International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Customer Churn Prediction

Dr. D. Esther Rani [1], Chaitanya T [2], Jasmine Sk [3], Maheshwari T [4], Sai Geethika V[5]

[1]Associate Professor, Department of CSE, N.B.K.R. Institute of Science & Technology Vidyanagar, Tirupati, Andhra Pradesh, India

[2,3,4,5]B.Tech Students, Department of CSE, N.B.K.R. Institute of Science & Technology Vidyanagar, Tirupati, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 01 June 2023

Published: 09 June 2023

Publication Issue

Volume 9, Issue 3

May-June-2023

Page Number

398-406

ABSTRACT

Customers are becoming more drawn to the standard of service (QoS) offered by businesses in the present. However, the present day shows greater rivalry in offering clients technologically cutting-edge QoS. However, efficient communication systems may help the organization attract new clients, preserve client connections, and enhance client retention by generating more revenue for the company's operations. Additionally, the client retention methods can benefit greatly from the use of machine learning models like support vector machines and Random Forest algorithms.

Keywords : Supervised learning, Machine Learning, Random Forest Algorithm and Support Vector Machines.

I. INTRODUCTION

Customers always play a crucial role in boosting the revenue and profit of any organization; as a result, it is crucial for managers at organizations to maintain an effective relationship with their customers by identifying their target clients and building strong relationships with them. The firm will also benefit from the CRM system's assistance with finding the most significant collection of consumers and their behavior, which will help it better understand its retention efforts. Additionally, lower customer turnover rates are associated with stronger customer loyalty, therefore applying machine learning algorithms like the support vector algorithm can help avoid customer churn. The use of supported vector machine learning to increase customer loyalty and retention will be the main topic of this paper.

The income a firm receives for a certain service or product is adversely affected by the effect of clients who left an organization in the past to those who are prepared to continue an exchange with it. Customers may have unrealistic expectations that are not met when they buy items from a certain firm, as seen by their preference for newer manufacturers. For corporations who manufacture machinery, it is difficult to persuade customers about a certain machine when they test it and discover it underperformed. If the device meets further requirements, the consumer will choose a different seller. Consumer management tactics fail, particularly under these circumstances, causing the company to increase the quality of their products and expand their brand portfolio.

Customer satisfaction has been seen as a key element of sustaining long-term customer connections in the relationship marketing literature. Therefore, when clients encounter a service failure, figuring out how to boost customer satisfaction and retain unhappy customers becomes a huge issue. Prior studies have stressed the importance of justice and post-recovery customer happiness in resolving this issue. For customers to believe that their unmet requirements have been addressed, they must think that the outcomes are correct or fair (Kau and Loh, 2006). According to Hoffman and Kelley (2000), each of the following elements—the recovery of service process itself, the outcomes, the interpersonal behaviors demonstrated through the method, and the delivery of the results—is essential. Their stance is in line with the three-dimensional definition of justice proposed by Tax et al. in 1998 (distributive, procedural in nature and interactional justice). But the bulk of studies use a static perspective. It is well acknowledged that firms must continuously increase customer pleasure to retain customers since it accumulates over time.

In this study, the idea of customer happiness is divided into two post-recovery components (i.e., contentment with recovery and comfort with organization) and previous satisfaction (before service failure). The first of this study's two main goals is to ascertain whether complaint justice mediates the link within prior fulfillment and post-recovery satisfaction (from with the organization's recovery and with the recovery process), and the second is to ascertain whether post-recovery satisfaction mediates the connections across the various aspects of complaint justice along with customer retention. It is said that mending the relationship and recovering consumers' faith in companies depends on views of complaint fairness. In the overall scheme of performance recovery, it helps a business capitalize on its prior customer satisfaction. Meanwhile, we suggest that two mechanisms—satisfaction with recuperation and contentment with organization—transform perceived fairness into a behavioral desire. Organizational satisfaction, however, may be more crucial in this process.

It also makes it possible to read classifiers more accurately, which is essential in business analytics. Many machine learning technologies are frequently referred to by practitioners and "black boxes," which discourages them from using the related techniques. For analytics in business decision-making, a greater understanding of the data creation process is therefore necessary, for example by identifying the aspects that enable the consumer decision-making process to be explained. In the past, statistically inspired techniques were the most often used approaches for testing classification and feature selection techniques. Recently, the use of profit-based criteria for classifier validation has been suggested. By creating multiple embedded strategies that include the Holdout supports vector machine (HOSVM) technique with a range of validation measures, we expand the idea of profit-driven metrics and implement it to the issue of feature selection in this work.

To the best of our knowledge, the special issue of driven by profits feature selection has not yet been addressed in the data analysis and machine learning literature. The majority of work in feature selection and business analytics employs conventional, statistically based approaches without taking considerations about financial gain into account. Our studies show that the proposed methods outperform rival tactics and provide classifiers with highly relevant features, reducing the risk of the overfitting while likewise boosting the related profit.

II. RELATED WORKS

[1] An example from service recovery on the roles of injustice and client fulfillment in customer retention Customers complain: when a service malfunction occurs because they're expecting the company to take care of them fairly. The relationship between perceived complaint fairness and consumer pleasure

has been investigated. However, a static viewpoint was employed in the majority of prior studies. Since satisfaction builds up over time, we suggest that it is critical to include both prior and post-recovery contentment when assessing claim fairness in the framework of service recovery. This study intends to fill the gap by examining the role of justice as a mediator in the relationship between prior happiness and post-recovery contentment (both with the process of recovery and for the organization), and to investigate the connection between the various elements of fair and customer retention. Hypotheses were tested using a group of 200 customers who had poor service at Chinese restaurants in the city of Hong Kong. The justice components of justice by distribution, justice by procedure, and interactional justice were shown to be the only factors that moderated the relationship between prior happiness and satisfaction with recovery. All traits, overall, the exception of interconnected fairness, were demonstrated to be partially mediating in the connection between prior happiness and post-recovery pleasure with organization. The findings also demonstrated that two post-recovery contentment variables, with essentially opposite roles, mediated the transformation of the justice characteristics into behavioral intention. Discussions are had on how to improve the process of building enduring relationships with clients, and recommendations are given.

[2] Profit based feature selection using support vector machines–General framework and an application for customer retention: Churn prediction, a critical application of classification models, identifies the clients who are most likely to attrite upon characteristics indicated by, for instance, socio-demographic and behavioral parameters. Because these variables are now more often being captured and retained in the pertinent computational systems, an efficient management of the resulting data overload becomes a very major challenge for building customer retention systems that utilize churn prediction models. Therefore, selecting features is an essential step in creating the right classifier. However, the bulk of feature selection techniques rely on statistically driven validated criteria, which don't necessarily provide models that are the most effective at achieving the goals stated by the organization that applies. In this study, we provide a profit-driven method for concurrent variable selection and SVM-based classifier building. Experimental results show that our simulations outperform conventional feature selection methods, producing better commercial success. Coordinated goals On February 1st, 2015, a preprint had been submitted to Computational Soft Computing.

[3] Churn prediction using comprehensible support vector machine: An analytical CRM application: Support vector machines (SVM) are currently state-of-the-art for classification issues due to their ability to model nonlinearities. SVM's primary drawback is that it develops "black box" models, meaning that it does not communicate the knowledge acquired throughout training in a manner that is intelligible to humans. Such opaque models are converted into transparent models through the process of rule extraction. In this study, a hybrid approach for deriving rules employing SVM for applications in CRM was proposed. The recommended hybrid method is composed of three phases. In the first stage, the feature set is reduced using SVM-RFE (SVM-recursive feature elimination). (ii) In the second stage, support vectors are extracted and an SVM model is built using a collection of data with less attributes. The Naively Bayes Tree (NB Tree) is employed to generate rules in the final stage. The forecasting of churn among bank debit card users was examined using the company's Business Informatics Cup 2004 dataset, and had a very unequal customer retention percentage of 93.24% and a turnover rate of 6.76%. We also employed a number of standard balancing techniques to reconcile the data and generated rules. The results of the empirical study demonstrate that the proposed hybrid outperformed all other strategies considered. It should be noted that the recommended method extracts rules of shorter length when the smaller feature dataset is used, which

improves the system's comprehension. The developed guidelines act as an early warning system for the bank management.

[4] Churn prediction in the telecommunications sector using support vector machines: Due to challenges imposed on by global competition, customer churn is one of the biggest problems for organizations in all industries today. With an annual turnover of 30%, the telecoms sector takes the top place on the ranking. To solve this problem, predictive algorithms must be utilized to identify clients who are most likely to depart. In order to estimate client turnover in the wireless communications industry, this study proposes a complex methodology. The dataset's 3333 entries for call information records each contain 21 features. We use a machine learning method featuring four kernel functions to build the prediction models. Gain metric is used to evaluate and contrast the performance of the models.

[5] Fast training of support vector machines using sequential minimal optimization: Even though sequencing maximum optimization (SMO) is a popular method for training neural networks using support vector machines (SVM), solving issues of a large scale still takes a lot of time. For SVM training, this study recommends using an individual parallel SMO implementation. The parallel SMO (MPI) is built using a message passing interface. More specifically, the parallel SMO splits the entire training information set into smaller segments before handling each of the split data sets with several CPU processors operating concurrently. Trials on the grown-up data set as well as the Merging the National Institutes of Science and Engineering (MNIST) data set show a considerable speedup when many processors are employed. Positive results are also included in the Web data collection.

[6] A fast SVM training algorithm: A rapid SVM (support vector machine) training strategy that successfully integrates kernel cached data, digested and atrophy rules, and stopping conditions is presented using the broken-down mechanism of the SVM technique. A number of experiments utilizing the MNIST cursive digit database are being conducted to show that the recommended method operates approximately 9 times faster than SMO developed by Keerthi et al. Principal component analysis and 10 one against the rest classifiers were trained on MNIST in about 0.77 hours total. The proposed scheme's projected scalability allows SVM to be used to a variety of engineering problems.

III. METHODOLOGY

Proposed system:

We employ machine learning methods in the proposed system to draw conclusions from the complex trends in the data. This method's straightforward architecture makes computations with it cheap.

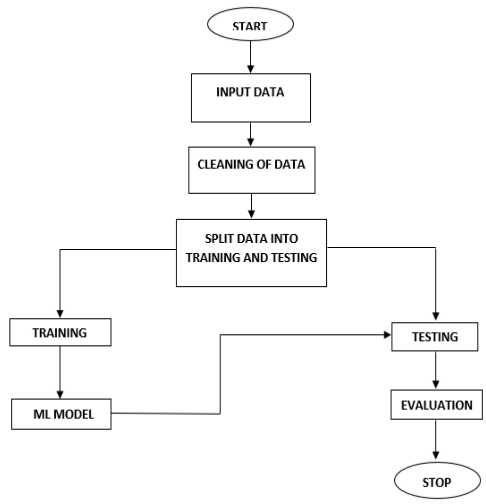


Figure 1: Block diagram

IV. IMPLEMENTATION

The project was carried out using the algorithm given below.

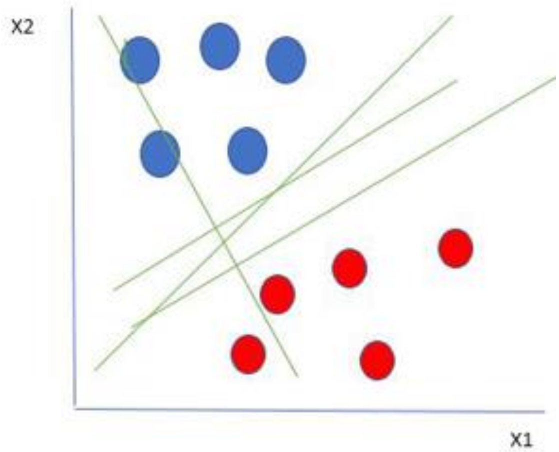
ALGORITHM:

SUPPORT VECTOR MACHINEs

A unsupervised machine learning approach called a Support Vector Machine, or SVM, is utilized for regression and classification. Although we often refer to regression concerns, categorization is the most appropriate term. Finding a hyper plane that exists in a space with dimensions N that readily separates the data points is the goal of the SVM method. The total number of features determines the hyperplane's size.

The hemi plane is only a line if there are just two input characteristics. The hyper plane turns into a 2-D plane if there are three input characteristics. Imagining something with more than three characteristics gets challenging.

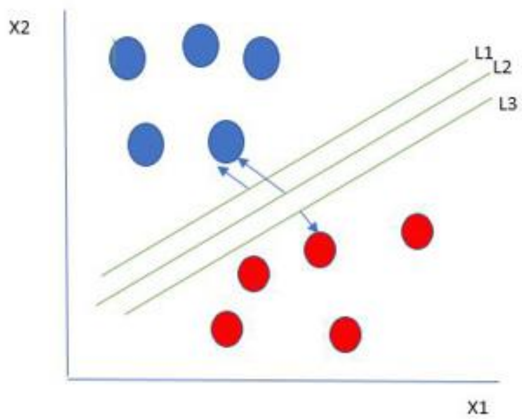
Let's look at two independent variables (x1, x2) and one dependent variable (either a blue circle or a red circle), which is the dependent variable. varieties of SVM.



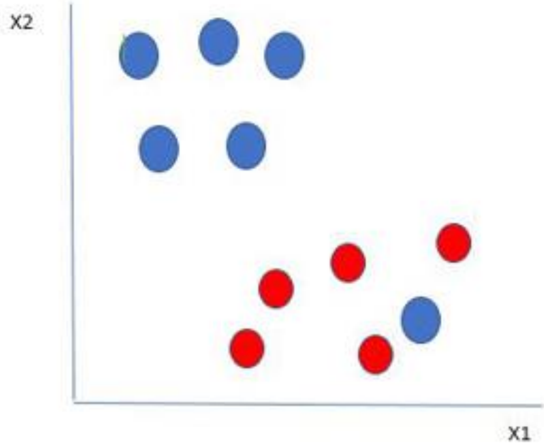
It is extremely obvious that there are several lines that separate each of our points or perform a categorization amongst red and blue spheres (our hyperplane in this case is a line as we are just taking into account two input characteristics, x1, x2). How do we select the best line, or best hyper plane in general, to separate our data points?

Selecting the best hyper-plane:

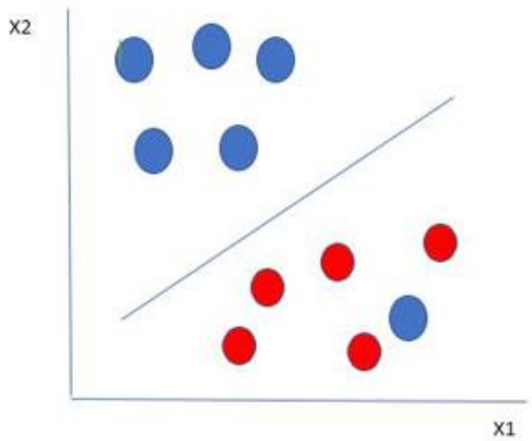
The hyperplane that indicates the greatest gap or margins amongst the two classes is a logical option for the best hyperplane.



Therefore, we select the hyper plane with the maximum distance between it to the closest data point on either side. The maximum-margin hypo plane/hard margin is referred to if precisely a hyper plane is present. Therefore, we select L2 from the given diagram.

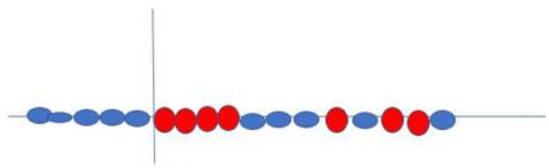


One blue ball is now within the red ball's perimeter. How does SVM categorize the data then? It's easy! An anomaly of blue balls is the blue ball that lies on the edge of the red ones. When searching for the ideal hyperplane to maximize the margin, the algorithm used by SVM has the ability to disregard outliers. SVM can withstand outliers.

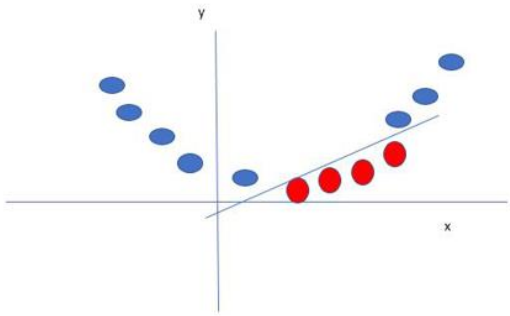


SVM determines the maximum margin for this particular kind of data like it did for earlier data sets and also applies an extra charge each time it crosses the limit. Therefore, in circumstances like this, the margins are referred to be soft margins. The SVM seeks to reduce 1/margin+(penalty) when the data set has a soft margin. Losing the hinge is a frequent type of punishment. No loss of hinge if no infractions. If infractions occur, there will be a loss based on the violation's length.

Up until this point, we have only discussed data that is linearly separable (the group of blue and red balls may be separated by a straight line or linear line). What should one do if data cannot be separated linearly?



Let's say that our data looks like the figure above. By employing a kernel to create a new variable, SVM resolves this. We designate an entirely novel variable yi as an indicator of the distance from the origin o. As a result, if we plot this, we obtain the result that is depicted below.



In this instance, the amount of time from the origin is used to construct the new variable y. Kernel is the name for a function that's not linear that produces a new variable.

SVM Kernel:

The SVM kernel is a function that converts non separable problems into separable problems by taking low-dimensional input space and transforming it into higher-dimensional space. It works best in non-linear separation issues. Simply explained, the kernel determines how to split the data depending on the outputs or labels that have been defined after performing some incredibly sophisticated data transformations.

Advantages of SVM:

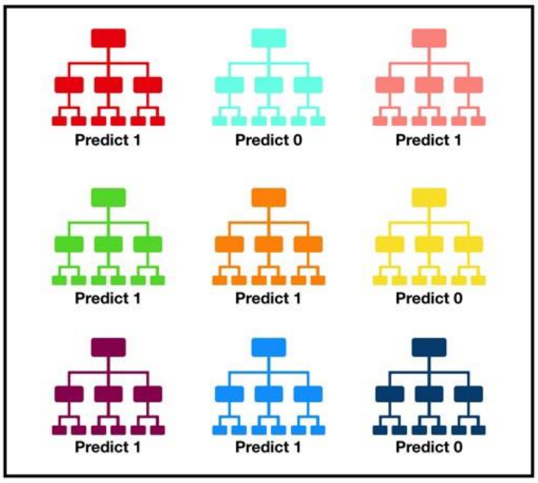
• Useful in instances with high dimensions

• Different kernel functions may be supplied for the decision functions, and it is possible to supply custom kernels.

• It is memory-efficient because it employs a subset for training points associated with the decision function that are called support vectors.

2. Random Forest:

A random forest, as its name indicates, is a collection of various autonomous decision trees that function as an ensemble. The categorization that obtains the most votes from the random forest's trees becomes the prediction produced by our model (see the graphic below).



Tally: Six 1s and Three 0s
Prediction: 1

Creating a Prediction Using the Random Forest Visualization

The fundamental principle of random forest is the wisdom of crowds, which is a simple yet powerful concept. In the context of data science, the random forest data collection model works so well because:

Working together, several usually uncorrelated models (trees) will perform better than any one of the individual component models.

The weak correlation between models is the key. Similar to how investments with low correlations (like stocks and bonds) collaborate to create an investment strategy that is more than the sum of its parts, indistinguishable models are capable of offering ensemble projections that are considerably more exact than any of each of the projections.

This wonderful effect is created when the trees protect one another from their own errors, provided that they don't all regularly make the same faults. A collection of trees will move in the appropriate direction since many of them will be correct but others may be mistaken. Therefore, for the random forest system to function effectively:

1. For the reason for models built using our features to perform better than guesses, they must contain some genuine signal.

Low correlation between the estimates (and therefore the errors) made by each tree are required.

It's crucial to comprehend the incredible effects of using several uncorrelated models, therefore allow me to illustrate with an example. Let's say we're playing the following game:

I use a randomly distributed number generator to generate a number; if the result is in excess of or equal to 40, you win (you have a 60% chance of winning), and I pay you some money; if it is less than 40, I win, and you give me the same amount.

• I'm going to give you the following choices right now. Options include:

s**1. Game 1: Play 100 times, betting $1 each.**

2. Game 2: Play 10 times with a $10 wager each.

3. Game 3: Place a single $100 wager.

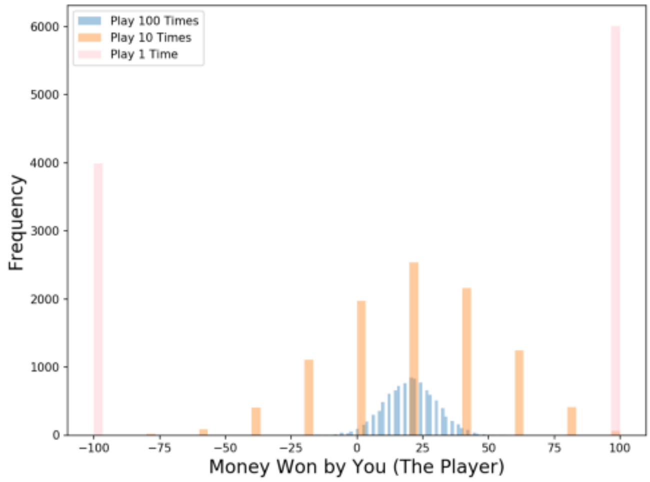
Which do you like best? The expected value for each game is the same:

Game 1's expected value is equal to (0.60*1 + 0.40*-1).*100 = 20

Game 2's expected value is equal to (0.60*10 + 0.40*-10)*10 = 20

Expected Value for Game 3 is 20 (or 0.60 x 100 + 0.40 x -100).

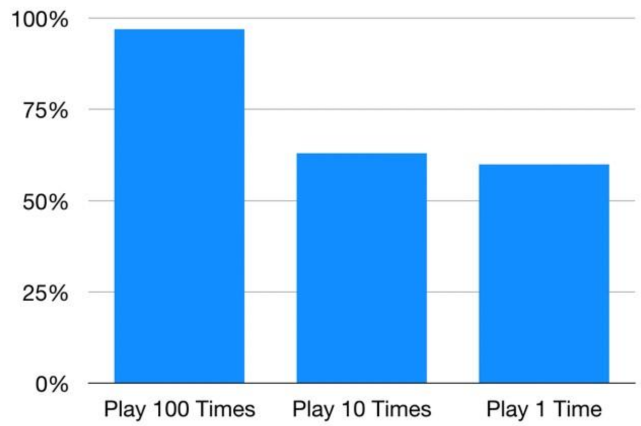
ssssssssssss



Breakdown of the Results for Each Game from 10,000 Simulations

What about the distributions? Let's conduct a simulation using Monte Carlo methods to determine the results (for example, let are going to replicate a thousand of each of the 100 matches of Game 1 for each type of game). Using the diagram on the left as a guide, which of the following would you pick? identical even when anticipated values, The result distributions considerably diverge (pink) when they go between optimistic and narrower (blue) to binary.

In Game 1 (which we play 100 times), you have the highest chance of making money because you succeed in 97% of the 10,000 models I ran! For the second round (when we play ten times), you score in 63% of the simulations, a dramatic decline (and a much higher likelihood that you would lose money). Additionally, 60% of what are called simulations in Game 3—which we have only played once—result in you winning money.



Probability of Making Money for Each Game

Even if the games' result distributions are the same, they are not the same since they possess the same expected value. The greater number of plays we spread out our $100 stake throughout, the more confident we may be in our ability to cash in. This works because, as was already mentioned, each play may be read independently of the others.

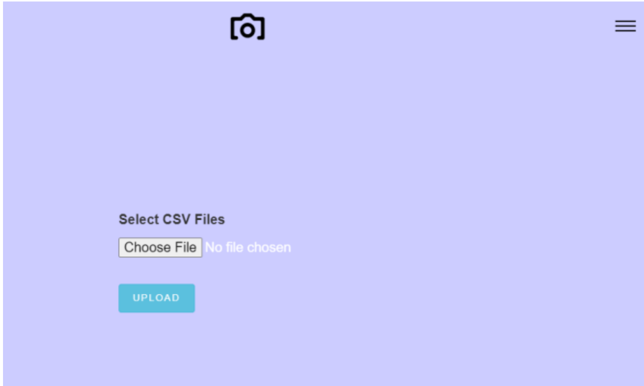
A play from our last game is analogous to every plant in a random forest. We just observed how playing more frequently increased our chances of winning. Comparable to the model of random forests, our approach increases the probability that our predictions would be accurate by including more uncorrelated trees.

V. RESULTS AND DISCUSSION

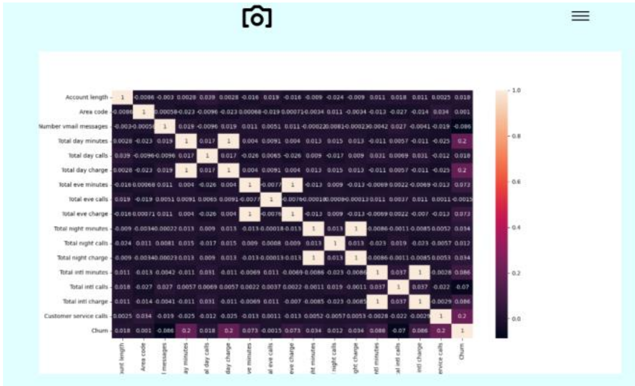
The project's flow and working method are shown in the following screenshots.

Home page: This is the churn prediction in the telecom industries main page. We are determining if the client will be utilizing the same networking or not in our project.

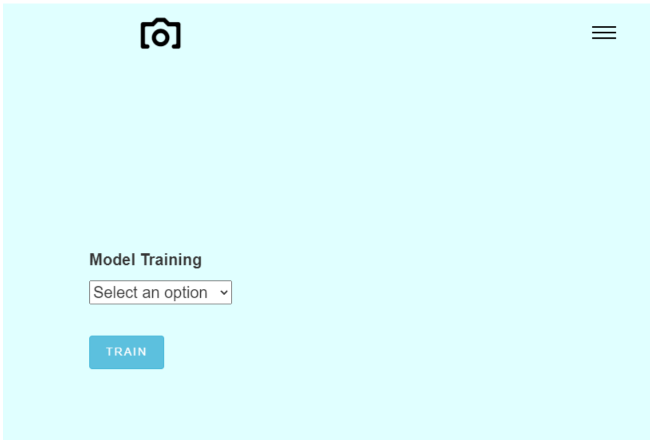
Upload file: The dataset that we are using is being uploaded here.



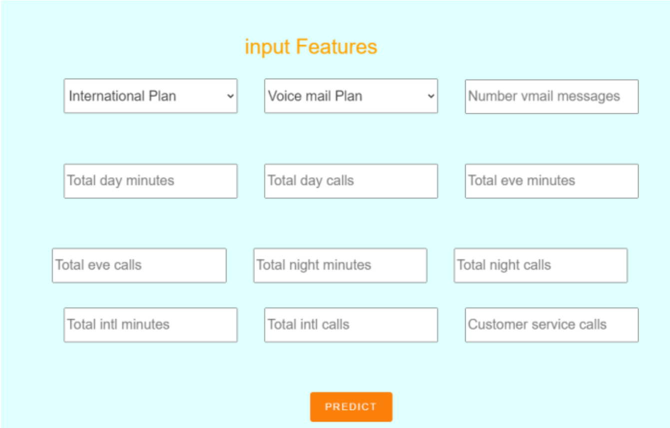
Visualizations: Visualizations using correlation plots are displayed on this page.



Training: On this page, you may choose the model for which you want to get training.



Prediction: The parameters used for prediction are shown on this page.



VI. CONCLUSION

In this application, we have successfully constructed a system to forecast if a consumer would leave or not. This is made in a user-friendly setting using Flask and Python programming. In order to determine if the user will remain with the current network or switch to a different one, the software is likely to collect data from the user.

VII. REFERENCES

[1]. Siu NY, Yau CY, and Zhang TJ. A lesson about service recovery on the function of morality and customer pleasure in customer retention. Journal of Business Ethics, 114(6), 675–686, June 1, 2013.

[2]. Suchy NJ and Hossain MM. Study on the mobile telecommunications industry's relationship between customer happiness and loyalty. 2013;9(2):73-80 in the peer-reviewed journal of the Social Sciences,

[3]. Maldonado S, Flores, Verbraken, Baesens, and Weber. Profit-based feature selection using SVMs: A general framework and a customer retention application. 2015 October 1;35:740-8. Applied Soft Computing.

[4]. The assignee is Accenture Global Services Ltd.; the inventors are Maga M, Canale P, and Bohe A. Churn management and prediction system. U.S. patent number US 8,712,828. 2014 Apr 29

[5]. "Knowledge maintenance and data mining for marketing," Decision-Supported Systems, Vol. 31, no. 1, 2001, pp. 127–137. M. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge.

[6]. "Turning telecommunications call details to churn prediction: a data mining approach," Expert Systems Development with Application development, Vol. 23, 2002, pp. 103–112.

[7]. "Customer churn analysis: Churn determinants and facilitation outcomes of partly defection in the Korean mobile telecommunications service industry," Telecommunications Policy, Vol. 30, Issues 10-11, 2006, pp. 552-568.

[8]. "Non-parametric Statistical Analysis of Machine Learning Methods for Credit Scoring," Advances in Intelligent Systems and Computing, Volume 171, 2012, pp. 263-272. V. Garca, I. Marqués, and J. Sánchez.

[9]. "Data Mining Curriculum: A Proposal," Version 1.0, 2006, S. Chakrabarti, M. Ester, U. Fayyad, J. Gehrke, J. Han, S. Morishita, G. Piatetsky-Shapiro, and W. Wang.

[10]. Voting-based q-generalized extreme learning machine, STOSIC D, STOSIC D, LUDERMIR T. 2016; 174: 1021–1030 in Neurocomputing.

Cite this article as :

Dr. D. Esther Rani, Chaitanya T, Jasmine Sk, Maheshwari T, Sai Geethika V, "Customer Churn Prediction", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.398-406, May-June-2023.