prefer fiber optics for internet services. They like paperless billing very much. Some of them use the automatic payment methods. This cluster signed shorter contracts and has shorter tenure, compared to Cluster 1. They have not contributed a lot to the company yet, since their tenure is not very long. This cluster is considered as a middle cluster between Cluster 1 and Cluster 2, in terms of services demands.

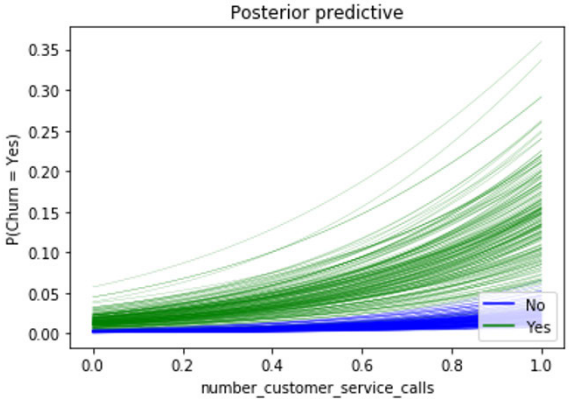


FIGURE 18. Probability of churn (international plan and number of customer service calls) for Dataset 2.

B. RESULTS OF DATASET 2

For Dataset 2, we show how the international plan and customer service calls affect one's churn as example. The probability to churn in customer groups with international plan and customer groups without international plan is illustrated in Figure 18. It is observed that, as customer service calls increase, the churn risk of customers increases. Customers who subscribe to the international plan are more likely to turnover than those who do not. Customers who have made many customer service calls and signed international plans can reach a 20% probability of churn.

TABLE 20. Odds ratio and percentage effect of each feature for Dataset 2.

Attributes	Weight	OR	Effect %
Account Length	0.0320	1.0325	**3.25**
Area Code	-0.0231	0.9772	**-2.28**
International Plan	2.0268	7.5898	658.98
Voice Mail Plan	-2.1640	0.1149	-88.51
Number Vmail Messages	1.4983	4.4741	347.41
Total Day Minutes	2.1919	8.9522	795.22
Total Day Calls	0.0424	1.0433	**4.33**
Total Day Charge	2.1772	8.8216	782.16
Total Eve Minutes	0.9867	2.6824	168.25
Total Eve Calls	-0.5633	0.5693	-43.07
Total Eve Charge	0.7583	2.1346	113.46
Total Night Minutes	0.6560	1.9271	92.71
Total Night Calls	-0.6526	0.5207	-47.93
Total Night Charge	0.4367	1.5476	54.76
Total Intl Minutes	0.4520	1.5715	57.15
Total Intl Calls	-0.9494	0.3870	-61.30
Total Intl Charge	0.5814	1.7885	78.85
Number Customer Service Calls	2.3601	10.5920	959.20

The odds ratio and percentage effect of each feature are shown in Table 20. It is observed that account length, area

code, and total day calls have little impact on churn in this dataset. Therefore, they are excluded when doing the customer segmentation. Total evening calls, total night calls, and international calls are protective factors to churn. It can be seen that for every additional unit, total day minutes, total day charges, and number customer service calls are very risky to churn.

TABLE 21. Features of Dataset 2 for segmentation.

#	Features	#	Features
1	International_plan	**9**	total_night_minutes
2	Voice_mail_plan	**10**	total_night_calls
3	Number_vmail_messages	**11**	total_night_charge
4	total_day_minutes	**12**	total_intl_minutes
5	total_day_charge	**13**	total_intl_calls
6	total_eve_minutes	**14**	total_intl_charge
7	total_eve_calls	**15**	number_customer_service_calls
8	total_eve_charge		

TABLE 22. Number of samples in each cluster for Dataset 2.

	Cluster 1	Cluster 2	Cluster 3	Total
Dataset 2	233	161	150	544

The features selected for segmentation are shown in Table 21. Customers are divided into 3 clusters by K-means clustering. From Table 22, it is observed that Cluster 1 has the largest number of samples (233), with the ratio of 42.83%. Cluster 2 has 161 samples, with the ratio of 29.60%. And Cluster 3 has 150 samples, with the ratio of 27.57%.

TABLE 23. Summary of clustering for Dataset 2.

Attributes	C1	C2	C3
International Plan	H	M	L
Voice Mail Plan	H	M	L
Number Vmail Messages	H	M	L
Total Day Minutes and Charges	L	M	H
Total Eve Minutes and Charges	L	H	M
Total Eve Calls	M	L	H
Total Night Minutes and Charges	L	M	H
Total Night Calls	H	L	M
Total Intl Minutes and Charges	H	L	M
Total Intl Calls	L	M	H
Number Customer Service Calls	H	M	L

The summary of clustering is also shown in Table 23. For service-related features, H represents more customers in this cluster have signed up this service among 3 clusters. For number of messages and calls, H stands for more number. For minutes, H means longer duration. And for charges, H means the customers spend more money.

The characteristics of 3 clusters for Dataset 2 are concluded. Customers in Cluster 1 have higher demands for services. Many customers in this cluster have subscribed to the international plan, but their average international charge is not much different from the other two clusters. The operator needs to investigate whether these customers who have signed up for the plan really need this paid plan. This cluster does not contribute much to the operator's revenue. They call customer service more frequently and may turnover because they are not satisfied with customer service feedback. They may be suitable for being issued some economic packages.

Cluster 2 is a middle cluster between Cluster 1 and Cluster 3, in terms of service demands and charges. The customers of this cluster do not have much demand for services. They also spend the least total international charges. In particular, it can be observed that the customers of this cluster have few evening calls, but there are the longest total call minutes and the highest total call charges. The operator can provide them with special evening packages and give call minutes discounts at this specific time, according to their call duration.

Cluster 3 has little demand for services. They make more domestic calls and international calls, but because of the shorter duration, they do not spend the most money. The characteristics of their phone calls are that calls seem to be frequent and short. Customer service calls are rarely made in this cluster. The operator can give certain discounts based on their call minutes characteristics. For example, the customers can be issued a 24-hour package, so that they can make unlimited calls at a fixed rate within 24 hours.

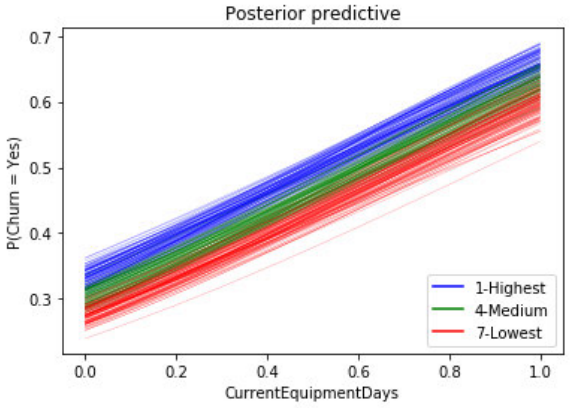


FIGURE 19. Probability of churn (credit rating and current equipment days) for Dataset 3.

C. RESULTS OF DATASET 3

For Dataset 3, we show how the credit rating and the days of current equipment affect one's churn. The probability to churn in customer groups with the highest, medium and lowest credit rating is illustrated in Figure 19. It is observed that as the days of current equipment increases, the probability of churn will increase linearly. It is interesting to find that customers with higher credit rating are more likely to churn, compared to those with lower credit rating. For the highest credit rating group, when the customers have been using the current equipment for a long time, the probability of churn may reach more than 60%, which means they are risky to turnover.

The odds ratio and percentage effect of each feature are shown in Table 24. It is observed that buy via mail order, home ownership, peak calls in out, and age have little impact

TABLE 24. Odds ratio and percentage effect of each feature for Dataset 3.

Attributes	Weight	Odds Ratio	Effect %
Made Call To Retention Team	0.567	1.763	76.262
Retention Calls	1.164	3.204	220.368
Current Equipment Days	1.369	3.931	293.063
Handset Refurbished	0.286	1.331	33.149
Retention Offers Accepted	-0.856	0.425	-57.501
Credit Rating	-0.272	0.762	-23.837
Handset Web Capable	-0.114	0.892	-10.783
Responds To Mail Offers	-0.138	0.871	-12.855
Buys Via Mail Order	0.039	1.040	**3.987**
Monthly Minutes	-0.451	0.637	-36.301
Handset Models	-1.276	0.279	-72.079
Off Peak Calls In Out	-0.317	0.728	-27.160
AgeHH1	-0.475	0.622	-37.812
Received Calls	0.458	1.581	58.107
Homeownership	-0.029	0.972	**-2.820**
Handsets	0.824	2.280	128.006
Peak Calls In Out	-0.052	0.949	**-5.077**
AgeHH2	0.016	1.016	**1.582**
Total Recurring Charge	-1.372	0.254	-74.650
Outbound Calls	0.452	1.571	57.114

on churn in this dataset. Therefore, they are excluded when doing the customer segmentation. Retention offers accepted, credit rating, handset web capable, responds to mail offers, monthly minutes, handset models, off peak calls in out, and total recurring charge are protective factors to churn. On the other hand, it can be seen that for every additional unit, retention calls, and current equipment days are very risky to churn.

TABLE 25. Features of Dataset 3 for segmentation.

#	Features	#	Features
1	Made Call To Retention Team	**9**	Monthly Minutes
2	Retention Calls	**10**	Handset Models
3	Current Equipment Days	**11**	Off Peak Calls In Out
4	Handset Refurbished	**12**	AgeHH1
5	Retention Offers Accepted	**13**	Received Calls
6	Credit Rating	**14**	Handsets
7	Handset Web Capable	**15**	Total Recurring Charge
8	Responds To Mail Offers	**16**	Outbound Calls

TABLE 26. Number of samples in each cluster for Dataset 3.

	Cluster 1	Cluster 2	Cluster 3	Total
Dataset 3	9424	4276	1011	14711

The features selected for segmentation are shown in Table 25. Customers are divided into 3 clusters by K-means clustering. From Table 26, it is observed that Cluster 1 has the largest number of samples (9424), with ratio 64.06%. Cluster 2 has 4276 samples, with ratio 29.07%. Cluster 3 has 1011 samples, with ratio 6.87%.

The summary of clustering is shown in Table 27. For call behaviours, H stands for more number of calls, longer call duration, and higher charges. For boolean attributes such as made call to retention team, handset refurbished, handset web capable, and responds to mail offers, H means more customers in this cluster are true with the specific attribute.

TABLE 27. Summary of clustering for Dataset 3.

Attributes	C1	C2	C3
Made Call To Retention Team	M	L	H
Retention Calls	M	L	H
Current Equipment Days	H	M	L
Handset Refurbished	L	M	H
Retention Offers Accepted	L	M	H
Credit Rating	H	M	L
Handset Web Capable	L	M	H
Responds To Mail Offers	H	M	L
Monthly Minutes	L	M	H
Handset Models	Model 1	Model 1	Model 1
Off Peak Calls In Out	L	M	H
AgeHH1	H	M	L
Received Calls	L	M	H
Handsets	Phone 1	Phone 1	Phone 1
Total Recurring Charge	L	M	H
Outbound Calls	L	M	H

Handsets and handset models are two attributes describing the issued handsets and models of customers, the values show the issued handsets and models of the highest frequency. Particularly, for credit rating, H represents that this cluster has the overall highest credit rating.

The characteristics of 3 clusters for Dataset 3 are concluded. The age of first member of household in Cluster 1 are relatively older than the other 2 groups. This cluster is not very into the retention campaign. They hardly accepted the retention offers, so maybe these offers are not attractive enough. They have been with current equipment for long time. There are more customers' handsets, in this cluster, are not web capable, which means they do not have internet service. They do not call often, and also they spend very little money. They also have overall the highest credit rating among these 3 clusters. They may be customers who do not use current mobile phones frequently, so they need some relatively simple packages with lower charges.

Cluster 2 is a middle group between Cluster 1 and Cluster 3, in terms of call behaviours. Customer in this cluster are much younger than Cluster 1. The overall credit rating of this cluster is High. They do not seem to be very interested in the retention campaign. They have contributed some to operator revenue, however, since their monthly minutes are not as high as Cluster 1, this cluster is not as valuable as Cluster 1. Operators may be able to launch some cost-effective packages to attract and retain the customers of this cluster, in a non-disturbing way such as text messages.

Cluster 3 engages more actively in the retention campaign. Customers in this cluster are younger. They have been with the current equipment for a shorter period and this cluster has more customers refurbished their handsets than the other 2 groups. They have overall lowest credit rating. They need internet service since more customers' handsets are web capable. This cluster also makes calls more frequently and has the longest call duration. They spend much money and have contributed to the operator a lot, so they are considered high-value customers. They use the service of the operator frequently, so they may be dissatisfied with some services and

cause their churn. The operator may need to make some calls back to the customers in this cluster.

VII. GENERAL FINDINGS AND DISCUSSION

For churn prediction, it can be seen from the experiments that the performance of the neural network learning model based on Multi-layer Perceptron is not outstanding, and the computational time of the model is much longer as shown in Figure 20. Neural network models are often effective for processing large volume data, but both Dataset 1 and Dataset 2 in our case have only thousands of records. The advantages of neural network model have not been observed. Thus, it is more practical and time-saving to use conventional machine learning models.

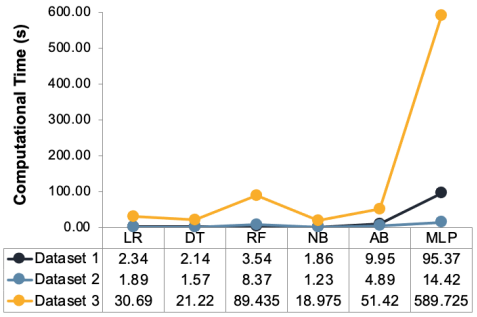


FIGURE 20. Computational time (s) for 3 Datasets.

Besides, It can be found though experiments that SMOTE does not improve the accuracy of prediction, however, this does not mean that SMOTE ineffective. The user preferences need to be considered to select a reasonable metric. The goal of this research is to identify customers with high risk to churn and give the operator more time to respond to it. The classifier with high accuracy may just identify few customers who are about to churn. After applying SMOTE, though the accuracy might drop, the models have successfully identified more customers who are about to leave, which is the goal of churn management. If a classifier did not master the characteristics of churn customers well during the learning process, then it cannot identify the churn customers correctly. Therefore, it is necessary to implement re-sampling methods to the imbalanced training set. SMOTE greatly improves the value of Recall, which means the models with SMOTE can identify more customers that are about to churn.

Regarding to the metric selection, particularly, in the confusion matrix, FN should be paid more attention since it reflects how many churn customers are failed to be identified. Precision, Recall, F1-Score are more suitable to evaluate our models. Precision represents how many of the customers that are predicted to churn are the ones that will really churn. Recall measures how many customers are identified from the customer base that actually churn. Recall is a relatively important metric, and operators should minimise the risk of missed determination. In other words, operators would rather adopt a retention strategy for customers who are not actually churn, rather than neglect customers who are about to churn. However, if the Recall is increased to 100%, there is a possibility that the classifier will classify all customers as churn customers, and the operator will apply retention strategies to all customers. In this case, not any churn customers will be missed, but the cost of management for the operator will increase greatly, making churn management ineffective. There is always a trade-off. F1-score, as a comprehensive metric, can be used to evaluate the performance of the classifier.

To prevent the bias of comparing the different classifiers before and after SMOTE, the statistical test is done. T-test is done for each classifier, in terms of F1-score. The null hypothesis (H_0) is as (13), and the alternative hypothesis (H_1) as (14), where *perf* is the performance of a classifier.

$$H_0 : perf(ClassifierA) = perf(ClassifierB) \quad (13)$$

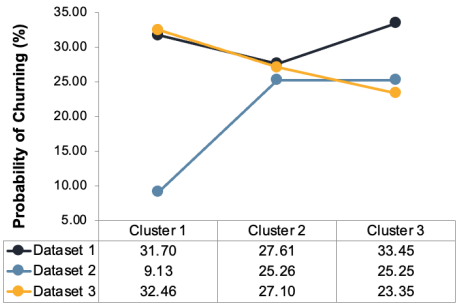
$$H_1 : perf(ClassifierA) \neq perf(ClassifierB) \quad (14)$$

If the p-value is less than 0.05, then the null hypothesis can be rejected, at 95% confidence level, indicating there is a significant difference of the performance between two classifiers. The results are shown in Table 28. It can be found that most of the classifiers result in low p-values, suggesting there are significant differences. For Dataset 1, Random Forest got the lowest p-value, at 7.55E-06. For Dataset 2, Logistic Regression obtained the lowest p-value of 5.95E-10. And for Dataset 3, AdaBoost got the lowest p-value, standing at 7.36E-14. For the average p-value, Multi-layer Perceptron is the lowest, indicating the most obvious difference in performance with and without SMOTE.

TABLE 28. P-values obtained from T-test.

	Dataset 1	Dataset 2	Dataset 3	Avg p-value
LR	0.01	**5.95E-10**	5.27E-13	3.47E-03
DT	1.45E-05	0.19	5.43E-12	0.06
RF	**7.55E-06**	0.12	2.34E-11	0.04
NB	0.43	1.11E-03	7.20E-05	0.14
AB	2.34E-03	1.56E-07	**7.36E-14**	7.81E-04
MLP	1.09E-03	5.89E-06	7.97E-14	**3.64E-04**

Since the operators may need more information on the overall probability of churning for the segmented clusters, the overall probability of churning for 3 datasets are calculated based on the Bayesian Logistic Regression model, as shown in Figure 21. For Dataset 1, among all churn customers, Cluster 3 obtained the highest overall probability of churning at 33.45%, indicating that customers who have similar characteristics with Cluster 3 will be more likely to churn. As mentioned in Section VI-A, Cluster 3 in Dataset 1 has large demand for basic services and has contributed to operators. The reason for their churn is probably because they did not receive satisfactory service. For Dataset 2, Cluster 2 has the highest overall probability of churning, standing at 25.26%. Cluster 2 may require a more suitable package due to their calls behaviour. For Dataset 3, Cluster 1 is more likely to churn in general, with 32.46%. Customers in Cluster 1 are relatively inactive users and use very less operator



	Cluster 1	Cluster 2	Cluster 3
Dataset 1	31.70	27.61	33.45
Dataset 2	9.13	25.26	25.25
Dataset 3	32.46	27.10	23.35

FIGURE 21. Overall probability of churning for 3 Datasets (%).

services. They have the highest credit rating but still churn. The operator needs to further explore the reasons behind the churning for different clusters with different customer behaviour. In this way, the operator can gain experience by analysing the customers who have already churn, so as to prevent more customers from churning in the future.

VIII. CONCLUSION

Since customers in the telco market always tend to be saturated, it is more beneficial for operators to propose retention strategies for customers who are about to leave. In this research, an integrated customer analytics framework is proposed. For Dataset 1, our model achieves the highest F1-score at 63.11% and the highest AUC value at 84.52%, using AdaBoost. For Dataset 2, our model achieves the highest F1-score at 77.20% and the highest AUC value at 91.40%, using Random Forest. For Dataset 3, our model achieves the highest F1-score at 42.84% and the highest AUC value at 58.66%, using Multi-layer Perceptron and Logistic Regression, respectively. Additionally, the importance of using SMOTE to deal with imbalanced datasets and better model evaluation metrics are also discussed.

Through the results of Bayesian Analysis, some features that are not important to churn are dropped for churn customer to achieve precise segmentation. Besides, the Elbow Criterion method and the Silhouette Coefficient method are used to determine the best K. The churn customers in Dataset 1, Dataset 2 and Dataset 3 are segmented into three groups, using K-means clustering. Furthermore, each segment is summarised and provided a cluster description for the marketers and strategists to better understand different groups of customers. The overall probabilities of churning are calculated for each cluster of the 3 datasets. Cluster 3 in Dataset 1 has the highest overall probability of churning, with 33.45%. Customers in Cluster 2 of Dataset 2 are more likely to churn, standing at 25.26%. Cluster 1 in Dataset 3 has higher overall probability of churning, at 32.46%.

This research contributes to the existing literature from three aspects. First, only a few of the existing literature considers both churn prediction and customer segmentation in the telco industry. This research fills this gap by proposing an integrated customer analytics framework to seamlessly connect these two components. Second, only limited research involves Bayesian Analysis. This research adopts it for conducting factor analysis, enabling it acts as an intermediary linking churn prediction to customer segmentation. Third, this research provides operators with the overall probability of churning of each cluster, allowing them to better understand the churning situation of each cluster.

In the future, other oversampling and undersampling methods to deal with the imbalanced datasets can be further explored. In particular, some oversampling methods including the variants of SMOTE will be studied. Besides, in order to improve the accuracy of prediction, ROC analysis can be further investigated to select a more reasonable threshold for churn prediction. The hyper-parameter tuning using Bayesian optimisation will also be studied to make the models with SMOTE more accurate.

REFERENCES

[1] G. Esteves and J. Mendes-Moreira, "Churn perdiction in the telecom business," in *Proc. 11th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Sep. 2016, pp. 254–259.

[2] C.-F. Tsai and Y.-H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12547–12553, Dec. 2009.

[3] G. F. Retana, C. Forman, and D. J. Wu, "Proactive customer education, customer retention, and demand for technology support: Evidence from a field experiment," *Manuf. Service Oper. Manage.*, vol. 18, no. 1, pp. 34–50, Feb. 2016.

[4] J. Bayer, "Customer segmentation in the telecommunications industry," *J. Database Marketing Customer Strategy Manage.*, vol. 17, nos. 3–4, pp. 247–256, Sep. 2010.

[5] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1414–1425, Jan. 2012.

[6] S. Induja and D. Eswaramurthy, "Customers churn prediction and attribute selection in telecom industry using kernelized extreme learning machine and bat algorithms," *Int. J. Sci. Res.*, vol. 5, no. 12, pp. 258–265, Dec. 2016.

[7] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *Eur. J. Oper. Res.*, vol. 218, no. 1, pp. 211–229, Apr. 2012.

[8] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," in *Proc. 4th Int. Conf. Rel., Infocom Technol. Optim. (ICRITO)*, Sep. 2015, pp. 1–6.

[9] J. Pamina, B. Raja, S. SathyaBama, S. Soundarya, M. S. Sruthi, S. Kiruthika, V. J. Aiswaryadevi, and G. Priyanka, "An effective classifier for predicting churn in telecommunication," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 10, pp. 221–229, Jun. 2019.

[10] S. Khotijah. (2020). *Churn Prediction*. [Online]. Available: https://www.kaggle.com/khotijahs1/churn-prediction

[11] D. D. Adhikary and D. Gupta, "Applying over 100 classifiers for churn prediction in telecom companies," *Multimedia Tools Appl.*, vol. 248, pp. 1–22, Aug. 2020.

[12] A. Idris, M. Rizwan, and A. Khan, "Churn prediction in telecom using random forest and PSO based data balancing in combination with various feature selection strategies," *Comput. Electr. Eng.*, vol. 38, no. 6, pp. 1808–1819, Nov. 2012.

[13] A. A. Q. Ahmed and D. Maheswari, "Churn prediction on huge telecom data using hybrid firefly based classification," *Egyptian Informat. J.*, vol. 18, no. 3, pp. 215–220, Nov. 2017.

[14] J. Vijaya and E. Sivasankar, "An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing," *Cluster Comput.*, vol. 22, no. S5, pp. 10757–10768, Sep. 2019.

[15] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer learning and meta classification based deep churn prediction system for telecom industry," 2019, *arXiv:1901.06091*. [Online]. Available: https://arxiv.org/abs/1901.06091

[16] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, and K. Huang, "Customer churn prediction in the telecommunication sector using a rough set approach," *Neurocomputing*, vol. 237, pp. 242–254, May 2017.
[17] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *J. Bus. Res.*, vol. 94, pp. 290–301, Jan. 2019.
[18] A. Amin, B. Shah, A. M. Khattak, F. J. L. Moreira, G. Ali, A. Rocha, and S. Anwar, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, Jun. 2019.
[19] A. Amin, F. Al-Obeidat, B. Shah, M. A. Tae, C. Khan, H. U. R. Durrani, and S. Anwar, "Just-in-time customer churn prediction in the telecommunication sector," *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020.
[20] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940–7957, 2016.
[21] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 3964–3973, Mar. 2012.
[22] L. Ye, C. Qiuru, X. Haixu, L. Yijun, and Z. Guangping, "Customer segmentation for telecom with the k-means clustering method," *Inf. Technol. J.*, vol. 12, no. 3, p. 409, 2013.
[23] A. Namvar, M. Ghazanfari, and M. Naderpour, "A customer segmentation framework for targeted marketing in telecommunication," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nov. 2017, pp. 1–6.
[24] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," *Expert Syst. Appl.*, vol. 31, no. 3, pp. 515–524, Oct. 2006.
[25] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019.
[26] T. J. Gerpott, W. Rams, and A. Schindler, "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market," *Telecommun. Policy*, vol. 25, no. 4, pp. 249–269, May 2001.
[27] G. Olle, "A hybrid churn prediction model in mobile telecommunication industry," *Int. J. e-Educ., e-Bus., e-Manage. e-Learn.*, vol. 4, no. 1, p. 55, Feb. 2014.
[28] P. Li, T. Bi, Y. Liu, and S. Li, "Telecom customer churn prediction method based on cluster stratified sampling logistic regression," in *Proc. Int. Conf. Softw. Intell. Technol. Appl. Int. Conf. Frontiers Internet Things*, Dec. 2014, pp. 282–287.
[29] A. Khalatyan, "Churn management in telecommunications: Challenging the innovative capability of data mining tools," *iSCHANNEL*, vol. 5, no. 1, pp. 21–26, Sep. 2010.
[30] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer churn prediction modelling based on behavioural patterns analysis using deep learning," in *Proc. Int. Conf. Smart Comput. Electron. Enterprise (ICSCEE)*, Jul. 2018, pp. 1–6.
[31] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 1985.
[32] R. Bååth. (2017). *Introduction to Bayesian Data Analysis—Part 1: What is Bayes?*. [Online]. Available: https://www.youtube.com/watch?v=3OJEae7Qb_o
[33] R. Bååth. (2017). *Introduction to Bayesian Data Analysis—Part 2: Why use Bayes?* [Online]. Available: https://www.youtube.com/watch?v=mAUwjSo5TJE
[34] R. Bååth. (2017). *Introduction to Bayesian Data Analysis—Part 3: How to do Bayes?* [Online]. Available: https://www.youtube.com/watch?v=Ie-6H_r7I5A
[35] IBM. (2018). *Telco Customer Churn*. [Online]. Available: https://www.kaggle.com/blastchar/telco-customer-churn
[36] Kaggle. (2018). *Customer Churn Prediction 2020*. [Online]. Available: https://www.kaggle.com/c/customer-churn-prediction-2020/data
[37] Cell2Cell. (2018). *Telecom Churn (Cell2Cell)*. [Online]. Available: https://www.kaggle.com/jpacse/datasets-for-churn-telecom
[38] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–50, Nov. 2016.
[39] J. M. Bland and D. G. Altman, "The odds ratio," *Brit. Med. J.*, vol. 320, no. 7247, p. 1468, 2000.

SHULI WU received the bachelor's degree in computer science and technology from Xiamen University Malaysia. She is currently pursuing the master's degree in business analytics with The University of Edinburgh, U.K. Her research interests include data analytics and machine learning.

WEI-CHUEN YAU (Member, IEEE) received the B.S. and M.S. degrees from National Cheng Kung University, Taiwan, and the Ph.D. degree from Multimedia University. He is currently an Associate Professor with the School of Electrical and Computer Engineering, Xiamen University Malaysia. He is also a Chartered Engineer (CEng) and a Certified Information Systems Security Professional (CISSP). His research interests include cryptography, security protocols, machine learning, and network security. He was the General Co-Chair of Mycrypt 2016. He has also served as a Guest Editor for the *ETRI Journal* Special Issue on Cyber Security and AI.

THIAN-SONG ONG (Senior Member, IEEE) received the M.Sc. degree from The University of Sunderland, U.K., in 2001, and the Ph.D. degree from Multimedia University, Malaysia, in 2007. He is currently an Associate Professor with the Faculty of Information Science and Technology, Multimedia University. He has published more than 60 international refereed journals and conference papers. His research interests include data analytics, machine learning, and biometric security. He was the General Chair of ICoICT2017 and ICoICT2019. He has served as an Editorial Board Member for IEEE Biometric Council Newsletter, from 2013 to 2015.

SIEW-CHIN CHONG (Senior Member, IEEE) received the B.IT. degree in software engineering, the M.Sc. degree in information technology, and the Ph.D. degree in IT from Multimedia University, Malaysia, in 2003, 2006, and 2018, respectively. She is currently serving as the Deputy Dean of staff and student affairs with the Faculty of Information Science and Technology, Multimedia University. Her research interests include machine learning, biometrics security, and mobile app development. She has served as an Editorial Board Member for *Progress in Human Computer Interaction*, from 2018 to 2020, and a committee member for several international conferences. She has passed the examinations of the IBM Certified Academic Associate (DB2 9 Database and Application Fundamentals) and the Certificate of Cloud Security Knowledge V4.