Telecom Customer Churn Prediction Using Machine Learning

Pothuraju Raju(✉), Sandam Swathi, Veeravasarapu Keerthi Sumana Sree, Veeramalla Lakshmi Durga, Pula Niharika, and Usarthi Pujitha

Department of CSE, Shri Vishnu Engineering College for Women (A), West Godavari, Bhimavaram, Andhra Pradesh, India
{rajupcse,20b01a05f4,20b01a05i3,20b01a05i2,20b01a05e1, 20b01a05h3}@svecw.edu.in

Abstract. Customer churn or Customer attrition is a serious issue in the telecommunication industry. Today, most of the companies in the telecom sector face this problem that leads them to lose revenue. To be successful in this business, the company has to acquire new customers or retain its users. Acquiring a new customer is significantly more expensive, ranging from five to seven times the cost of retaining an existing one. Retention of the customer is highly required for the company to continue to be successful in this highly competitive sector. As it directly shows an impact on the revenue, telecom companies are in search of new methods and strategies to minimize the switching of customers from one service provider to another and to increase the loyalty of the customers. We proposed a prediction model that helps these companies identify the churn of the customers and the reasons for the churn so that they can come up with new techniques that aid in the retention of the customers. The proposed model which is helpful in the early detection of the customers who are likely to churn uses efficient algorithms in machine learning such as Random Forest (RF) and Support Vector Machine (SVM). The telecom customer churn prediction model uses the most suitable approach and according to our experimental results, the model can achieve 95% accuracy with the Random Forest classifier algorithm.

Keywords: Dark Channel Prior · Dehazing · Image Dimension · Image Quality · Object Detection · Transmission Map · Visibility · Wavelet Transform

1 Introduction

Data has been rapidly increasing over the recent past years because of advancements in the technologies in the telecommunication sector. The technical progress led to a competitive environment which made it hard for telecom companies to gain new customers. Most importantly, for businesses to continue making money, they must keep their clients. The identification of the reasons that made customers leave the company and minimizing them helps in retaining customers. Customer churn is a major concern in the telecommunication service sector. On the other hand, customer churn prediction

P. Pareek et al. (Eds.): IC4S 2024, LNICST 599, pp. 207–220, 2025.
https://doi.org/10.1007/978-3-031-77081-4_17

involves identifying customers who might churn and will improve revenue if addressed early on.

1.1 Customer Churn

Customer Churn [1] is the tendency of customers to abandon a business and stop being clients of a particular business. The churn rate, also referred to as the attrition rate, is the degree to which consumers leave the company over time. Churn refers to the situation where the customer leaves the company and switches to another which can happen for many reasons mainly dissatisfaction with the service provided by the company, high costs, unattractive plans, and poor support. In the telecommunication industry, it refers to the situation where the user terminates their relation with their respective telecom service provider and moves to another provider or drops their services. This can severely affect telecom businesses in terms of revenue as gaining new customers costs high when compared to retaining existing customers. It is critical to determine the causes influencing churn because they have a direct impact on a company's revenue.

1.2 Methods for Handling Customer Churn

It was shown by Dirk Van den Poel and Jonathan Burez that there are two approaches to managing client attrition: proactive and reactive. Reactively, the business offers no retention [2] incentives until the customer requests to terminate their subscription. The business takes a proactive stance by identifying clients who are most likely to go and providing them with exclusive deals to keep them around.

1.3 Machine Learning Approach

Machine learning [3] is highly efficient when it comes to predictions. To perform the analysis [4], it provides the models [5] to automatically learn from past experiences. By using algorithms that learn from data in iterations, machine learning enables systems to uncover hidden patterns without explicit programming to do so. Machine learning consists of three main approaches: unsupervised, semi-supervised, and supervised. Supervised learning involves uncovering hidden patterns from labeled datasets, while unsupervised learning entails finding patterns within unlabelled data [6].

2 Data Set

We downloaded the dataset churnBigML from the Kaggle website. It was imported in two parts.

1) Churn-bigml-80 accounts for 80% of the data used in training.
2) Churn- bigml-20 represents 20% of the data used in testing.

This dataset has 19 attributes.

- The target parameter 'churn' has two values: 'true' and 'false'.

- If the value is accurate, it indicates that the specific customer is a churner—that is, he moves to a different service provider. If the value is false, it indicates a non-churner who stays with the service provider.
- The dataset also includes the 'international plan' and 'voice mail plan'.
- 'Voice mail plan' and 'international plan' are categorical values with values 'yes' or 'no', where 'yes' indicates that the client has subscribed to the specific plan and 'no' indicates that the customer has not subscribed to the plan.
- The 'Customer service calls' function displays the number of calls received from customers regarding services.

We also noticed that, except 'churn', 'international plan', 'voice mail plan', and states, all features in the dataset are numerical.

2.1 Problem Definition

Churn Attrition or Churn rate prediction [7] is essential in the telecom industry as it directly impacts the economic status of the company. If we can predict the churners at an early stage, it will be a great help to telecom companies to retain them. Acquiring a new customer is associated with high expenses. For the retention of customers, it is necessary to predict the churn rate so that the company can analyze the disadvantages underlying their services and identify the factors that make customers churn. To serve this purpose, we are using some of the efficient algorithms in predictions like SVM and Random Forest [8]. Performing required training on the dataset and using this data to be trained by the above algorithms to get accurate results.

2.2 Objective

The primary objective of this paper is an early prediction of churning of the customer i.e. to predict whether a customer is a churner or non-churner using appropriate machine learning techniques [9] to achieve this we used algorithms like Support Vector Machine and Random Forest. Also, we incorporate machine learning methods [10] to provide suitable visualizations that help understand the dataset well.

3 Literature Survey

Siu, N.YM., Zhang [11] proposed an analytical CRM application that employs a simple support vector machine technique. This proposal presents a hybrid approach for extracting interpretable rules from Support Vector Machine (SVM) models in customer relationship management (CRM), utilizing SVM-RFE, SVM modeling, and Naive Bayes Tree; the method is applied to an imbalanced dataset on bank credit card customer churn prediction, demonstrating superior performance and improved rule comprehensibility [12].

[13] Hossain, M. M. & Suchy, N. J. wrote a journal article titled "The Roles of Justice and Customer Satisfaction in Customer Retention". Using a randomly selected group of 200 consumers who experienced [14] poor customer service at Hong Kong

restaurants, they attempted to investigate the cumulative nature of customer satisfaction, highlighting the mediating role of justice in the relationship between before-satisfaction and after-recovery comfort. They also explore how post-recovery satisfaction affects the characteristics of equity in determining customer retention.

[15] Verbraken, Maldonado, and Flores suggested a profit-driven feature selection strategy utilizing SVM. They proposed a profit-driven approach for classifier development and feature selection based on Support Vector Machines, which outperformed traditional strategies in terms of commercial goals. Churn prediction, which is crucial for customer retention, confronts issues in dealing with information overload caused by sociodemographic and behavioural characteristics.

[16] A. Krzyzak and J.X. Dong introduced a fast Support Vector Machine [17] (SVM) training technique, which incorporates kernel caching, and reducing strategies, as well as efficient stopping situations within the SVM decomposition framework, resulting in a 9-fold speed improvement over Keerthi et al.'s enhanced [18] SMO on MNIST. Training 10 one-one-against-the-rest classifiers on MNIST using PCA provides flexibility for a range of engineering applications.

[19] J.C. Platt demonstrated a fast-training strategy for support vector machines using SMO. This study describes a parallel implementation of SMO (Sequential Minimal Optimization) for training SVM using MPI, which reduces computing time for huge problems by partitioning the data and concurrently processing subsets, resulting in considerable speedups on datasets such as adult, MNIST, and Web data [20].

[21] Michael Haenlein proposed SVM-based attrition prediction in the telecom sector. In response to increased global competition, customer attrition, which is notably high in the telecommunications sector at 30%, necessitates the use of predictive models. This work presents an advanced methodology for predicting churn in the mobile telecommunications industry using Support Vector Machines with four kernel functions. Model performance is evaluated using a gain measure on a dataset having 21 attributes for each of its 3333 entries.

[22] Benlan suggested a Support Vector model-based approach for predicting customer churn and used a random sampling technique to improve [23] the SVM model by addressing the uneven nature of customer data. An SVM builds a hyperplane in an n-feature space that can be used for classification. The sampling approach [24] can be used to change the distribution of the data and alleviate the imbalance caused by the low number of churners.

[19] To address data distribution issues, Y. Xie and colleagues employed an extended balanced random forest (IBFR) model that incorporated both balanced randomly produced forests and scaled random forests. IBRF includes iteratively learning the best characteristics by changing the class distribution and imposing larger penalties for misclassifying minority classes. The study used a Chinese bank dataset and found that IBRF beat decision trees and support vector machines in terms of accuracy.

4 Working Model

4.1 Data Collection

The first step of working is to collect data that is suitable for analysis. We used the ChurnBigMl a set containing 3335 rows and 19 columns. This dataset contains prameters like state, account length, customer service calls, etc.

4.2 Data Preprocessing

We proceeded to the preparation step after thoroughly understanding the data. Preprocessing is an important first step in any machine-learning project since it prepares data for future machine-learning models. This preprocessing step in our model includes two steps.

4.2.1 Cleaning of Data

The data cleaning process involves handling missing values and null values. However, our chosen data contains no missing or null values. So, we are not removing any rows in our dataset.

4.2.2 Encoding of Categorical Values

As mentioned above, our data contains an international plan, voice mail plan, and churn having categorical values in the form of either yes/no or true/false. We used a label encoder from Scikit-learn to serve the purpose of transforming the categorical values into numerical values for further analysis. The following are the observations from this preprocessing step.

- Firstly, the 'voice mail plan' and 'international plan' having yes/no values are converted into 1/0 respectively using a label encoder.
- Using a label encoder, the 'state' column is transformed into distinct numerical values.
- Then, 'churn' has values as 'true' and 'false' converted into 0 and 1 respectively.

4.3 Feature Selection

After performing the preprocessing step, now we are finally able to find the relation among multiple features that are there in our dataset. In this step, we find the important features that can affect the target variable churn and this step helps us drop some columns that are not related to churn. We use bar plot graphs to find the relation existing between the churn and other parameters in our dataset.

The relation between the International Plan and Churn (see Fig. 1), most users have not subscribed to the international plan; yet, those who have are less likely to leave the service.

The relation between Voicemail Plans and Churn (see Fig. 2), most customers have not subscribed to the voicemail plan, but those who have are less likely to discontinue service.

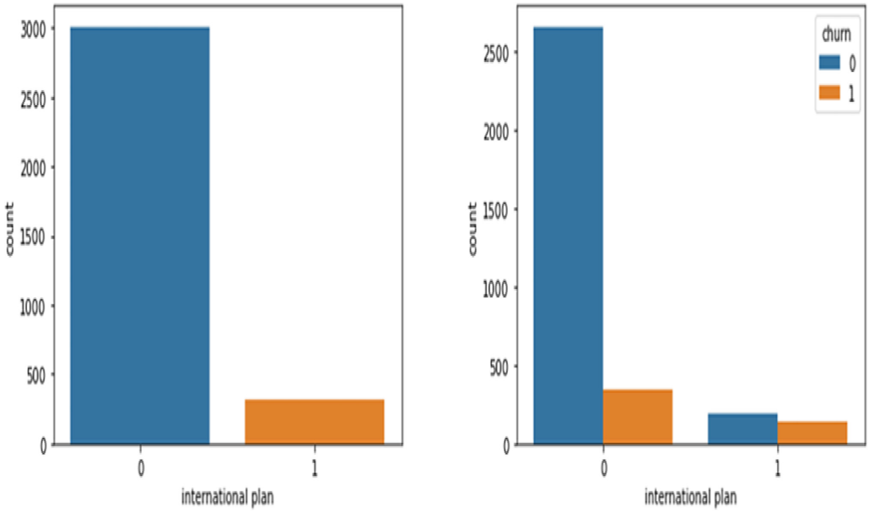

Fig. 1. Relation Between International Plan and Churn

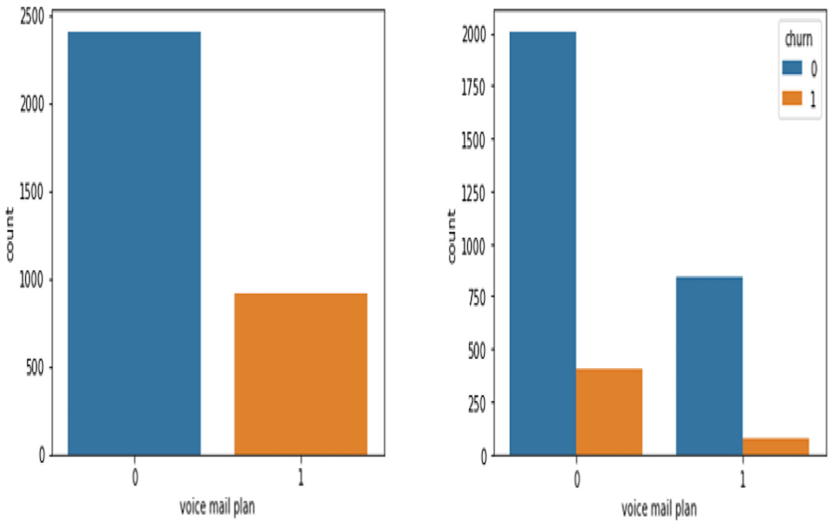

Fig. 2. Relation Between Voicemail Plans and Churn

The relation between Area Code and Churn (see Fig. 3), the second area has the largest user count. While there are many churners in this region, it cannot be definitively claimed that the area code influences churn behavior, as the increased churn rate could be attributed to the larger number of users compared to other areas. Customer service calls' is an influential feature among all the features in the dataset as it affects attrition in a significant way.

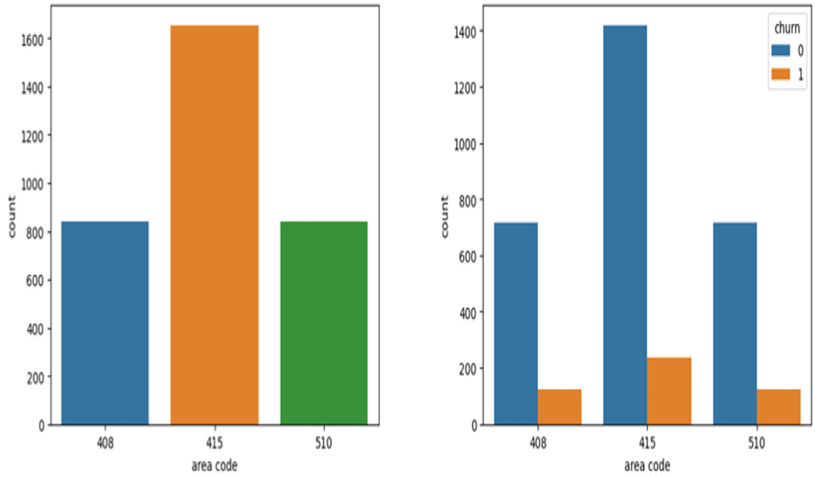

Fig. 3. Relation Between Area Code and Churn

Below (see Fig. 4) is a very high digit of users who have communicated with the service provider's customer service at least once. Customers who have only been approached once are much more likely to remain loyal to their telecom provider. We can conclude that customer support calls are directly related to the desired churn feature.

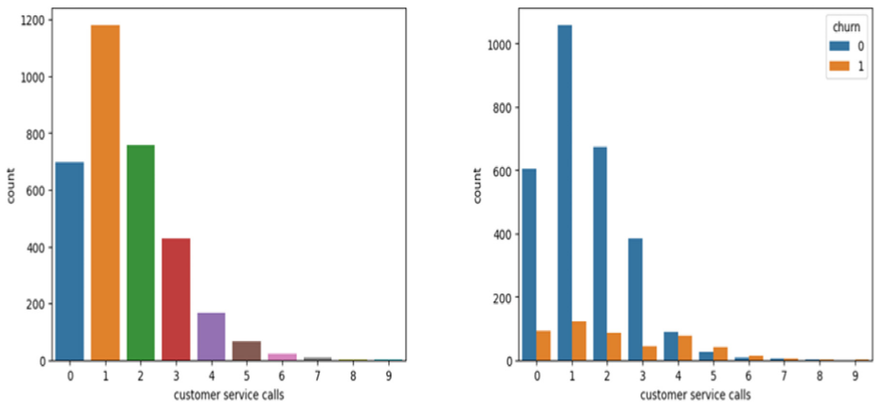

Fig. 4. Relation Between Customer Service Calls and Churn

The Correlation Matrix (see Fig. 5) shows how each attribute relates to the churn variable. Based on the matrix, we can conclude that customer service calls, international plans, total day charges, and total day minutes are strongly connected with turnover and have an impact on it. Total overseas calls, amount of voicemail messages, and voicemail have a lower correlation with churn.

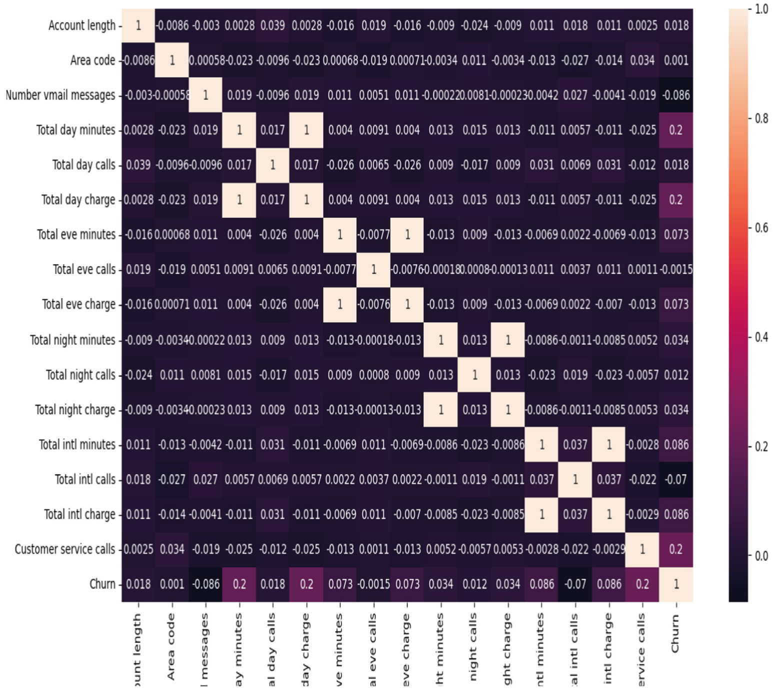

Fig. 5. Correlation Matrix

4.4 Model Selection and Development

After analyzing our data, we determined that some characteristics, such as state, area code, and account length, should be removed because they have a low impact on churn and add little value to the churn. We will go on to the rest of our data when we have dropped the columns. Moving on to the next phase, model creation, which comprises defining the model and selecting the necessary algorithms to build it. As previously noted, we used 80% of the data for training and 20% for testing. We train the model on 80% of the data using the machine learning methods Random Forest (RF) and Support Vector Machine (SVM).

4.4.1 Support Vector Machine (SVM)

SVM [25] works by generating a dimensional space with several dimensions from the input provided to it. The hyperplane that increases the distance between the various classes is then identified. The positions closest to the optimal boundary—the support vectors—are critical to the model's development. The decision boundaries, or support vectors, define the hyperplane itself. The margin is the distance between the nearest data points or support vectors in each class and the hyperplane. A higher margin suggests improved model performance. To use this algorithm, we imported it from sklearn.ensemble. While working with this approach to develop the model, we utilize random state = 10.

4.4.2 Random Forest Classifier (RFC)

Random Forest Classifier is an ensemble technique that generates several new datasets from our original dataset and trains a decision tree on each bootstrapped dataset independently. Each tree is based on a randomly selected subset of features. Random Forest Classifier is a good algorithm in terms of accuracy. The difficulties associated with decision trees, such as excessive variance and overfitting, can be considerably mitigated by utilizing a random forest classifier. RFC decreases variance by using random data sampling, which entails creating several random samples of data for training and using random feature subsets. Instead of employing a single tree, RFC employs many trees to avoid the overfitting problem. To use this algorithm, we imported it from sklearn.ensemble. While working with this algorithm, we generated ten random trees with max_depth=4, n_estimators=6, and random_state=0, using customer service calls, international plans, total day minutes, and so on as root nodes for random decision tree nodes.

4.5 Model Evaluation

After creating the models, it is critical to evaluate their efficiency. To do this, we employed the evaluation metric 'accuracy'. Accuracy is the most used criterion for assessing model efficiency. After applying this statistic to our SVM and Random Forest models, we discovered that both achieved an accuracy of 85% and 95%, respectively. So we found that in our project study, Random Forest outperforms SVM and delivers more accurate forecasts. The statistics below depict the accuracy we achieved after applying accuracy.

5 Implementation

After learning about the project, we are ready for the implementation, which includes installing an IDE like Visual Studio Code to work with, importing required libraries and software, performing required actions discussed in the project, and finally developing a useful interface that a user will find useful to work with.

- Install Visual Studio Code and get Python extensions to run the project.
- Importing fundamental packages and libraries (e.g., numpy, pandas, matplotlib, seaborn) is necessary for reading CSV files, running calculations, and creating visualizations.
- Importing needed algorithms: After preprocessing, we train models using training data using SVM and RF techniques. After comparing the models using assessment measures, we select the best model in terms of accuracy. We compare different models, Random Forest and SVM, to get maximum accuracy.

- Using the Flask Framework: Flask is a small and lightweight Python framework that provides useful features and tools for the creation of Python web applications. Because training a machine learning [19] model requires computational resources, we do not want our model to repeat the training process. Python's pickle module allows us to serialize and deserialize objects so that we can save our classifier in its current state and reload it whenever we need to categorize new data without having to retrain the model using the training data.
- To work with the framework, we must import Flask and Pickle into app.py. Pickle.load() serializes our object, then pickle.dump() deserializes it. After saving this application as app.py, we can either type 'python app.py' in the terminal or launch the.py file directly in the Visual Studio Code editor by selecting the 'launch without Debugging' option. Then a URL will be displayed; copy it and paste it into any browser to view the app in action.

6 Results

This section exhibits the project's web application results displayed below (see Fig. 6) and (see Fig. 7) displayed below. First and foremost, our project's primary purpose is to predict client attrition. To serve this, when we pass some parameters to the application's interface, it will process them and display the predicted results. The prediction results are based on the model trained with the best techniques available. We employed the SVM and RF algorithms, and the best one was declared as the RF in terms of accuracy.

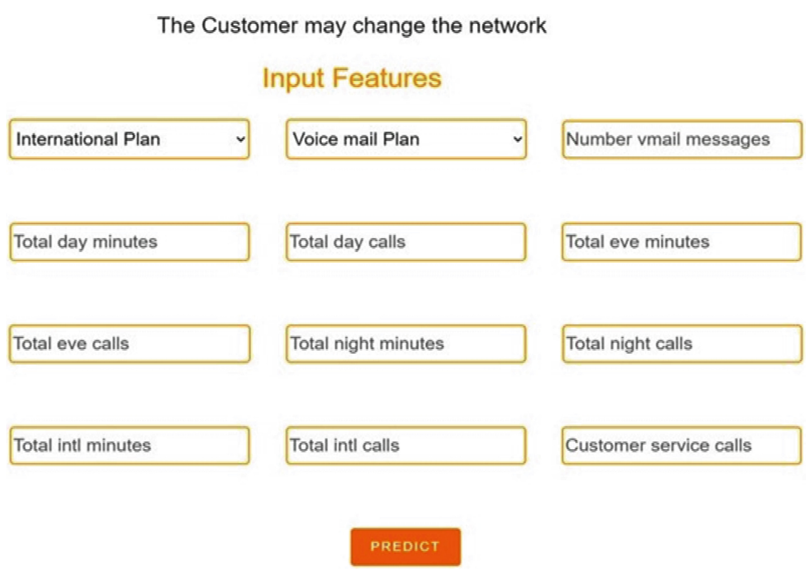

Fig. 6. Customer Changes in the network based on the input features

Two distinct algorithms Random Forest depict trumps SVM producing more accurate predictions. The accuracy of SVM is 85.90 (see Fig. 8), whereas the accuracy of Random Forest is 94.75 (see Fig. 9).

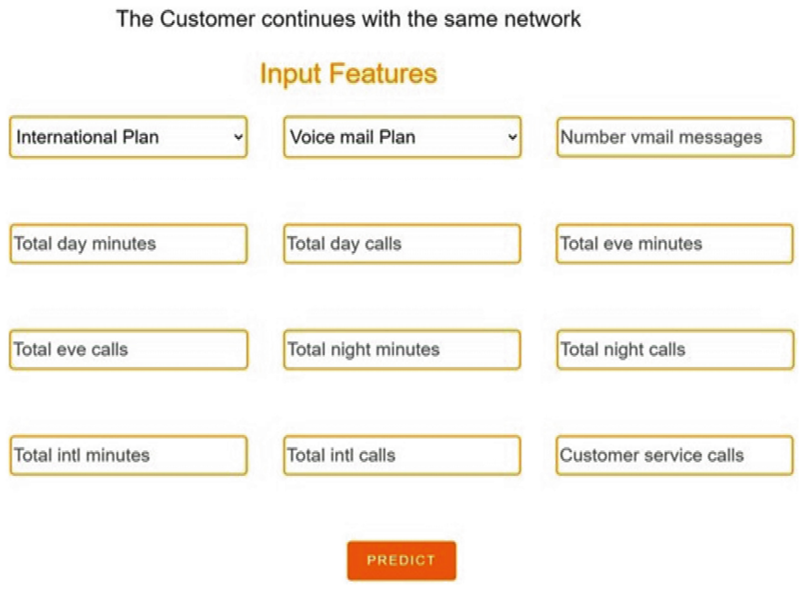

Fig. 7. Customer Remains in the network based on the input features

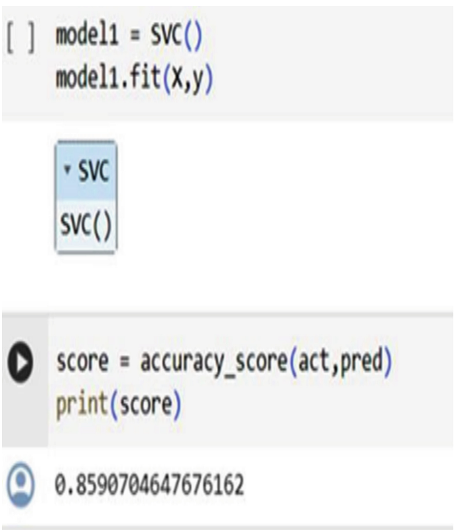

Fig. 8. Accuracy attained with SVM.

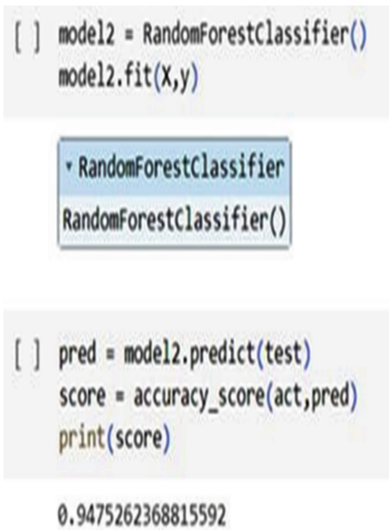

Fig. 9. Accuracy attained with RF

7 Conclusion and Future Work

The primary goal of this Telecom Customer Churn Model is to detect churn among their subscribers in advance using effective Machine Learning techniques. The proposed approach shows how to use machine learning algorithms such as SVM and Random Forest (RF). Using these approaches, we discovered that the models trained on them are efficient in predicting churners and non-churners, with SVM accuracy of 85% and RF of 95%. According to our experimental results, the Random Forest algorithm is effective at predicting churners with an outstanding accuracy of 95%. The significance of this study resides in its evaluation of multiple machine learning algorithms to determine their efficacy in churn prediction in the telecom business. These findings give telecom firms valuable insights into the most successful algorithms for predicting customer turnover, influencing their decision-making processes when developing retention tactics. Telecommunications firms can use this model to devise strategies that help them retain consumers while also increasing revenues.

To improve the prediction model's accuracy, the proposed methodology suggests adding new variables and employing feature selection methods. Improving the dataset's balance may also benefit the model's overall performance.

References

1. Jadhav, R.J.: Churn prediction in telecommunication using data mining technology. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **2**(2) (2011). https://doi.org/10.14569/IJACSA.2011.020204
2. Al-Hawari: The relationship between service quality and retention within the automated and traditional contexts of retail banking. J. Serv. Manag. **20**(4), 455–472 (2009). https://doi.org/10.1108/09564230910978539
3. Rao, A.P., Bhaskar, J., Kumar, G.P.: Machine learning framework for identification of abnormal EEG signal. In: Gupta, N., Pareek, P., Reis, M. (eds.) IC4S 2022. LNICSSITE, vol. 472, pp. 42–54. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28975-0_4
4. Phadke, C.: Prediction of subscriber churn using social network analysis. Bell Labs Tech. J. **17**(4), 63–76 (2013). https://doi.org/10.1002/bltj.21575
5. Sree, P.K.: Review on healthcare quality using machine learning methods. In: Khang, A. (ed.) AI and IoT-Based Technologies for Precision Medicine, Hershey, PA, pp. 413–430. IGI Global (2023). https://doi.org/10.4018/979-8-3693-0876-9.ch024
6. Maddula, P., Srikanth, P.: COVID-19 prediction with Chest X-Ray images using CNN. In: International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), pp. 568–572 (2023). https://doi.org/10.1109/IITCEE57236.2023.10090951
7. Rodan, A.: A support vector machine approach for churn prediction in telecom industry. Int. J. Inf. **17**(8), 3961–3970 (2014)
8. Raju, P.: Dynamic load balancing framework for context sensitive offloading scheme in mobile cloud computing. In: International Conference on Advanced Computing & Communication Technologies (ICACCTech), Banur, India, pp. 142–149 (2023). https://doi.org/10.1109/ICACCTech61146.2023.00031
9. Mallesh, A.S., Murty, P.T.S.: Smart system for early detection of agricultural plant diseases in the vegetation period. In: 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI), Wardha, India, pp. 1–6 (2023). https://doi.org/10.1109/IDICAIEI58380.2023.10406672

10. Pamarthi, N., Satyanarayana Murty, P.T.: A research study of heart health monitoring using deep learning and IoT. In: 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI), Wardha, India, pp. 1–6 (2023). https://doi.org/10.1109/IDICAIEI58380.2023.10406326
11. Siu, N.YM.: The roles of justice and customer satisfaction in customer retention: a lesson from service recovery. J. Bus. Ethics **114**, 675–686 (2013). https://doi.org/10.1007/s10551-013-1713-3
12. Satyanarayana Murty, P.T., Prasad, M., Raja Rao, P.B.V., Kiran Sree, P., Ramesh Babu, G., Phaneendra Varma, C.: A hybrid intelligent cryptography algorithm for distributed big data storage in cloud computing security. In: Morusupalli, R., Dandibhotla, T.S., Atluri, V.V., Windridge, D., Lingras, P., Komati, V.R. (eds.) MIWAI 2023. LNCS, vol. 14078, pp. 637–648. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-36402-0_59
13. Hossain, M.M.: Influence of customer satisfaction on loyalty: a study on mobile telecommunication industry. J. Soc. Sci. **9**(2), 73–80 (2013). https://doi.org/10.3844/jssp.2013.73.80
14. Boddapati, M.S.D., Desamsetti, S.A., Adina, K., Uppalapati, P.J., Murty, P.T.S., RajaRao, P.B.V.: Creating a protected virtual learning space: a comprehensive strategy for security and user experience in online education. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (eds.) IC4S 2023, LNICSSITE, vol. 536, pp. 350–361. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-48888-7_30
15. Maldonado, S.: Profit-based feature selection using support vector machines – general framework and an application for customer retention. Appl. Soft Comput. **35**, 740–748 (2015). ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2015.05.058
16. Maga, M.: Accenture Global Services Ltd, assignee. Churn prediction and management system. United States patent US 8,712,828 (2014)
17. Surbakti, H., Kumar, A.N., Tara, S., Keshta, I., khare, A., Prasad, G.N.R.: Support vector machine for multiclass classification of redundant instances. In: Balas, V.E., Semwal, V.B., Khandare, A. (eds.) IC-ICN 2023. LNNS, vol. 699, pp. 413–423. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3177-4_30
18. Raja Rao, P.B.V., Prasad, M., Kiran Sree, P., Venkata Ramana, C., Satyanarayana Murty, P.T.: Enhancing the MANET AODV forecast of a broken link with LBP. In: Reddy, V.S., Prasad, V.K., Wang, J., Rao Dasari, N.M. (eds.) ICISSC 2022. SIST, vol. 363, pp. 51–69. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-4717-1_6
19. Vafeiadis, T.: A comparison of machine learning techniques for customer churn prediction. Simul. Model. Pract. Theory **55**, 1–9 (2015). ISSN 1569-190X. https://doi.org/10.1016/j.simpat.2015.03.003
20. Mangalampalli, S., Pokkuluri, K.S., Raju, P., Shalem Raju, P.J.R., Usha Devi N, S.S.S.N., Mangalampalli, V.K.: Energy efficient VM consolidation technique in cloud computing using cat swarm optimization. In: Skala, V., Singh, T.P., Choudhury, T., Tomar, R., Abul Bashar, M. (eds.) Machine Intelligence and Data Science Applications. LNDECT, vol. 132, pp. 457–467. Springer, Singapore (2022). https://doi.org/10.1007/978-981-19-2347-0_36
21. Haenlein, M.: Social interactions in customer churn decisions: the impact of relationship directionality. Int. J. Res. Mark. **30**(3), 236–248 (2013). ISSN 0167-8116. https://doi.org/10.1016/j.ijresmar.2013.03.003
22. Farquad, M.A.H.: Churn prediction using comprehensible support vector machine: an analytical CRM application. Appl. Soft Comput. **19**, 31–40 (2014). ISSN 1568-4946. https://doi.org/10.1016/j.asoc.2014.01.031
23. Pitchai, R.: MRI image analysis for cerebrum tumor detection and feature extraction using 2D U-ConvNet and SVM classification. Pers. Ubiquit. Comput. **27**, 931–940 (2023). https://doi.org/10.1007/s00779-022-01676-y

24. Rosenberg, L.J.: A marketing approach for customer retention. J. Consum. Mark. **1**(2), 45–51 (1984). https://doi.org/10.1108/eb008094
25. Brandusoiu: Churn prediction in the telecommunications sector using support vector machines (2013). https://doi.org/10.15660/AUOFMTE.2013-1.2772