NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696 Website: ijiert.org
VOLUME 8, ISSUE 5, May. -2021

A REVIEW ON CUSTOMER CHURN PREDICTION USING MACHINE LEARNING APPROACH

Kiran Dhangar
Department of Computer Engineering,PUSA Institute Of Technology, Delhi,India
* krndhangar2@gmail.com

Prateek Anand
Phd Scholar NSUT East Campus Delhi
anand.prateek91@gmail.com

ABSTRACT

Customer churn is a problem that affects firms in a variety of industries. Every time a client departs, there will be significant loss of a firm. Churn prediction refers to determining which consumers are most likely to cancel a service subscription based on how they use it. It's a crucial prediction for many firms because getting new customers is sometimes more expensive than keeping old ones. There are six faces to our suggested methodology. Data pre-processing and exploratory data analysis are performed in the first two faces. In the third phase, feature selection is considered; after that, the data is divided into two portions, train and testset, in a ratio of 80% and 20%, respectively. The most prominent prediction models, such as logistic regression, naive bayes, support vector machine, and random forests, were used on the train set, and ensemble approaches were used to evaluate how they impacted model accuracy. In addition, for hyperparameter tuning and to avoid overfitting of models, k-fold cross validation was applied. Finally, the AUC/RUC curve was used to analyse the findings obtained on the test set. Random Forest and SVM were shown to have the highest accuracy of 87 percent and 84 percent, respectively. Random Forest achieves the greatest AUC score of 94.5 percent, while SVM classifiers obtain 92.1 percent, outperforming others.

INTRODUCTION

The telecommunications industry's globalisation As the number of operators in the market grows, so does the level of competition. [9] It's not simple to stay competitive in the digital age. Customer dissatisfaction with consumer service and assistance is the primary cause of churn. Forecasting clients who are likely to churn is the key to solving this problem. [18,27,34].With the increased competition in the market for providing services, the goal of customer churn prediction is to assist in the development of customer retention strategies. Customer retention is a set of activities that a business undertakes to improve the number of repeat customers and profits. The customer churn prediction model attempts to forecast voluntary attrition by identifying churn signals [43]. Churn is a metric that measures how many customers have abandoned your brand by cancelling their subscriptions or discontinuing to pay for your services. This is bad news for any firm because attracting a new customer costs five times as much as keeping an old one. A high customer churn rate will have a significant financial impact on your business. You can predict future churners who are going to depart your services by applying advanced artificial intelligence approaches such as machine learning (ML) [11,1]. As a result, many businesses have discovered that their existing database is one of their most valuable assets [11], and churn prediction, according to Abbasdimehr,[1] is a beneficial tool for predicting at-risk clients.

LITERATURE SURVEY

This article provides a brief overview of churn prediction in the telecom business, as well as related studies [2,7,12,20,21,23,27,28,31,35,38–40].

The authors of Adbelrahim et al. [3] used tree-based algorithms to forecast customer turnover, including decision trees, random forests, GBM tree methods, and XGBoost.

In a comparison, XGBoost outperformed the competition in terms of AUC accuracy. However, the accuracy of the feature selection process can be increased further by employing optimization algorithms. Support vector machine, decision tree, naive bayes, and logistic regression were used in a comparative comparison of machine learning models for customer churn prediction by Praveen et al. [5]. Following that, they looked at the impact of boosting algorithms on classification accuracy. SVM-POLY 123P showed up in the results. Lalwani et al., who used AdaBoost, outperformed the rest. However, combining feature selection procedures such as uni-variate selection and others can enhance classification accuracy even further. Horia Beleiu et al. [7] used three machine learning algorithms to predict customer churn: neural networks, support vector machines, and bayesian networks. Principle component analysis (PCA) is used to minimise the dimensionality of the data during the feature selection phase. However, by employing an optimization technique to improve the feature selection process, the classification accuracy can be enhanced. Gain measure and ROC curve were employed in the performance evaluation. The authors of J. Burez et al. [8] attempted to represent the problem of class imbalance. They used logistic regression and random forest techniques, as well as re-sampling. Additionally, boosting methods were used. AUC and Lift are taken into account in the performance analysis. They also looked at the impact of modern sampling techniques like CUBE, however the results showed no improvement in performance. However, employing optimization-based sampling strategies, the problem of class imbalance can be tackled more effectively.

The authors of K Coussement et al. [11] used support vector machines, logistic regression (LR), and random forest to try to capture the churn prediction problem (RF). SVM's performance was initially comparable to that of LR and RF, however when optimal parameter selection is taken into account, SVM beats both LR and RF in terms of PCC and AUC.On the churn prediction data-set, K. Dahiya et al. [12] used two machine learning models, namely decision tree and logistic regression. The WEKA tool was employed in the experimentation. Other machine learning techniques, on the other hand, can be used to tackle the aforementioned problem in a more efficient manner.

The authors of Umman et al. [16] used logistic regression and decision tree machine learning models to examine a large data set, but the accuracy attained was low. As a result, further work is needed before other machine learning and feature selection techniques may be used.J. Hadden et al. [17] investigate the factors that influence reverence churn. They also conducted a comparison of three machine learning methods, including neural networks, regression trees, and regression. Because of its rule-based architecture, the acquired results confirm that decision tree is better than others. Using existing feature selection approaches, the obtained accuracy can be enhanced even more.J. Hadden et al. [18] provided a comprehensive evaluation of all machine learning models considered, as well as a detailed study of existing feature selection strategies. They discovered that the decision tree outperformed the others in the prediction models. Optimization strategies are also important in feature selection since they improve prediction algorithms. Following a comparison of existing methodologies, the authors proposed a research route for future research.

1.2 Existing Problem

In order to capture the aforementioned issue, the organisation must correctly predict the customer's behaviour. There are two approaches to managing customer churn: (1) reactive and (2) proactive. In a reactive strategy, the company waits for the consumer to submit a cancellation request, following which it offers the consumer enticing retention options. The likelihood of churn is projected under the proactive method, and clients are provided plans accordingly. It's a binary classification problem in which churners and non-churners are separated.

To address this issue, we applied the following Machine Learning approaches in this study: Logistic Regression , Naive Bayes, Support Vector Machine (SVM), Random Forest Classifier. Furthermore, the data was pre-processed and significant feature vectors were identified using the gravitational search technique for a better comprehension of the data (GSA). The linearity of the data has also been examined and examined in order to employ appropriate Machine learning methods.

1.3 Methodology to be used

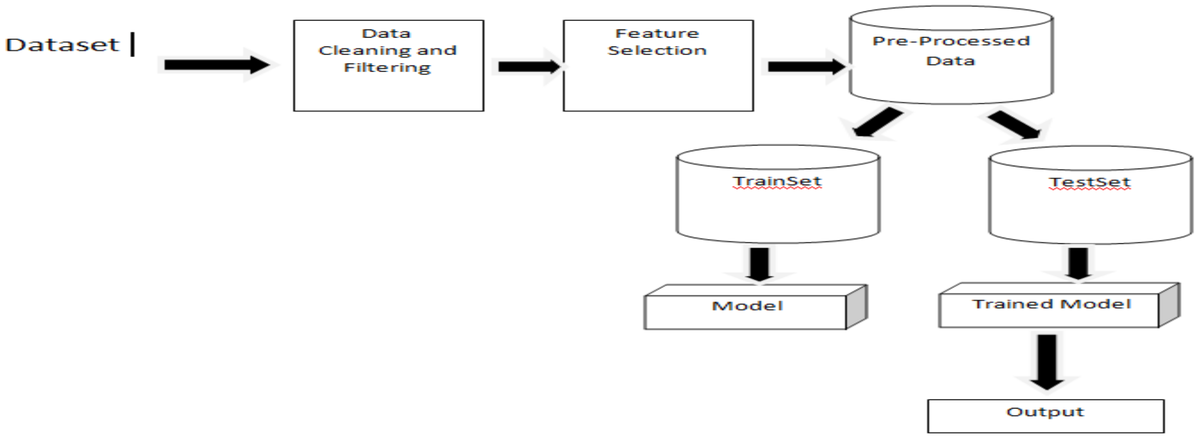
The following is a summary of our contribution:

• We used data cleaning, data pre-processing, data analysis, feature selection, and data reduction to lower the data-dimensions. set's

• We employed some of the most well-known machine learning techniques for prediction, such as logistic regression, SVM, and Random Forest, as well as k-fold cross validation to avoid overfitting.

• After that, we harnessed the power of ensemble learning to improve algorithms and obtain better outcomes**.**

• The algorithms were then evaluated on a test set using RUC/AUC curves, which were presented in the form of graphs and tables to compare which algorithm performed best for this data set.



Exploratory Data Analysis is a method of evaluating or comprehending data in order to derive insights or key characteristics. As we can see, loading the data into the pandas data frame is one of the most crucial phases in EDA, as the values from the data set are comma-separated. All we have to do now is read the CSV into a data frame, and pandas will take care of the rest.

The data will then be loaded into a Pandas data frame. We'll utilise the "Telco Customer Churn Dataset" dataset for this research. We examined the dataset by inspecting a few of the rows with the head() method, which retrieves the dataset's first five records. The dimensions of the data were observed using shape.

(7043, 21)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

The info () method displays information about the data, such as Column Name, Number of non-null entries in our columns, Data Type, and Memory Usage.

Missing values in the dataset are dealt with. Fortunately, there are no missing values in this dataset, but the real world isn't as simple as our example. isna.df (). total () We'll utilise the describe() method, which displays the following basic statistical properties for each numerical feature (int64 and float64 types): number of non-missing values, mean, standard deviation, range, median, 0.25, 0.50, 0.75 quartiles.

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

We look for duplicate values in our dataset since duplicate values will reduce the accuracy of our machine learning model. In our scenario, there are no duplicate rows.

number of rows with duplicates: (0, 21)

This step is essential in every EDA since there are occasions when we have a large number of columns that we never utilise; in these circumstances, dropping is the only option. The columns such as customerID don't make sense in this scenario, so they were removed for this instance; now the total number of rows and columns is:

(7043, 20)

We determined the percentage of customers who stayed with the company and those who left.

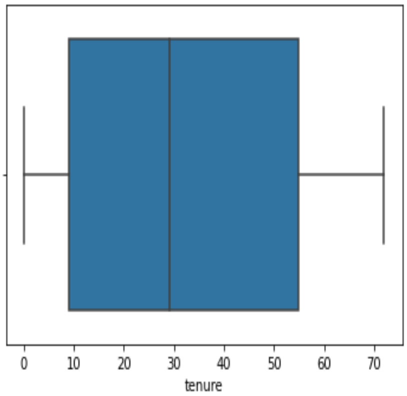
Customers stayed with the company 73.4630129206304 percent .

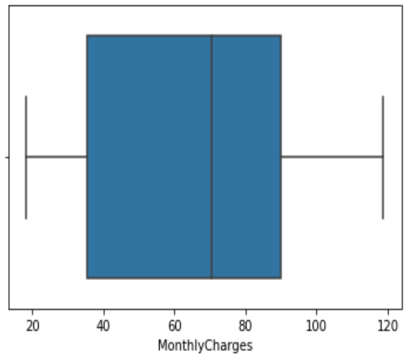
26.536987079369588% of customers abandoned the business.

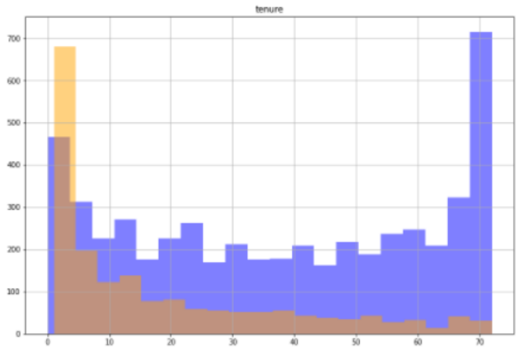
We check for data types because numerical data can sometimes be saved as a string or object; in that case, we must convert the string to integer data before plotting the data on a graph.

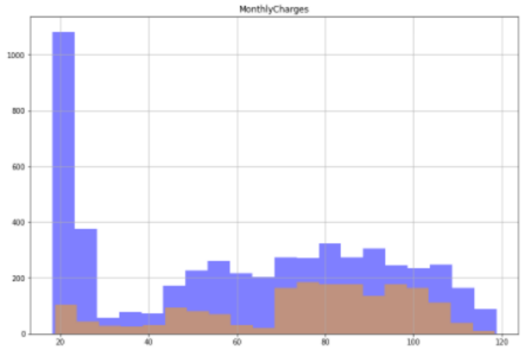
```
gender                 int32
SeniorCitizen          int64
Partner                int32
Dependents             int32
tenure                 int64
PhoneService           int32
MultipleLines          int32
InternetService        int32
OnlineSecurity         int32
OnlineBackup           int32
DeviceProtection       int32
TechSupport            int32
StreamingTV            int32
StreamingMovies        int32
Contract               int32
PaperlessBilling       int32
PaymentMethod          int32
MonthlyCharges       float64
TotalCharges           int32
Churn                  int32
dtype: object
```

Handling outliers in the data, i.e. the data's extreme values A Boxplot can be used to discover outliers in our data.









Compare and contrast different features (scatter), as well as frequency (histogram)

The frequency of occurrence of variables in an interval is referred to as a histogram. In this example, we can examine several characteristics such as gender, InternetService, Tenure, and MonthlyCharges.

Data Normalization and Scaling : Normalization, also known as feature scaling, is a method of standardising a data set's range of features, which might vary greatly. As a result, we can use machine learning methods to preprocess the data. So, for the numerical values, we'll utilise StandardScaler, which employs the formula x-mean/std deviation.

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSu
0	0	0	1	0	0.013889	0	1	0	0	2	0	
1	1	0	0	0	0.472222	1	0	0	2	0	2	
2	1	0	0	0	0.027778	1	0	0	2	2	0	
3	1	0	0	0	0.625000	0	1	0	2	0	2	
4	0	0	0	0	0.027778	1	0	1	0	0	0	

There is a class imbalance when observation in one class is higher than observation in other class.

In our scenario, we discovered an unbalanced data set by measuring the number of customers who churned and those who did not.

(1869, 20) (5174, 20)

Oversampling the minority class is one way to deal with unbalanced datasets. Duplicating instances in the minority class is the easiest way, but these examples don't provide any new information to the model. Instead, new examples can be created by synthesising old ones. The Synthetic Minority Oversampling Technique, or SMOTE for short, is a method of data augmentation for the minority population.

```
Original dataset shape Counter({0: 5174, 1: 1869})
Resampled dataset shape Counter({0: 4804, 1: 4804})
```

1.4Train Test Split

When machine learning algorithms are used to make predictions on data that was not used to train the model, the train-test split process is used to measure their performance.

It's a quick and simple technique that allows you to compare the performance of different machine learning algorithms for your predictive modelling problem. In our scenario, the dataset was split into two parts: 80 percent training and 20 percent test.

1.5Training of Dataset

Predictive models, such as logistic regression, naive bayes support vector machine, and random forests, were used on the train set, and ensemble approaches were used to see how they impacted model accuracy.In addition, for hyperparameter tuning and to avoid overfitting of models, k-fold cross validation was used over the train set under the guise of data such as dependability, efficiency, and popularity in the research community [16,17,22,30,33,36].

2.Performance Analysis

Different metrics, such as precision, recall, accuracy, and F-measure [39], have been used to evaluate the performance of applied models or throughput of Customer Churn Prediction on the test set. It assesses the predictive models' ability to accurately estimate churning customers [10].

Table 1 Comparative analysis of different Machine Learning model

Model	Precision	Recall	F-1 Score	Support
Logistic Regression				
0	0.86	0.79	0.82	971
1	0.80	0.86	0.83	951
accuracy			0.83	1922
macro avg	0.83	0.83	0.83	1922
weighted avg	0.83	0.83	0.83	1922
GaussianNB ()				
0	0.85	0.79	0.82	971
1	0.80	0.86	0.83	951
accuracy			0.83	1922
macro avg	0.83	0.83	0.83	1922
weighted avg	0.83	0.83	0.83	1922
SVM				
0	0.79	0.92	0.85	971
1	0.90	0.75	0.82	951
accuracy			0.84	1922
macro avg	0.85	0.84	0.83	1922
weighted avg	0.84	0.84	0.83	1922
Random Forest				
0	0.88	0.86	0.87	971
1	0.86	0.89	0.87	951
accuracy			0.87	1922
macro avg	0.87	0.87	0.87	1922
weighted avg	0.87	0.87	0.87	1922

Finally, the AUC/RUC curve was used to analyse the findings obtained on the test set. Random Forest and SVM were shown to have the highest accuracy of 87 percent and 84 percent, respectively. Random Forest achieves the greatest AUC score of 94.5 percent, while SVM classifiers obtain 92.1 percent, outperforming others.

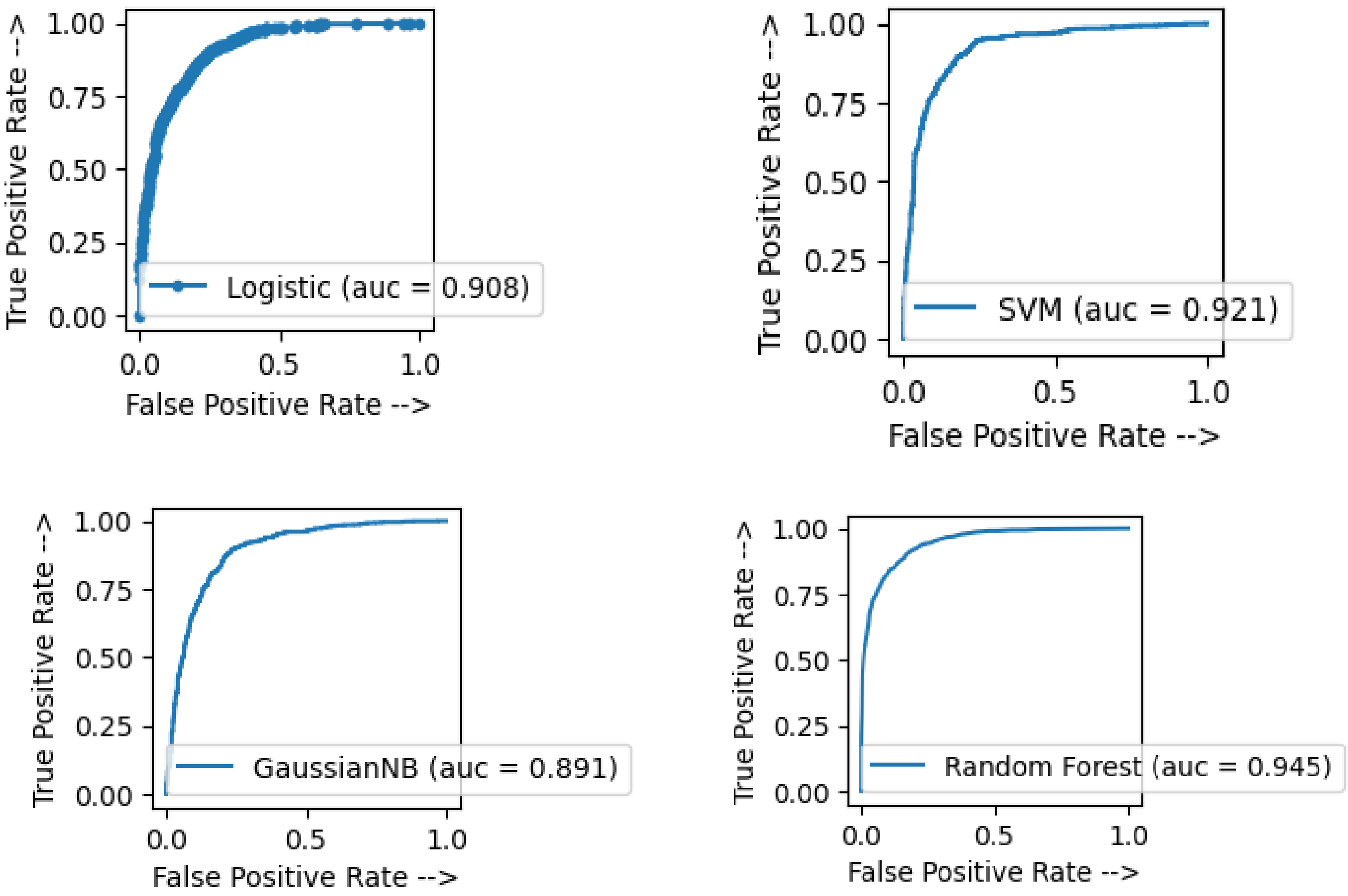


Fig 1 Model AUC curve a) Logistic regression b)SVM c)Gaussian NB d)Random Forest

CONCLUSION

The experimental results reveal that two learning techniques, Random Forest and SVM classifier, provide the highest accuracy when compared to other models, with AUC scores of 94.5 percent and 92 percent for the churn prediction problem, respectively. They outperformed other algorithms across the board, including accuracy, precision, F-measure, recall, and AUC score.

REFERENCES

1) Abbasimehr H, Setak M, Tarokh M (2011) A neuro-fuzzy classifier for customer churn prediction. International Journal of Computer Applications 19(8):35–41
2) Adwan O, Faris H, Jaradat K, Harfoushi O, Ghatasheh N (2014) Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis. Life Science Journal 11(3):75–81
3) Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data 6(1):28
4) Archambault, D., Hurley, N., Tu, C.T.: Churnvis: visualizing mobile telecommunications churn on a social network with attributes. In: 2013 IEEE/ACM International Conference on Advances in Social
5) Networks Analysis and Mining (ASONAM 2013), pp. 894–901. IEEE (2013)
6) Asthana P (2018) A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics 119(10):1149–1169

7) Aziz R, Verma C, Srivastava N (2018) Artificial neural network classification of high dimensional data with novel optimization approach of dimension reduction. Annals of Data Science 5(4):615–635
8) Brându¸soiu, I., Toderean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016 International conference on communications (COMM), pp. 97–100.IEEE (2016)
9) Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Systems with Applications 36(3):4626–4636
10) Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: From big data to big impact.MIS quarterly pp. 1165–1188 (2012)
11) Coussement K, De Bock KW (2013) Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. Journal of Business Research 66(9):1629–1636
12) Coussement K, Van den Poel D (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications 34(1):313–327
13) Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1–6 (2015)
14) Dong, T., Shang, W., Zhu, H.: Naïve bayesian classifier based on the improved feature weighting algorithm. In: International Conference on Computer Science and Information Engineering, pp. 142–147. Springer (2011)
15) Fawcett T (2006) An introduction to roc analysis. Pattern recognition letters 27(8):861–874
16) García S, Fernández A, Herrera F (2009) Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems. Applied Soft Computing 9(4):1304–1314
17) Gürsoy U ¸S (2010) Customer churn analysis in telecommunication sector. ˙Istanbul Üniversitesi ˙I¸sletme Fakültesi Dergisi 39(1):35–49
18) Hadden J, Tiwari A, Roy R, Ruta D (2006) Churn prediction: Does technology matter. International Journal of Intelligent Technology 1(2):104–110 123Customer churn prediction system: a machine learning approach
19) Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: State-of the-art and future trends. Computers & Operations Research 34(10):2902–2917
20) Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier, Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., Zeng, J.: Telco churn prediction with big data. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607–618 (2015)
21) Idris, A., Khan, A., Lee, Y.S.: Genetic programming and adaboosting based churn prediction for telecom. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328–1332. IEEE (2012)
22) Kirui, C., Hong, L., Cheruiyot, W., Kirui, H.: Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. International Journal of Computer Science Issues (IJCSI) 10(2 Part 1), 165 (2013)
23) Kisioglu P, Topcu YI (2011) Applying bayesian belief network approach to customer churn analysis: A case study on the telecom industry of turkey. Expert Systems with Applications 38(6):7151–7157
24) Lalwani P, Banka H, Kumar C (2017) Crwo: Clustering and routing in wireless sensor networks using optics inspired optimization. Peer-to-Peer Networking and Applications 10(3):453–471

25) Lalwani, P., Banka, H., Kumar, C.: Gsa-chsr: gravitational search algorithm for cluster head selection and routing in wireless sensor networks. In: Applications of Soft Computing for the Web, pp. 225–252. Springer (2017)
26) Lalwani P, Banka H, Kumar C (2018) Bera: a biogeography-based energy saving routing architecture for wireless sensor networks. Soft Computing 22(5):1651–1667
27) Lejeune MA (2001) Measuring the impact of data mining on churn management. Internet Research
28) Massey AP, Montoya-Weiss MM, Holcom K (2001) Re-engineering the customer relationship: leveraging knowledge assets at ibm. Decision Support Systems 32(2):155–170
29) Musheer RA, Verma C, Srivastava N (2019) Novel machine learning approach for classification of high-dimension al microarray data. Soft Computing 23(24):13409–13421
30) Nath SV, Behara RS (2003) Customer churn analysis in the wireless industry: A data mining approach.
31) Proceedings-annual meeting of the decision sciences institute 561:505–510
32) Petrison LA, Blattberg RC, Wang P (1997) Database marketing: Past, present, and future. Journal of Direct Marketing 11(4):109–125
33) Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A., Rehman, A.: Telecommunication subscribers' churn prediction model using machine learning. In: Eighth International Conference on Digital Information Management (ICDIM 2013), pp. 131–136. IEEE (2013)
34) Radosavljevik D, van der Putten P, Larsen KK (2010) The impact of experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter? Trans. MLDM 3(2):80–99
35) Rajamohamed R, Manokaran J (2018) Improved credit card churn prediction based on rough clustering and supervised learning techniques. Cluster Computing 21(1):65–77
36) Rodan A, Faris H, Alsakran J, Al-Kadi O (2014) A support vector machine approach for chur prediction in telecom industry. International journal on information 17(8):3961–3970
37) Shaaban E, Helmy Y, Khedr A, Nasr M (2012) A proposed churn prediction model. International Journal of Engineering Research and Applications 2(4):693–697
38) Sharma H, Kumar S (2016) A survey on decision tree algorithms of classification in data mining.
39) International Journal of Science and Research (IJSR) 5(4):2094–2097
40) Simons, R.: Siebel systems: Organizing for the customer (2005)
41) Sokolova, M., Japkowicz, N., Szpakowicz, S.: Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: Australasian joint conference on artificial intelligence,pp. 1015–1021. Springer (2006)
42) Tamaddoni Jahromi, A.: Predicting customer churn in telecommunications service providers (2009)
43) Ultsch A (2002) Emergent self-organising feature maps used for prediction and prevention of churn in mobile phone markets. Journal of Targeting, Measurement and Analysis for Marketing 10(4):314–324
44) 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 (IRJET) 4(4):1065–1070
45) Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert systems with applications 23(2):103–11212