Customer Churn Prediction through Attribute Selection Analysis and Support Vector Machine

Jia Yi Vivian Quek
Multimedia University

Ying Han Pang
Multimedia University

Zheng You Lim
Multimedia University

Shih Yin Ooi
Multimedia University

Wee How Khoh
Multimedia University

Abstract: An accurate customer churn prediction could alert businesses about potential churn customers so that proactive actions can be taken to retain the customers. Predicting churn may not be easy, especially with the increasing database sample size. Hence, attribute selection is vital in machine learning to comprehend complex attributes and identify essential variables. In this paper, a customer churn prediction model is proposed based on attribute selection analysis and Support Vector Machine. The proposed model improves churn prediction performance with reduced feature dimensions by identifying the most significant attributes of customer data. Firstly, exploratory data analysis and data preprocessing are performed to understand the data and preprocess it to improve the data quality. Next, two filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance (ANOVA), are applied to the pre-processed data to select relevant features. Then, the selected features are input into a Support Vector Machine for classification. A real-world telecom database is used for model assessment. The empirical results demonstrate that ANOVA outperforms the Chi-squared filter in attribute selection. Furthermore, the results also show that, with merely ~50% of the features, feature selection based on ANOVA exhibits better performance compared to full feature set utilization.

Keywords: Churn Prediction, Attribute Selection, Machine Learning, Filter Methods, Support Vector Machine.

http://doi.org/10.18080/jtde.v11n3.777

Introduction

The term "churn" describes a scenario in which a customer discontinues a company's services. This can occur as a consequence of unforeseeable events. For example, the Covid 19 outbreak has led to businesses going above and beyond to entice customers to stay loyal (Johny & Mathai, 2017). Predicting customer churn is a good strategy to reduce customer churn. With the statistics, businesses can identify those potential churn customers as well as the reasons. Johnny & Mathai (2017) claimed that churn rate is affected by a variety of factors, including demographic details, such as age, gender, marital status, and location, and customer behaviour, such as frequency of interaction with service providers, monthly revenue, and total recurring charges. While predicting customer churn is useful and helpful for a business, it can be difficult due to a massive database. Thus, attribute selection is an important process in machine learning that aids in comprehending the complicated relationships between attributes and identifying the essential factors while eliminating irrelevant or redundant ones (Albulayhi *et al.*, 2022).

In this paper, a customer churn prediction model based on attribute selection analysis and Support Vector Machine is proposed. With an effective attribute selection technique, the most significant customer data attributes can be identified, leading to enhanced churn prediction performance while minimizing the feature dimension. In this work, exploratory data analysis and data preprocessing are first performed to understand the customer data. Then, the data is pre-processed for data quality improvement, which helps improve the classification performance. Next, two filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance, are performed on the pre-processed data to analyse and select relevant features. Then, the selected feature sets are input into a Support Vector Machine for data classification. In this study, a real-world telecom database, i.e., Cell2Cell (2018), is used for model assessment.

The contributions of this study are listed as follows:

- A machine learning-based customer churn prediction framework is proposed for the telecommunication industry.
- The performance of filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance, in identifying the most important attributes for churn prediction is examined.
- The performance of the proposed customer churn prediction system is assessed based on real- world telecommunication customer data.

http://doi.org/10.18080/jtde.v11n3.777

Related Work

There are numerous works for customer churn prediction in the telecommunication business. Since this study is using the Cell2Cell (2018) dataset, the literature review in this study focuses on the previous research that was conducted on the Cell2Cell dataset. In the literature, it had been demonstrated that Support Vector Machine produced a better model performance than other classification algorithms, such as Decision Tree and neural networks, in churn classification (Umayaparvathi & Iyakutti, 2016; Vaidya & Nigam, 2022).

Shuli Wu and Wei-Chuen Yau *et al.* (2021) explored several machine learning algorithms for predicting customer churn. In this work, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set to address the issue of imbalanced data. Due to the large number of attributes in the dataset, a feature selection technique known as the Chi-squared test was performed to reduce the data dimensionality. The tested machine learning techniques adopted in this work were Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, AdaBoost, and Multi-Layered Perceptron (MLP). The empirical results showed that Multi-layer Perceptron demonstrated the best performance in terms of F1-score with 42.84%, whereas Random Forest was ranked the best in terms of accuracy with a score of 63.09% (Wu *et al.*, 2021).

Fujo *et al.* (2022) proposed Deep-BP-ANN using two feature selection techniques (i.e., Variance Thresholding and Lasso Regression). Furthermore, the authors also adopted the Random Oversampling (ROS) technique to solve the issue of imbalanced data in the Cell2Cell dataset. In this study, the performance was evaluated using a holdout set and 10-fold cross-validation. Different classifiers, such as Naïve Bayes, Logistic Regression, XG-Boost, and the KNN algorithm, were explored. From the experimental results, the XG-Boost algorithm surpassed the other machine learning algorithms (Fujo *et al.*, 2022).

Jain *et al.* (2022) performed churn prediction on a subset of the Cell2Cell database. This work mainly focused on the feature importance and feature engineering for churn data. Random Forest and Gradient Boosted Tree were examined for classification. From the empirical results, Gradient Boosted Tree performed better in terms of accuracy and sensitivity. In this work, additional new features based on specific rules were produced. It was shown that these new features achieved high importance for churn prediction (Jain *et al.*, 2022).

The Proposed System

There are four phases involved in the proposed system: (1) data collection and retrieval; (2) exploratory data analysis (EDA) and data preprocessing; (3) feature selection; and (4)

http://doi.org/10.18080/jtde.v11n3.777

model generation and training. Figure 1 illustrates the overview of the proposed customer churn prediction. The details of each phase will be explained in the following subsections.

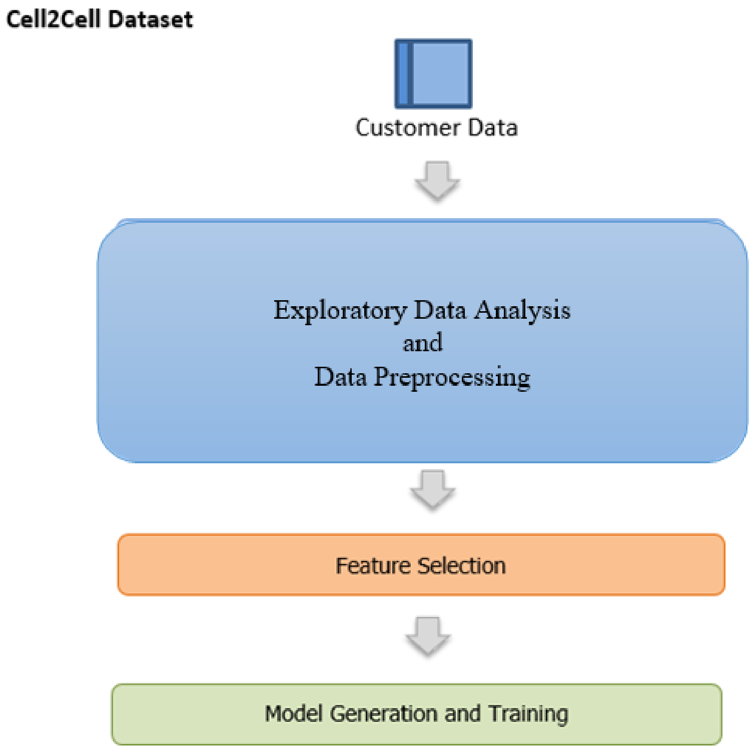


Figure 1. Flowchart of the proposed system

Customer data collection

Cell2Cell dataset is a telecommunication database containing 71047 instances and 58 attributes (Cell2Cell, 2018). The attributes include customer demographic information, product information, marketing-related data, service-related data, and payment-related data, etc., which includes numerical variables such as monthly revenue and roaming calls; binary variables, such as owned computer and new cell phone users; ordinal variables, such as credit rating and occupation, and so on. Table 1 records the details of the attributes. The attribute "Churn" is used as the class label.

Table 1. Details of the Attributes in the Cell2Cell dataset

Numerical Attributes	Data Format	Categorical Attributes	Data Format
CustomerID	int	Churn	Yes/No
Monthly Revenue	float	ServiceArea	string
MonthlyMinutes	float	ChildrenInHH	Yes/No
TotalRecurringCharge	float	HandsetRefurbished	Yes/No
DirectorAssistedCalls	float	HandsetWebCapable	Yes/No
OverageMinutes	float	TruckOwner	Yes/No
RoamingCalls	float	RVOwner	Yes/No
PercChangeMinutes	float	Homeownership	Known/Unknown
PercChangeRevenues	float	BuyViaMailOffers	Yes/No
DroppedCalls	float	RespondsToMailOffers	Yes/No

Numerical Attributes	Data Format	Categorical Attributes	Data Format
BlockedCalls	float	OptOutMailings	Yes/No
UnansweredCalls	float	NonUSTravel	Yes/No
CustomerCareCalls	float	OwnsComputer	Yes/No
ThreewayCalls	float	HasCreditCard	Yes/No
ReceivedCalls	float	NewCellphoneUser	Yes/No
OutboundCalls	float	NotNewCellphoneUser	Yes/No
InboundCalls	float	OwnsMotorcycle	Yes/No
PeakCallsInOut	float	HandsetPrice	string
OffPeakCallsInOut	float	MadeCallToRetentionTeam	Yes/No
DroppedBlockingCalls	float	CreditRating	string
CallForwardingCalls	float	PrizmCode	Other/Suburban/ Town/Rural
CallWaitingCalls	float	Occupation	Other/Professional/ Crafts/Clerical/Self/ Retired/Student/ Homemaker
MonthsInService	int	MaritalStatus	Unknown/Yes/No
UniqueSubs	int		
ActiveSubs	int		
Handsets	float		
HandsetModels	float		
CurrentEquipmentDays	float		
AgeHH1	float		
AgeHH2	float		
RetentionCalls	int		
RetentionOffersAccepted	int		
ReferralsMadeBySubscriber	int		
IncomeGroup	int		
AdjustmentsToCreditRating	int		

Exploratory data analysis and data preprocessing

The Cell2Cell dataset is saved as a CSV file and the dataset variable type is a Pandas DataFrame, which is a two-dimensional structure used to examine a wide range of tabular data with associated labels. In this database, there are missing values and outliers, particularly in variables such as monthly revenue and total recurring charge. Moreover, some variables require data transformation to achieve the desired form for further processes. Therefore, data preprocessing is critical for cleaning and transforming these variables in order to improve prediction performance. The dataset is checked for missing values using the isnull() method. Instead of dropping the null values, the np.nan technique is used to clean the data by replacing those missing variables with zero. In the Cell2Cell database, there are 14 attributes with missing values, as recorded in Table 2.

After analyzing the data information, two attributes are eliminated from consideration for data learning and analysis in this study. The attributes are "CustomerID", which is the identifier for each client and is not useful in predicting churn behaviour, and "ServiceArea" attribute. The removal of "Service Area" is because it could cause a lot of noise in the dataset (Jain *et al.*, 2022). Most values in this attribute are unique after the label encoding process, resulting in ~740 different category labels.

Table 2. Missing Variables in the Dataset

Attribute Name	Total Missing Values
Monthly Revenue	156
MonthlyMinutes	156
TotalRecurringCharge	156
DirectorAssistedCalls	156
OverageMinutes	156
RoamingCalls	156
PercChangeMinutes	367
PercChangeRevenues	367
ServiceArea	24
Handsets	1
HandsetModels	1
CurrentEquipmentDays	1
AgeHH1	909
AgeHH2	909

By utilizing log transformation, the skewness of numerical variables is decreased. Next, the values are further transformed to the range from 0 to 1. Next, the categorical variables are converted into numerical variables because the machine learning model only accepts numerical input data. Thus, a label encoding method, known as one-hot encoding, is applied to convert those categorical variables from string format into numerical variables. In the Cell2Cell database, the attribute of "Prizm Code" has four categories: other, suburban, town, and rural. After performing label encoding, the categories are represented in numerical formats of 0, 1, 2, and 3, respectively. Figure 2 illustrates the example of label encoding on the attribute of "Prizm Code".

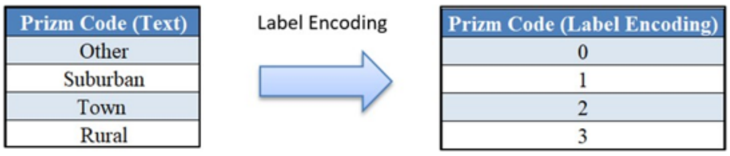


Figure 2. Label encoding on a categorical variable

In Exploratory Data Analysis, experimental results show that numerous strategies can be used to analyze data and improve the performance of predictive models (Zheng, 2022). A correlation heatmap is used to identify the correlations between variables, as it allows easier comparison between different pairs of variables. The colour palette in the legend represents the degree of correlation between the factors. In this study, we adopt the colour "Blues" to represent the heatmap; see Figure 3. The darker shade indicates a higher correlation, indicating that the variables tend to move in the same direction; whereas the lighter shade indicates a lower correlation, indicating that the variables are not closely related to one another. We can swiftly and easily determine which variables are most strongly correlated and which variables are independent by using a heatmap. Figure 4 illustrates the correlation heatmap between the attributes/variables of the Cell2Cell database. From the map, we can discover that there are a few attributes that are highly correlated. Attribute selection analysis will be performed to determine a subset of relevant and informative attributes for further processes.

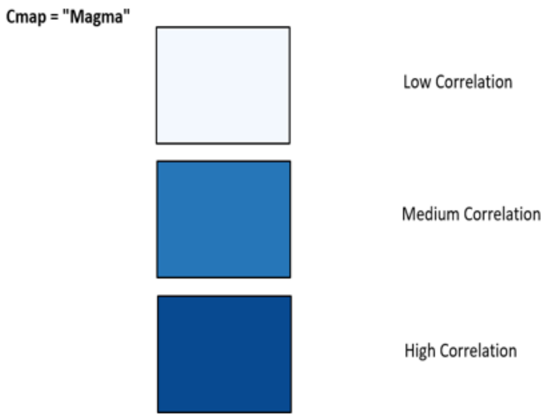


Figure 3. Correlation heatmap with different correlations identified using different colours

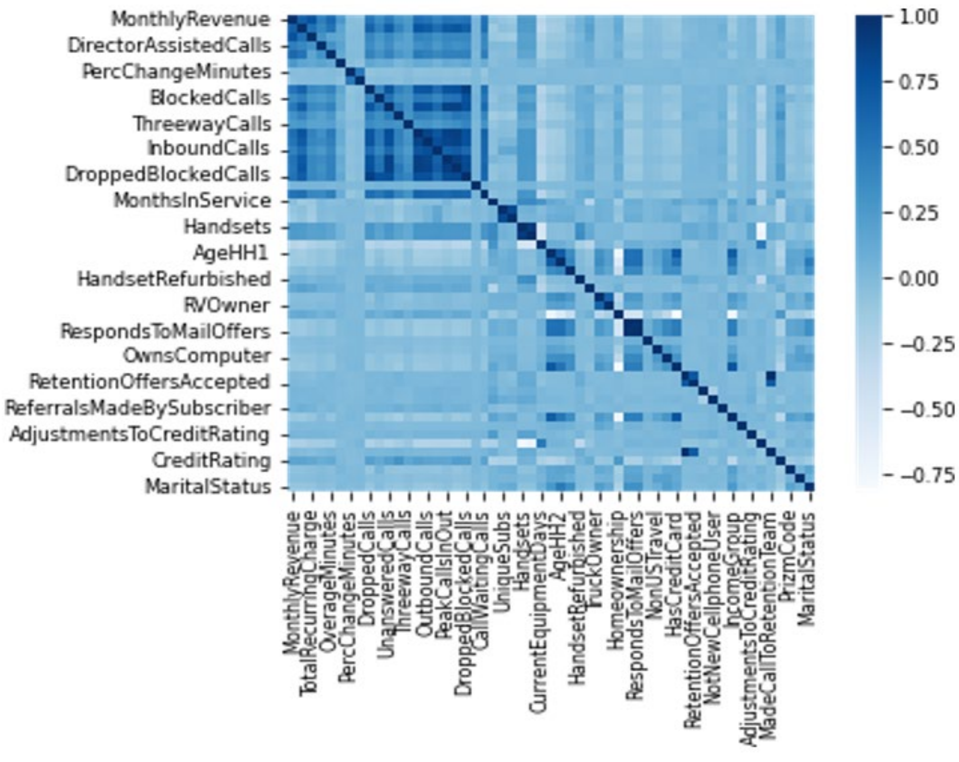


Figure 4. Correlation heatmap of Cell2Cell dataset

In some research, there are a few researchers who have neglected the issue of class imbalance, and the majority of the studies described above used historical data, primarily from Kaggle, and were conducted in wealthy countries. Such data, however, may not adequately reflect the issues in the real world. Besides that, the bulk of the research relied on a small number of datasets, which may have restricted the development and selection of better models that could

http://doi.org/10.18080/jtde.v11n3.777

define the overall issue (Seid & Woldeyohannis, 2022). In this study, it is observed that the churn rate of the Cell2Cell dataset provided by Kaggle is imbalanced, i.e., having more non-churn samples (i.e., majority class) than churn samples (i.e., minority class), with ~70% and ~30% in each class. Imbalanced data might lead to bias towards the majority class in the dataset. Thus, a data sampling technique, known as the Synthetic Minority Oversampling method (SMOTE), is used to balance the dataset's churn rate. This technique is an oversampling technique. By increasing the representation of the minority class, data sampling could balance the dataset and provide a more representative sample for training and testing the model. A balanced dataset is required for training the models, because it guarantees that the model is not biased towards one class and can make accurate predictions for both. SMOTE creates synthetic minority class observations by interpolating between minority class observations that already exist. This is accomplished by randomly selecting an observation from a minority class, then locating its k-nearest neighbours in the feature space. Next, one of the neighbours is randomly selected and a new observation is produced by interpolating between the selected observation and the randomly selected neighbour. Before implementing the SMOTE method, our training dataset contains 31265 samples with 9010 churn samples (28.8%) and 22255 non-churn samples (71.2%). After implementing the SMOTE method, our training dataset contains 44510 samples with 22255 churn samples (50%) and 22255 non-churn samples (50%).

Attribute selection analysis

Given the growing sample size of the database, predicting customer churn may not be simple. Research has stated that, even though the predictive models perform well at first, the addition of the proposed feature selection approach recursive feature elimination (RFE) results in a significant improvement in their performance. Regardless of the machine learning algorithms used to predict customer churn, a feature selection method should be included. However, establishing the best feature selection strategy for customer churn prediction in the telecoms business remains a difficult task (Naing *et al.*, 2022). As a result, one of the crucial methods to understand complicated attribute interactions is through attribute selection. Attribute selection analysis is able to discover crucial variables while removing those unnecessary and redundant ones. In other words, attribute selection can aid in the selection of the best representative features that might contribute to analyzing customer churn behaviour.

In this study, two filter-based approaches for attribute selection are examined: Chi-Squared Test and Anova (analysis of variance) Test method. The Chi-squared test is a statistical approach for comparing actual outcomes to predicted results. The primary goal of this test is to establish whether the difference between observed data (actual results) and predicted data

http://doi.org/10.18080/jtde.v11n3.777

(predictions) is due to chance or a real relationship between the variables under consideration. This method is frequently used to select highly related sample data, and then uses a minimum redundancy algorithm to further remove redundancy and select features (Wang & Zhou, 2021). When two features are unrelated, the actual and anticipated results are likely to be comparable, resulting in a lower Chi-squared score. A high Chi-squared score, on the other hand, indicates that the independence claim is unjustified, when there is a strong link between the variables. In other words, attributes with higher Chi-squared values are more reliant on the response variable and can be used to train the model. Figure 5 illustrates the Chi-squared score (representing the feature significance score in this study) of the Cell2Cell attributes, ranked from the highest score to the lowest score. In this study, different numbers of selected features (i.e., 10, 20, 30, 40 and 50) are examined. The results indicated that the top 30 features yielded better performance, so the top 30 significant features are selected for the subsequent processes (see Table 3). Note that the p-value is the area under the density curve of the Chi-squared distribution to the right of the value of the test statistic.

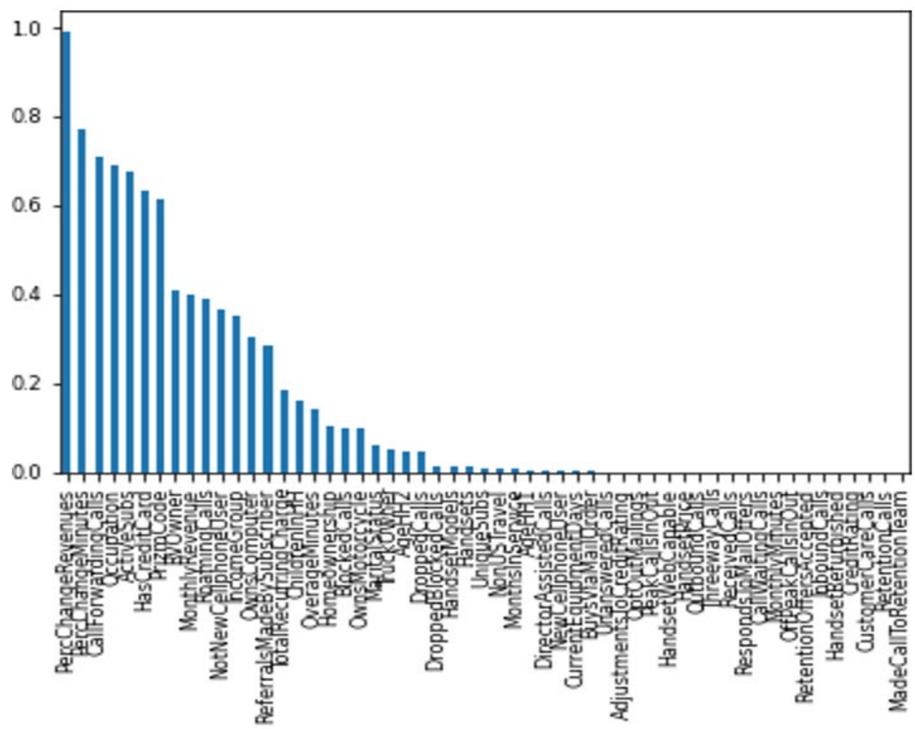


Figure 5. Chi-Squared Test Attribute Selection Results

The Anova Test, often known as the SelectKBest method, is a technique for choosing features based on their scores. It is used to perform the ANOVA statistical test and select the 30 most important features from the original dataset (Lazaros *et al.*, 2022). The technique deliberately excludes features with lower scores and selects the top *k* features with the highest scores. This is because it can help minimize data dimensionality while keeping the most important features; it is frequently used as a feature selection approach in machine learning for classification performance improvement. Table 4 records the top 30 significant features based on the Anova test. These features will be considered for the next processes.

Table 3. Top 30 Attributes selected by Chi-Squared Test Method

No.	Attribute Name	Scores	No.	Attribute Name	Scores
1	PercChangeRevenues	9.881e-01	16	ChildrenInHH	1.607e-01
2	PercChangeMinutes	7.693e-01	17	OverageMinutes	1.411e-01
3	CallForwardingCalls	7.072e-01	18	Homeownership	1.021e-01
4	Occupation	6.889e-01	19	BlockedCalls	9.925e-02
5	ActiveSubs	6.753e-01	20	OwnsMotorcycle	9.889e-02
6	HasCreditCard	6.324e-01	21	MaritalStatus	6.055e-02
7	PrizmCode	6.140e-01	22	TruckOwner	5.109e-02
8	RVOwner	4.100e-01	23	AgeHH2	4.711e-02
9	MonthlyRevenue	4.001e-01	24	DroppedCalls	4.405e-02
10	RoamingCalls	3.875e-01	25	DroppedBlockedCalls	1.444e-02
11	NotNewCellphoneUser	3.638e-01	26	HandsetModels	1.319e-02
12	IncomeGroup	3.488e-01	27	Handsets	1.220e-02
13	OwnsComputer	3.035e-01	28	UniqueSubs	8.466e-03
14	ReferralsMadeBySubscriber	2.844e-01	29	NonUSTravel	6.916e-03
15	TotalRecurringCharge	1.844e-01	30	MonthsInService	5.468e-03

Table 4. Top 30 Attributes selected by Anova Test Method

No.	Attribute Name	Scores	No.	Attribute Name	Scores
1	CurrentEquipmentDays	775.376	16	ReceivedCalls	121.937
2	MonthlyMinutes	433.309	17	HandsetPrice	101.028
3	TotalRecurringCharge	263.868	18	ThreewayCalls	96.762
4	CustomerCareCalls	224.992	19	CallWaitingCalls	89.230
5	HandsetModels	200.445	20	UniqueSubs	88.353
6	CreditRating	196.644	21	MonthsInService	75.534
7	OffPeakCallsInOut	186.874	22	DirectorAssistedCalls	73.334
8	PeakCallsInOut	175.178	23	MonthlyRevenue	72.644
9	MadeCallToRetentionTeam	165.937	24	AgeHH1	58.101
10	InboundCalls	165.011	25	DroppedBlockedCalls	55.591
11	Handsets	164.313	26	AdjustmentsToCredit-Rating	53.442
12	RetentionCalls	155.184	27	PercChangeMinutes	51.419
13	HandsetWebCapable	144.790	28	HandsetRefurbished	42.499
14	UnansweredCalls	135.294	29	RetentionOffers-Accepted	40.074
15	OutboundCalls	122.552	30	DroppedCalls	32.330

Classification model

The Support Vector Machine (SVM) classifier is a supervised machine learning approach which can classify both linear and nonlinear data. In other words, the SVM classifier can be categorized into:

- Linear SVM — operates on linearly separable data where the statistics that can be divided into groups by a single straight line;

- Non-linear SVM — operates on non-linearly separable data where the statistics cannot be divided into groups by a straight line.

This classifier adopts kernel approaches to transform data from a low-dimensional space to a higher-dimensional space where a distinct separation can be made. It aims to find the best separation line (known as a hyperplane — a decision boundary) that could optimally separate the classes of data points. In this work, SVM Radial Basis Function is adopted. An SVM classifier is considered in this study due to the following advantages:

- It can work well for this study case where the sample number exceeds the number of feature dimensions;
- Its regularization parameter aids in preventing overfitting;
- It can handle non-linearly separable data (real-world data is often nonlinear). By utilizing kernel functions, the SVM classifier is able to transform the real-world nonlinear data into a higher-dimensional space where the data can be separated linearly, as illustrated in Figure 6.

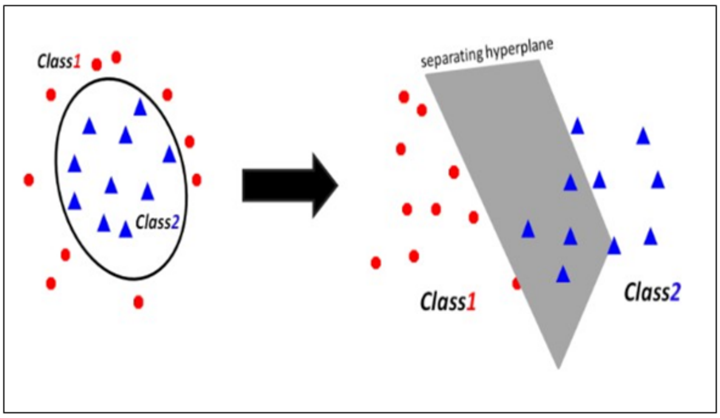


Figure 6. Nonlinear SVM: mapping data into the higher dimensional feature space (right). (The figure is extracted from the work of Mahmoodi *et al.* (2011))

Experimental Results and Discussion

In this study, the experiments are conducted using a train-test split protocol to evaluate the performance of the proposed customer churn prediction model. The Cell2Cell dataset is split into two separate subsets: a training set for training the model; and a testing set to assess the model's performance and its generalization ability to unseen data. After employing the train-test split protocol, there are 44510 training samples and 15315 testing samples from the database. We adopt several performance metrics for performance evaluation, such as precision, recall, and F1 score, as well as the confusion matrix. Precision, recall, and F1 score are formulated as below:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

where *TP* is the number of true positive classifications; *FP* is the number of false positives; *FN* is the number of false negatives.

Before training the model, the top significant 30 attributes are selected for evaluating the model's performance. After selecting the attributes, the Support Vector Machine (SVM) model results are analyzed with three performance metrics, which are precision, recall, and F1 score, by using the two filter-based feature selection techniques, Chi-squared Test and Anova Test. Moreover, the achieved confusion matrix is also provided for reference purposes.

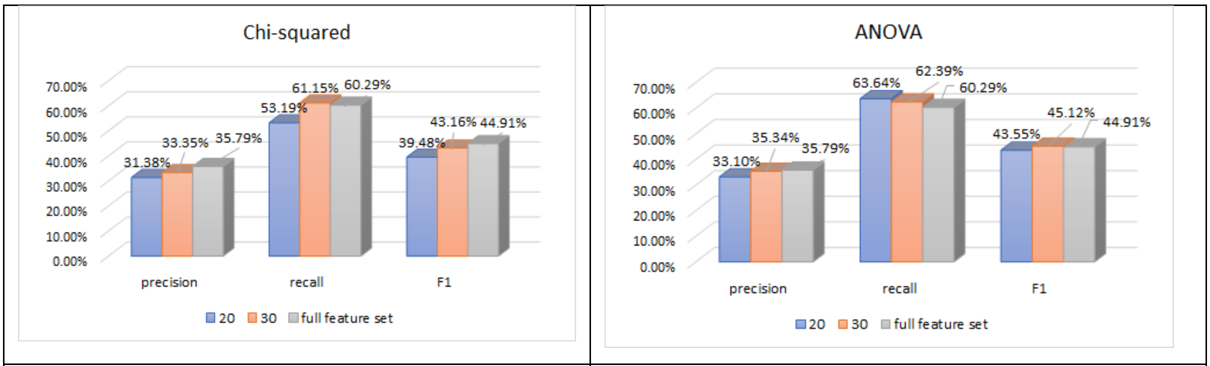
Table 5 records the classification performance (in terms of precision, recall, and F1 score) of the proposed customer churn prediction model with different attribute selection analyses. Figure 7 shows the performance for better illustration. From the obtained empirical results, we can observe that the proposed system using the Anova Test as attribute selection analysis performs better than that using the Chi-squared test. The former model achieves a precision score of 35.54%, a recall score of 62.39%, and an F1 score of 45.12%; whereas the latter model obtains a precision score of 33.35%, a recall score of 61.14%, and an F1 score of 43.16%. Furthermore, the experimental results show that the proposed model using a full feature set for classification attains a precision score of 35.79%, a recall score of 60.29%, and an F1 score of 44.91%.

Table 5. Classification Performance of the Proposed Customer Churn Prediction Model with Different Attribute Selection Analyses

Attribute Selection Analysis	Feature Dimension	Precision (%)	Recall (%)	F1-Score (%)
Chi-Squared	20	31.38	53.19	39.48
	30	33.35	61.15	43.16
Anova Test	20	33.1	63.64	43.55
	30	35.34	62.39	45.12
Full Feature Set	55	35.79	60.29	44.91

Figure 8 illustrates the confusion matrices of the proposed models (using the Chi-squared test and Anova test) with different feature selection dimensions. It is understood that the core objective of a customer churn prediction model is to identify customers who are very likely to leave a business (i.e., stopping purchasing/subscribing to the company's product or service), so that necessary actions can be taken to stop them from churning. From the figure, we can observe that the models can predict customer churn with a feature dimension, *K*, of 30, compared to that of 20. In the model with the Chi-squared test, 2699 out of 4414 churn customers are able to be identified by using K=30, compared to the model with K=20 having

2348 churn customers detection. Similar to the models with the Anova test, K=30 allows for the identification of 2809 of 4414 churning customers, as opposed to 2754 with K=20.



(a)	(b)

Figure 7. Performance of the proposed prediction model with different attribute selection analysis: (a) Chi-squared and (b) Anova

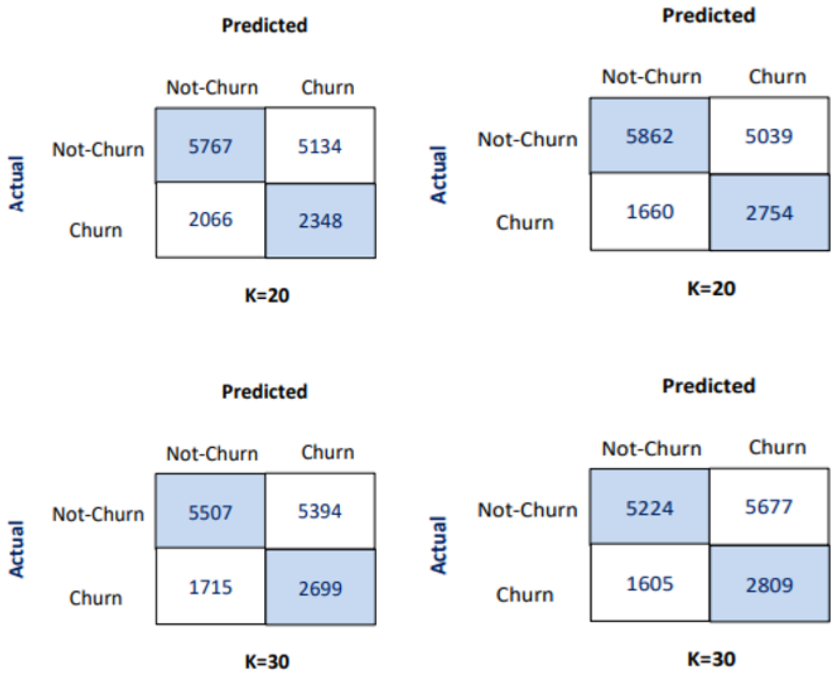


Figure 8. Confusion Matrix for Chi-squared Test (Left) and Anova Test (Right) with K values of 20 and 30

Conclusions

In this paper, a machine learning-based customer churn prediction model is proposed with filter-based attribute selection analysis and Support Vector Machine (SVM). In this research, a real-world telecommunication database, i.e., Cell2Cell dataset, is employed to assess the performance of the proposed churn prediction model. In order to better comprehend the customer data and prepare it for feature analysis, exploratory data analysis and data preprocessing are first carried out. These processes help increase classification performance by enhancing data quality. Next, relevant and representative features are selected using two filter-based attribute selection techniques, namely Chi-squared and Anova. The top 30 significant features are then passed into a Support Vector Machine to classify the data. From the obtained empirical results, it is observed that the proposed model with the Anova test attains a higher F1 score of 45.12% compared with the model with the Chi-squared test (with an F1 score of 43.16%) and the model with a full feature set (with F1 score of 44.91%). It is found that the Anova test is a better attribute strategy compared to the Chi-squared test. The finding shows the importance of using an adequate attribute selection strategy to comprehend complex attribute relations and obtain representative features for improved data classification. Furthermore, by eliminating irrelevant and redundant attributes, it helps in effective computation.

Acknowledgements

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

References

Albulayhi, K., Abu Al-Haija, Q., Alsuhibany, S. A., Jillepalli, A. A., Ashrafuzzaman, M., & Sheldon, F. T. (2022). IoT intrusion detection using machine learning with a novel high performing feature selection method. *Applied Sciences*, *12*(10), 5015. https://doi.org/10.3390/app12105015

Cell2Cell. (2018). *Telecom Churn (Cell2Cell)*. [Online]. Available at https://www.kaggle.com/jpacse/datasets-for-churn-telecom

Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, *11*(1), 24. http://dx.doi.org/10.18576/isl/110120

Jain, H., Khunteta, A., & Srivastava, S. (2022). Telecom Churn Prediction Using an Ensemble Approach with Feature Engineering and Importance. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(3), 22–33. https://ijisae.org/index.php/IJISAE/article/view/2134

Johny, C. P., & Mathai, P. P. (2017). Customer churn prediction: A survey. *International Journal of Advanced Research in Computer Science*, 8(5), 2178-2181. http://www.ijarcs.info/index.php/Ijarcs/article/view/4079

Lazaros, K., Tasoulis, S., Vrahatis, A., & Plagianakos, V. (2022). Feature Selection For High Dimensional Data Using Supervised Machine Learning Techniques. *IEEE International Conference on Big Data (Big Data)*, 2022 (pp. 3891–3894). IEEE. https://doi.org/10.1109/BigData55660.2022.10020654

Mahmoodi, D., Soleimani, A., Khosravi, H., & Taghizadeh. M. (2011). FPGA Simulation of Linear and Nonlinear Support Vector Machine. *Journal of Software Engineering and Applications*, *5*(4), 320–328. http://dx.doi.org/10.4236/jsea.2011.45036

Naing, Y. T., Raheem, M., & Batcha, N. K. (2022). Feature Selection for Customer Churn Prediction: A Review on the Methods & Techniques applied in the Telecom Industry. *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022 (pp. 1–5). IEEE. https://doi.org/10.1109/ICDCECE53908.2022.9793315

Seid, M. H., & Woldeyohannis, M. M. (2022). Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia. *International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 2022 (pp. 1–6). IEEE. https://doi.org/10.1109/ICT4DA56482.2022.9971224

Umayaparvathi, V., & Iyakutti, K. (2016). A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. *International Research Journal of Engineering and Technology*, *3*(4), 1065-1070. https://www.irjet.net/archives/V3/i4/IRJET-V3I4213.pdf

Vaidya, S., & Nigam, R. K. (2022). An Analysis of Customer Churn Predictions in the Telecommunications Sector. *International Journal of Electronics Communication and Computer Engineering*, *13*(4), 37–43. https://www.ijecce.org/index.php/component/jresearch/?view=publication&task=show&id=1382&Itemid=437

Wang, Y., & Zhou, C. (2021). Feature selection method based on chi-square test and minimum redundancy. In Emerging Trends in Intelligent and Interactive Systems and Applications: *Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020)* (pp. 171–178). Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-63784-2_22

Wu, S., Yau, W. C., Ong, T. S., & Chong, S. C. (2021). Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access*, 9, 62118–62136. https://doi.org/10.1109/ACCESS.2021.3073776

Zheng, K. (2022). Identifying Churning Employees: Machine Learning Algorithms from an Unbalanced Data Perspective. *5th International Conference on Machine Learning and Machine Intelligence*, 2022 (pp. 14–22). https://doi.org/10.1145/3568199.3568202

http://doi.org/10.18080/jtde.v11n3.777