PAPER • **OPEN ACCESS**

Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes

To cite this article: Y Yulianti and A Saifudin 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **879** 012090

IOP Conf. Series: Materials Science and Engineering **879** (2020) 012090 doi:10.1088/1757-899X/879/1/012090

Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes

Y Yulianti[1*], A Saifudin[2]
[1,2]Informatics Engineering, Pamulang University, Jalan Raya Puspitek 46, Banten 15310, Indonesia

Email : [*]aries.saifudin@unpam.ac.id

Abstract. The customer churn prediction is an important business strategy for the company. The ease of switching operators is one of the serious challenges that must be faced by the telecommunications industry. To get new customers requires a much higher cost than maintaining existing customers. Customer churn refers to the periodic loss of customers in an organization. To retain existing customers, organizations must improve customer service, improve product quality, and must be able to know in advance which customers have the possibility of leaving the organization. By predicting customer churn, companies can immediately take action to retain customers. Prediction can be done by analysing customer data using data mining techniques. This study proposes to implement feature selection to select relevant features and can provide improved performance in customer churn prediction models. Some proposed feature selection techniques are Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Forward Floating Selection (SFFS), Sequential Forward Floating Selection (SBFS), Sequential Backward Floating Selection (SBFS). The classification algorithm used to classify is Naive Bayes. The model that provides the best performance value is the model that implements Sequential Backward Selection (SBS) and Sequential Backward Floating Selection (SBFS) feature selection technique with feature number 19.

1. Introduction

Research on customer churn is increasingly important and attracts much attention of researchers[1]. Customers can easily exercise their rights to switch service providers from one operator to another. The large number of cellular operators encourages fierce business competition. The ease of switching operators is one of the serious challenges that must be faced by the telecommunications industry[2]. Given the fact that the telecommunications industry experiences an average annual churn rate of 30-35 percent, and the cost of recruiting new customers is 5-10 times more expensive than retaining existing ones, retaining customers becomes more important than acquiring customers[3]. Customer churn refers to periodic customer loss in an organization[4,5]. To retain existing customers, organizations must improve customer service, improve product quality, and must be able to know in advance which customers have the possibility of leaving the organization.

Telecommunication industries are trying to develop ways to predict customers who have the potential to churn so that action can be taken to prevent it because of the effect of direct churn on the decline in company revenue[6]. Churn activity is very influential on total profits and business image, so it is better if it can be predicted and prevented[7]. Churn prediction can be used to identify churners early before they move, and can help the CRM (Customer Relationship Management) department to retain them, so that the potential loss of the company can be avoided[8]. Thereby the service providers must optimise the performance of churn prediction model and employ churn prediction techniques and apply appropriate marketing strategies to retain the existing subscriber[9]. The customer churn prediction is

IOP Conf. Series: Materials Science and Engineering **879** (2020) 012090 doi:10.1088/1757-899X/879/1/012090

an important business strategy for the company. By predicting customer churn, companies can immediately take action to retain customers.

To make predictions using data mining techniques required past data that has been collected. Consumer data is widely available in company databases, how to use it to predict customer churn is a challenge for researchers[10]. Many classification algorithms have been applied, such as rule based classification (Ripper, PART), decision tree approach (C4.5, CART, Alternating Decision Trees), neural networks (Multilayer Perceptron, Radial Basis Function Network), k-Nearest Neighbour (k-NN), ensemble methods (Random Forests, Logistic Model Trees, Bagging, Boosting), and classical statistical methods (logistic regression, Naïve Bayes, Bayesian Networks). Classical statistical methods such as logistic regression and Naïve Bayes provide good grades, strong results, and are easy to apply[11], but have not yet achieved excellent values (excellent).

In a study conducted by Keramati[12] stated that to stay in the telecommunications business must be able to distinguish between customers who have the possibility to move to competitors, and customers who are reluctant to move. Therefore, the prediction of customer churn has become an important issue in the telecommunications business. In a competitive business, reliable predictors of customers are considered invaluable. In this study applying data mining classification techniques including Decision Tree (DT), Artificial Neural Networks (ANN), k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM), then compare its performance. The dataset used was obtained from telecommunications companies in Iran. In addition, this study also proposes a hybrid method that is intended to make improvements to the value of evaluation measures. The results showed that the value of precision and recall can achieve very good values.

In research conducted by Rodan, Fayyoumi, Faris, Alsakran, and Al-Kadi[13] stated that currently telecommunications companies have paid much attention to the problem of identifying behaviors customer churn. In business, it is known that attracting new customers is far more expensive than retaining existing ones. Therefore, implementing an accurate customer churn prediction model can effectively help in retaining customers and increasing profits. In this research, the Mutilayer Perceptrons (MLP) ensemble is used with training obtained using Negative Correlation Learning (NCL) to predict customer churn in telecommunications companies. The results of the study confirm that NCL-based MLP ensemble can achieve better general performance compared to MLP ensemble without NCL (flat ensemble) and general data mining techniques used to analyse customer churn.

The dataset used for customer churn predictions has many features, but those features have the possibility of redundancy or are irrelevant so that it can cause a decrease in classifier performance[14]. This study proposes to implement feature selection to select relevant features and can provide performance improvements to customer churn prediction models. Some proposed feature selection techniques are Sequential Forward Selection (SFS), Sequential Backward Selection (SBS)[15], Sequential Forward Floating Selection (SFFS), Sequential Forward Floating Selection (SBFS), Sequential Backward Floating Selection (SBFS)[16]. The classification algorithm used to classify is Naive Bayes.

2. Method

This research is an experimental study conducted by proposing a customer churn prediction model, then applying it to the telecommunications customer churn dataset. Experiments carried out by developing applications to implement the proposed model using the Python programming language. Every model that has been applied is measured and compared to find the best model.

In this experiment use secondary data, namely datasets that had been collected by other researchers. The dataset which used for experiment is Telco Customer Churn dataset that obtained from https://www.kaggle.com/blastchar/telco-customer-churn. The dataset contain 20 features, one label and 7043 records. The dataset contains features with categorical and numeric types. The proposed model can only process numeric data, so preprocessing needs to be done to convert all categorical features into numeric types.

IOP Conf. Series: Materials Science and Engineering **879** (2020) 012090 doi:10.1088/1757-899X/879/1/012090

This research proposes a model that applies feature selection to select relevant features and can improve the performance of customer churn prediction models. The proposed churn prediction model frame work is shown in Figure 1. The customer churn dataset used contains categorical and numerical features. In order to be processed by the proposed model, all categorical features are converted to numeric types. Furthermore, standardization is performed using the scalar min-max algorithm. Then only selected using the proposed feature selection algorithm. The dataset of feature selection results is used to train and test the proposed model by applying validation using 10-fold cross validation.

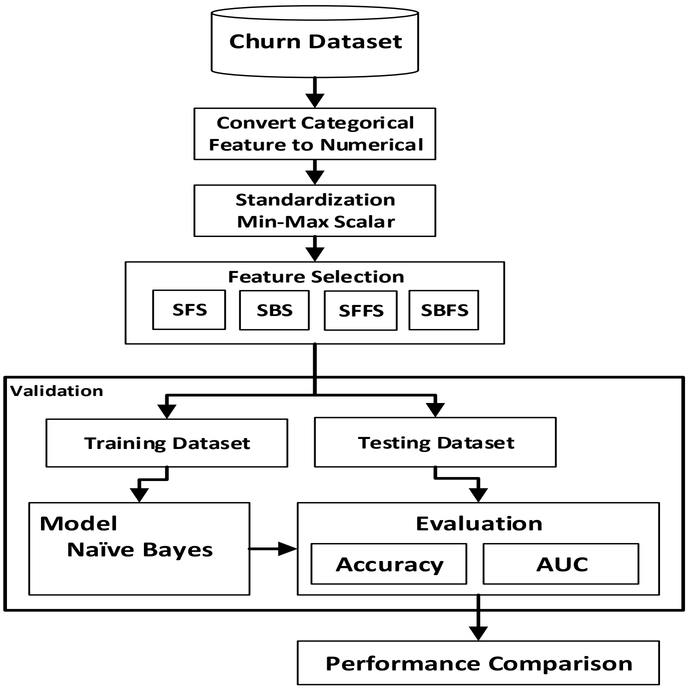

Figure 1. Churn Prediction Model Framework

The test results are entered in the confusion matrix table and calculate the performance of classifiers is carried out in the form of accuracy and AUC (Area Under the Curve). The confusion matrix is a very useful tool for analysing the performance of classifying models and being able to recognize tuples and features from different classes [17]. Analysis using a confusion matrix is done by counting the number of objects that are predicted correctly and incorrectly to determine the performance of the model [18]. Table 2 is a table of confusion matrix.

Table 1 Table of Confusion Matrix

Class		**Actual**	
		True	**False**
Prediction	**True**	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	TN (True Negative)

Validation values that have been entered into the confusion matrix are used to calculate the Accuracy or AUC value of each model and measure the performance of the model. To calculate the Accuracy and AUC values, the following equation can be used [17]:

$$\boldsymbol{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$TP_{rate} = \frac{TP}{TP+FN} \tag{2}$$

$$FP_{rate} = \frac{FP}{FP+TN} \tag{3}$$

$$AUC = \frac{1+TP_{rate}-FP_{rate}}{2} \tag{4}$$

Based on the proposed model, there will be 5 models, namely NB, SFS, SBS, SFFS, and SBFS. The performance of the five models is compared to get the best model. SFS is a deterministic feature selection method that uses hill-climbing search to add and assess all possible single attribute expansions to the present subset[19]. While SBS works in the opposite direction to SFS[20]. SFS and SBS select features in one-way, so the features that have been evaluating cannot be selected again, but these weaknesses avoided in SFFS and SBFS[21].

Naïve Bayes algorithm is a machine learning method that analyzes data based on calculating probabilities. When used to analyze large-sized datasets, Naïve Bayes can provide high accuracy and speed. Naïve Bayes can also be used even if it doesn't have enough datasets with accurate results[22]. The Bayes theorem is the basis of the Naïve Bayes equation, which is:

$$P(C|x) = \frac{P(C)P(x|C)}{P(x)} \tag{5}$$

In the above equation, C is a class whose probability value will be calculated because it is influenced by the value of feature x. When feature x has a continuous value, it is assumed to have a Gaussian distribution with averages (μ) and standard deviations (σ) [23]. So the equation becomes:

$$P(x|C) = \frac{1}{\sqrt{2\pi}\cdot\sigma_C}\, e^{-\frac{(x-\mu)^2}{2(\sigma_C)^2}} \tag{6}$$

3. Results and Discussion

Based on the model proposed in Figure 1, to find out the performance of the basic model applied by the Naïve Bayes algorithm as a classification without being optimized. The second model is the integration of Naïve Bayes with Sequential Forward Selection (SFS). The third model integrates Naïve Bayes and Sequential Backward Selection (SBS). The fourth model integrates Naïve Bayes and Sequential Forward Floating Selection (SFFS). The fifth model integrates Naïve Bayes and Sequential Backward Floating Selection (SBFS).

Table 2 Model Performance

Feature Number	Naïve Bayes		SFS		SBS		SFFS		SBFS	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
1	65.98%	0.729	76.09%	0.690	76.09%	0.690	76.09%	0.690	76.09%	0.690
2	65.98%	0.729	**77.74%**	0.696	**77.74%**	0.696	**77.74%**	0.696	**77.74%**	0.696
3	65.98%	0.729	66.85%	0.737	66.85%	0.737	66.85%	0.737	66.85%	0.737
4	65.98%	0.729	68.48%	0.745	68.48%	0.745	68.48%	0.745	68.48%	0.745
5	65.98%	0.729	68.38%	0.744	68.38%	0.744	68.38%	0.744	68.38%	0.744
6	65.98%	0.729	68.68%	0.748	68.68%	0.748	68.68%	0.748	68.68%	0.748
7	65.98%	0.729	68.93%	0.750	68.93%	0.750	68.93%	0.750	68.93%	0.750

Feature Number	Naïve Bayes		SFS		SBS		SFFS		SBFS	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
8	65.98%	0.729	68.99%	0.751	68.99%	0.751	68.99%	0.751	68.99%	0.751
9	65.98%	0.729	68.91%	0.749	69.50%	0.746	68.91%	0.749	69.50%	0.746
10	65.98%	0.729	69.71%	0.747	68.12%	0.745	69.71%	0.747	68.12%	0.745
11	65.98%	0.729	68.05%	0.744	68.44%	0.745	68.44%	0.745	68.44%	0.745
12	65.98%	0.729	68.51%	0.745	69.27%	0.747	69.27%	0.747	69.27%	0.747
13	65.98%	0.729	69.22%	0.747	70.38%	0.749	70.38%	0.749	70.38%	0.749
14	65.98%	0.729	70.31%	0.749	70.57%	0.750	70.57%	0.750	70.57%	0.750
15	65.98%	0.729	70.48%	0.750	70.48%	0.750	70.48%	0.750	70.48%	0.750
16	65.98%	0.729	70.01%	0.748	69.71%	0.748	70.48%	0.750	69.71%	0.748
17	65.98%	0.729	70.57%	0.750	70.07%	0.748	70.04%	0.747	70.07%	0.748
18	65.98%	0.729	70.40%	0.749	70.54%	0.751	70.44%	0.750	70.54%	0.751
19	65.98%	0.729	69.96%	0.745	**71.33%**	**0.753**	70.99%	0.752	**71.33%**	**0.753**
20	65.98%	0.729	70.30%	0.748	70.75%	0.750	70.48%	0.748	70.75%	0.750
21	65.98%	0.729	70.03%	0.746	70.52%	0.748	70.27%	0.748	70.52%	0.748
22	65.98%	0.729	69.81%	0.746	70.23%	0.747	69.90%	0.746	70.23%	0.747
23	65.98%	0.729	69.59%	0.744	69.59%	0.744	69.59%	0.744	69.59%	0.744
24	65.98%	0.729	68.98%	0.742	68.98%	0.742	68.98%	0.742	68.98%	0.742
25	65.98%	0.729	68.21%	0.739	68.21%	0.739	68.21%	0.739	68.21%	0.739
26	65.98%	0.729	67.58%	0.736	67.58%	0.736	67.58%	0.736	67.58%	0.736
27	65.98%	0.729	67.06%	0.734	67.06%	0.734	67.06%	0.734	67.06%	0.734
28	65.98%	0.729	66.55%	0.732	66.55%	0.732	66.55%	0.732	66.55%	0.732
29	65.98%	0.729	65.98%	0.729	65.98%	0.729	65.98%	0.729	65.98%	0.729
Average	**65.98%**	**0.729**	**69.46%**	**0.741**	**69.59%**	**0.741**	**69.60%**	**0.741**	**69.59%**	**0.741**

Base on Table 3 of model performance, accuracy can improve using 2 selected features using all four feature selection algorithm. Then the performance model visualize using graph in Figure 2 and Figure 3.

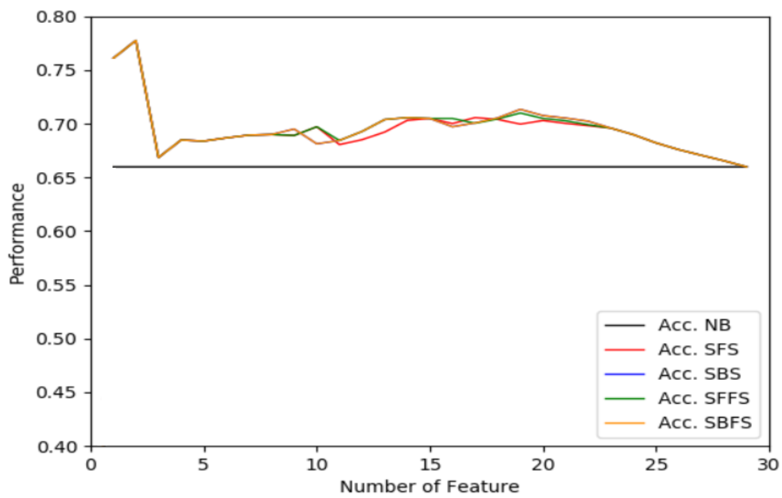

Figure 2. Graph of Accuracy Models

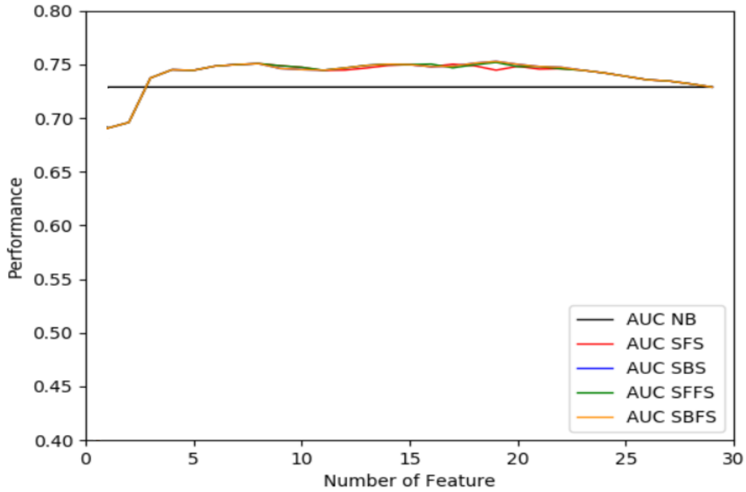

Figure 3. Graph of AUC Models

Based on the performance graph of the proposed model, it shows that models that implement feature selection can provide better results. From the experiments show that the model that can provide the highest results is a model with SBS and SBFS feature selection.

The AUC and accuracy performance has high performance when select 19 features using SBS and SBFS. The model which implement feature selection SBS and SBFS select the same feature, namely SeniorCitizen, tenure, MonthlyCharges, gender_Male, Partner_Yes, PhoneService_Yes, MultipleLines_No phone service, MultipleLines_Yes, InternetService_Fiber optic, OnlineSecurity_Yes, OnlineBackup_Yes, DeviceProtection_Yes, TechSupport_Yes, StreamingTV_Yes, StreamingMovies_Yes, Contract_One year, Contract_Two year, PaperlessBilling_Yes, PaymentMethod_Electronic check. So, 10 features that degrade performance are Dependents_Yes, InternetService_No, OnlineSecurity_No internet service, OnlineBackup_No internet service, DeviceProtection_No internet service, TechSupport_No internet service, StreamingTV_No internet service, StreamingMovies_No internet service, PaymentMethod_Credit card (automatic), PaymentMethod_Mailed check.

These results indicate that feature selection can select relevant features to train customer churn prediction models and produce better performance. Its confirms that the feature selection should use to choose the relevant features to train the prediction model[24]. Moreover, it also proves that not all features in the dataset are relevant for customer churn predictions. The performance result proves that to build a high-quality customer churn prediction model need to implement feature selection[13].

4. Conclusion

The cost of getting new customers is much higher than keeping existing customers. The customer churn prediction model is an important research topic because it can be used to reduce operational costs in the telecommunications industry. In this study, it has been proposed the application of feature selection to select features that are relevant and have a positive effect on prediction models. Based on research conducted, models that implement feature selection can provide better performance. The proposed model can detect customers who will unsubscribing is better, so action can be taken to prevent it.

References

[1] Lee E B, Kim J and Lee S G 2017 Predicting Customer Churn in Mobile Industry Using Data Mining Technology *Ind. Manag. Data Syst.* **117**, pp. 90–109

[2] Huang B, Kechadi M T and Buckley B 2012 Expert Systems with Applications Customer churn prediction in telecommunications *Expert Syst. Appl.* **39**, pp.1414–25

[3] Lu J Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS â
[4] Churi A, Divekar M, Dashpute S and Kamble P 2015 Analysis of Customer Churn in Mobile Industry using Data Mining **5** 1–6
[5] Yu X, Guo S, Guo J and Huang X 2011 Expert Systems with Applications An extended support vector machine forecasting framework for customer churn in e-commerce *Expert Syst. Appl.* **38** pp.1425–30
[6] Ahmad A K, Jafar A and Aljoumaa K 2019 Customer churn prediction in telecom using machine learning in big data platform *J. Big Data* **6**
[7] Pamina J, Dhiliphan Rajkumar T, Kiruthika S, Suganya T and Femila F 2019 Exploring hybrid and ensemble models for customer churn prediction in telecom sector *Int. J. Recent Technol. Eng.* **8**, pp.299–309
[8] Umayaparvathi V and Iyakutti K 2012 Applications of Data Mining Techniques in Telecom Churn Prediction **42**, pp.5–9
[9] Misra R, Singh S and Mahajan R 2019 An empirical study on the cellular subscribers churn, selection factors and satisfaction with the services *Int. J. Prod. Dev.* **23**, pp.105–21
[10] Chen Z Y, Fan Z P and Sun M 2012 A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data *Eur. J. Oper. Res.* **223**, pp.461–72
[11] Verbeke W, Dejaeger K, Martens D, Hur J and Baesens B 2012 New insights into churn prediction in the telecommunication sector : A profit driven data mining approach *Eur. J. Oper. Res.* **218** pp.211–29
[12] Keramati A, Aliannejadi M, Ahmadian I, Mozzafari M and Abbasi U 2014 Improved churn prediction in telecommunication industry using data mining techniques *Appl. Soft Comput. J.*
[13] Rodan A, Fayyoumi A, Faris H, Alsakran J and Al-kadi O 2015 Negative Correlation Learning for Customer Churn Prediction : A Comparison Study **2015**
[14] Turabieh H, Mafarja M and Li X 2019 Iterated feature selection algorithms with layered recurrent neural network for software fault prediction *Expert Syst. Appl.* **122**, pp.27–42
[15] Paul S and Das S 2015 Simultaneous feature selection and weighting - An evolutionary multi-objective optimization approach *Pattern Recognit. Lett.* **65**, pp.51–9
[16] Homsapaya K and Sornil O 2018 Modified Floating Search Feature Selection Based on Genetic Algorithm *MATEC Web Conf.* **164**, p.01023
[17] Jiawei H, Kamber M, Han J, Kamber M and Pei J 2012 *Data Mining: Concepts and Techniques*
[18] Gorunescu F 2011 *Data mining: concepts and techniques*
[19] Hira Z M and Gillies D F 2015 A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data *Adv. Bioinformatics* **2015**, pp.1–13
[20] Liu H, Jiang H and Zheng R 2016 The Hybrid Feature Selection Algorithm Based on Maximum Minimum Backward Selection Search Strategy for Liver Tissue Pathological Image Classification *Comput. Math. Methods Med.* **2016**
[21] Xue B, Zhang M, Browne W N and Yao X 2016 A Survey on Evolutionary Computation Approaches to Feature Selection *IEEE Trans. Evol. Comput.* **20**, pp.606–26
[22] Alpaydin E 2013 *Introduction to Machine Learning*, **53**
[23] Jain M and Richariya V 2012 An Improved Techniques Based on Naive Bayesian for Attack Detection *Int. J. Emerg. Technol. Adv. Eng.* **2**, pp.324–31
[24] Tadist K, Najah S, Nikolov N S, Mrabti F and Zahi A 2019 Feature selection methods and genomic big data: a systematic review *J. Big Data* **6**