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

Predict Your Customer Through Customer Behavior with Dynamic Churn Prediction Using Machine Learning Algorithms

K. Harisesha Chandana[1] , G. Lakshmikanth[2*]
M. Tech Student[1], Associate Professor[2]
[1&2]Department of Computer Science and Engineering,
Sree Rama Engineering College, Tirupati, Andhra Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 5
September-October-2022

Page Number : 112-121

Article History
Accepted: 10 Sep 2022
Published: 25 Sep 2022

ABSTRACT

In current days, the customers are getting more attracted towards the quality of service (QoS) provided by the organizations. However, the current era is evidencing higher competition in providing technologically advanced QoS to the customers. Nevertheless, efficient customer relationship management systems can be advantageous for the organization for gaining more customers, maintaining customer relationships and improve customer retention by adding more profit to the organizational business. Furthermore, the machine learning models such as support vector machine Random Forest algorithms can add more value to the customer retention strategies.

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

I. INTRODUCTION

Customers always play vital role in increasing profit and revenue of every organization; hence, to gain customer satisfaction it is important for the organizational managers to maintain one efficient customer relationship management system by selecting the target customers and maintaining effective relationship with them. Moreover, the CRM system will be helpful for the organization in identifying the most prominent group of customers and their behavior; which will become beneficial for the organization in understanding the retention strategies in a better way. Additionally, higher the customer loyalty, lesser is the customer churn rate; hence using machine learning algorithm such as support vector algorithm can add value in preventing the customer churn. This report will focus on the customer retention with the usage of support vector machine learning in gaining customer loyalty and increasing retention.

The influence by customers who churned earlier to those who are willing to carry on exchange with an organization negatively impacts the revenue collected by a company regarding a given service or product. The expectations of consumers may not go in line with what they find after purchasing products from a given company as depicted by customers opting for newer producers. The performance of a commodity basically for machinery producing companies, it is hard to convince buyers about a given machine that they test only to find it underperformed. If the gadget has other compliances, the customer will opt for another seller. Consumer management strategies fall into pay, especially in such instances hence making the organization to improve their quality production, raising their brand portfolio.

In the literature on relationship marketing, client pleasure has been viewed as a crucial component in maintaining long-term customer relationships. How to increase customer happiness and keep dissatisfied consumers is therefore a major task when customers experience a service failure. Prior research has emphasized the significance of post-recovery consumer satisfaction and the significance of justice in addressing this issue. Customers must believe that the results are right or fair for them to feel that their dissatisfied needs have been met (Kau and Loh 2006). Hoffman and Kelley (2000) contend that each of the following factors—the service recovery process itself, the results, the interpersonal behaviours displayed throughout the process, and the delivery of the outcomes—is crucial. Their position is consistent with the Tax et al. suggested three-dimensional notion of justice (distributive, procedural, and interactional justice) (1998). But the majority of research take a static approach. Because it builds up over time, it is commonly recognized that businesses must continually raise customer happiness to keep consumers.

The concept of customer satisfaction is differentiated in this study between prior satisfaction (before service failure) and two post-recovery satisfaction factors (i.e., satisfaction with recovery and contentment with organisation). The two main objectives of this study are to first determine whether complaint justice plays a mediating role in the relationship between prior satisfaction and post-recovery satisfaction (both with the recovery and with the organisation), and second, to determine whether post-recovery satisfaction plays a mediating role in the relationship between the dimensions of complaint justice and customer retention. It is asserted that perceptions of complaint fairness are important for repairing the connection and regaining customers' trust in businesses. In the context of service recovery, it assists a company to take advantage of its previous customer pleasure. Meanwhile, we propose that perceived fairness is transformed into a behavioural intention through two mechanisms: satisfaction with recovery and contentment with organization. But in this process, organizational satisfaction can be more important.

Additionally, it enables a better classifier interpretation, which is crucial in business analytics. Practitioners typically refer to many machine learning methodologies as "black boxes," which makes them hesitant to apply the corresponding techniques. Therefore, a deeper comprehension of the data generation process is essential for business analytics decision-making, such as by identifying the features that allow for the explanation of consumer decisions. The most popular methods for validating classifiers and feature selection methods in the past have been those that were statistically inspired. Profit-based metrics have recently been proposed for classifier validation. In this study, we take the concept of profit-driven metrics a step further and apply it to the problem of feature selection by developing numerous embedded approaches that combine the Holdout Support Vector Machine (HOSVM) method with a variety of validation measures.

To the best of our knowledge, data mining and machine learning literature has not yet addressed the unique problem of profit-driven feature selection. The majority of work in business analytics and feature selection use conventional, statistically based methodologies without taking profit-related concerns into consideration. Our tests demonstrate that the suggested approaches perform better than competing strategies and give classifiers highly relevant features, hence lowering the danger of overfitting while simultaneously raising the associated profit.

II. RELATED WORKS

The roles of justice and customer satisfaction in customer retention: A lesson from service recovery: When a service breakdown happens, customers complain because they want the firm to treat them fairly. It has been examined and studied how perceived complaint fairness affects consumer happiness. But the majority of earlier work used a static perspective. We contend that since satisfaction is cumulative, it is important to include both previous and post-recovery satisfaction when evaluating complaint fairness in the context of service recovery. By studying the mediating function of justice in the link between previous contentment and post-recovery satisfaction (both with the recovery and with the organization), as well as the relationship between the dimensions of justice and customer retention, this study aims to close the gap. A sample of 200 customers who experienced service failure in Chinese restaurants in Hong Kong was used to test hypotheses. The association between previous contentment and satisfaction with recovery was shown to be entirely mediated by the justice aspects of distributive justice, procedural justice, and interactional justice. The association between previous contentment and post-recovery satisfaction with organization was likewise shown to be partially mediated by all characteristics, with the exception of interactional fairness. The results also showed that two post-recovery satisfaction factors had mediating functions in converting the justice dimensions into behavioural intention, with the two variables having nearly opposing functions. The future development and enhancement of establishing long-term relationships with clients are discussed, and suggestions are made.

Profitbased feature selection using support vector machines–General framework and an application for customer retention: Churn prediction is a crucial use of classification models that pinpoints the clients most likely to attrite based on the traits denoted by, for example, socio-demographic and behavioural factors. When constructing customer retention systems based on churn prediction models, an effective handling of the ensuing information overload becomes a very significant issue because these attributes are now increasingly being recorded and kept in the relevant computational systems. As a result, choosing features is a crucial step in building the appropriate classifier. the majority of feature picking methods; nonetheless, are founded on statistically motivated validation criteria, which don't always result in models that best achieve the objectives set out by the relevant organization. In this research, we present a profit-driven strategy for SVM-based classifier development and simultaneous variable selection. Experimental findings demonstrate that our models outperform traditional feature selection strategies, yielding higher performance in terms of business. Associated objectives Preprint submitted to Applied Soft Computing on February 1, 2015.

Churn prediction using comprehensible support vector machine: An analytical CRM application: Due to its capacity to simulate nonlinearities, support vector machines (SVM) are now state-of-the-art for classification problems. The fundamental disadvantage of SVM is that it creates "black box" models, which means that it does not convey the information learned during training in a way that is understandable to humans. Rule extraction is the process of transforming such opaque models into transparent models. In this research, we suggested a hybrid technique for extracting rules from SVM for CRM applications. Three steps make up the suggested hybrid strategy. I SVM-RFE (SVM-recursive feature elimination) is used in the first phase to condense the feature set. (ii) In the second phase, a dataset with less characteristics is utilised to create an SVM model, and support vectors are retrieved. In the last stage, the Naive Bayes Tree (NBTree) is used to produce rules. The Business Intelligence Cup 2004 dataset, which was used to examine the churn prediction among bank credit card users, has a highly uneven customer retention rate of 93.24% and a churn rate of 6.76%. In order to balance the data and derived rules, we also used a variety of typical balancing procedures. The empirical findings show that the suggested hybrid outperformed every other strategy examined. It is also noted that the suggested technique extracts lesser length rules when the reduced feature dataset is employed, increasing the system's understandability. For the bank management, the created rules serve as an early warning expert system.

Churn prediction in the telecommunications sector using support vector machines: Customer turnover is one of the major issues for businesses across all sectors these days as a result of difficulties brought on by global competition. The telecommunications industry earns the top spot on the list with a turnover rate of 30%. Predictive models must be used to find clients who are likely to leave in order to tackle this issue. This study presents a sophisticated strategy for estimating customer turnover in the mobile telecommunications sector. Each of the 3333 entries in the dataset's call information records includes 21 characteristics. To create the prediction models, we employ a Support Vector Machines technique with four kernel functions. Gain measure is used to assess and compare the models' performance.

Fast training of support vector machines using sequential minimal optimization: Although sequential minimum optimization (SMO) is a well-liked approach for training support vector machines (SVM), it still takes a long time to solve problems of a big scale. This study suggests a single parallel SMO implementation for SVM training. Message passing interface is used to construct the parallel SMO (MPI). In further detail, the parallel SMO divides the total training data set into smaller subsets before running several CPU processors simultaneously to deal with each of the divided data sets. When multiple processors are deployed, experiments on the adult data set and the Mixing National Institute of Standard and Technology (MNIST) data set demonstrate a significant speedup. On the Web data set, there are also positive outcomes.

A fast SVM training algorithm: The decomposition structure of the SVM technique is used to present a quick support vector machine (SVM) training approach that successfully combines kernel caching, digest and shrinkage rules, and halting conditions. Numerous tests using the MNIST handwritten digit database have been done to demonstrate that the suggested technique is about 9 times quicker than Keerthi et almodified.'s SMO. The overall training time for ten one versus the rest classifiers on MNIST combined with principal component analysis was just 0.77 hours. Due to the suggested scheme's promised scalability, SVM may be used to a number of engineering challenges.

III. METHODOLOGY

Proposed system:

In proposed system, we implement a Machine Learning algorithms for getting insights from the complex patterns in the data. This technique is computationally inexpensive because of its simple architecture.

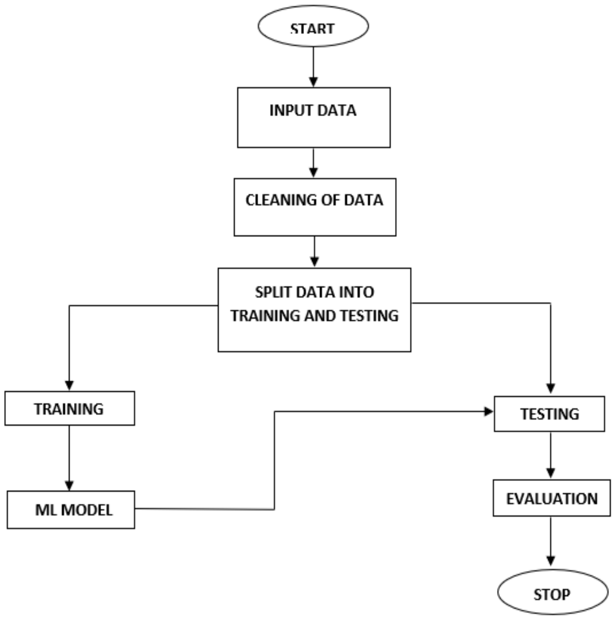


Figure 1: Block diagram

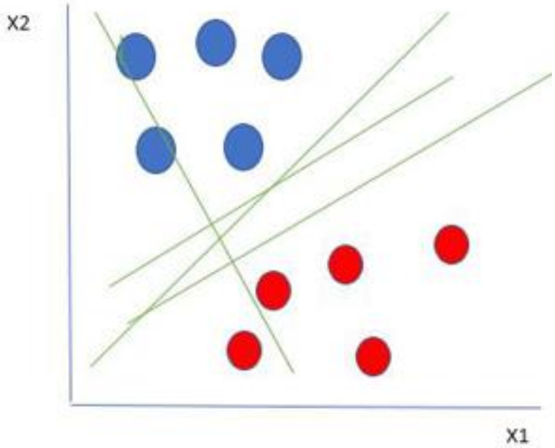
IV. IMPLEMENTATION

The project has implemented by using below listed algorithm.

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyper plane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyper plane depends upon the number of features. If the number of input features is two, then the hyper plane is just a line. If the number of input features is three, then the hyper plane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

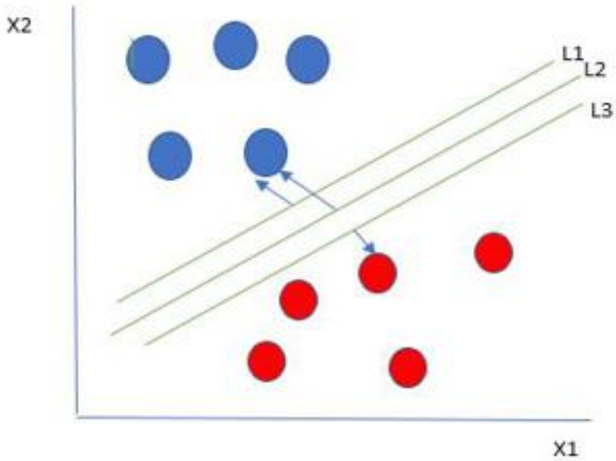
Let's consider two independent variables x1, x2 and one dependent variable which is either a blue circle or a red circle. Types of SVM.



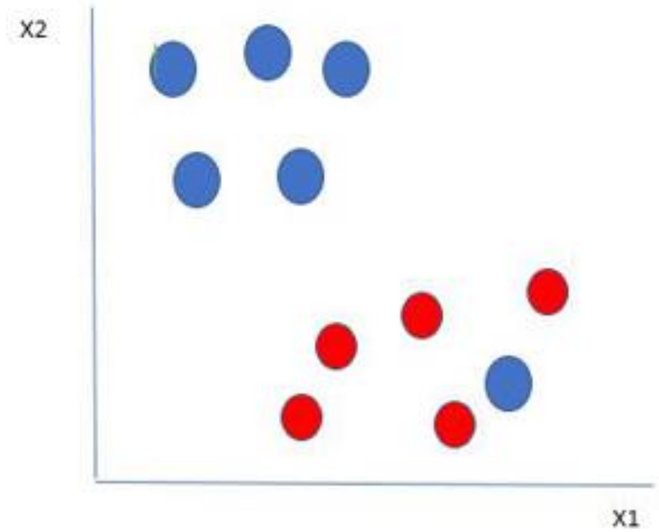
its very clear that there are multiple lines (our hyper plane here is a line because we are considering only two input features x1, x2) that segregates our data points or does a classification between red and blue circles. So how do we choose the best line or in general the best hyper plane that segregates our data points.

Selecting the best hyper-plane:

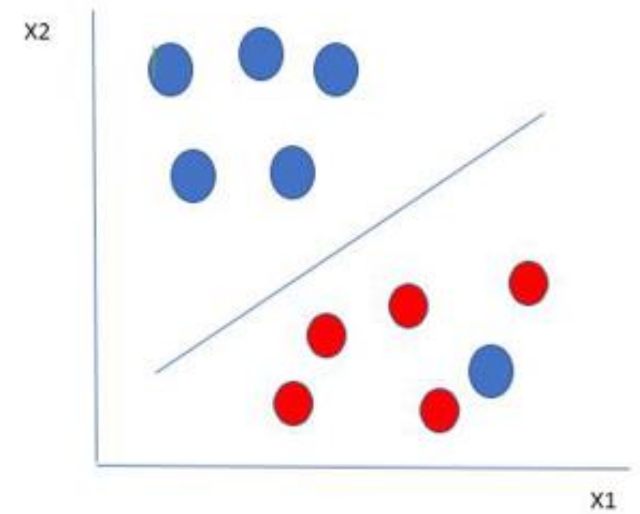
One reasonable choice as the best hyper plane is the one that represents the largest separation or margin between the two classes.



So we choose the hyper plane whose distance from it to the nearest data point on each side is maximized. If such a hyper plane exists it is known as the maximum-margin hyper plane/hard margin. So from the above figure, we choose L2.

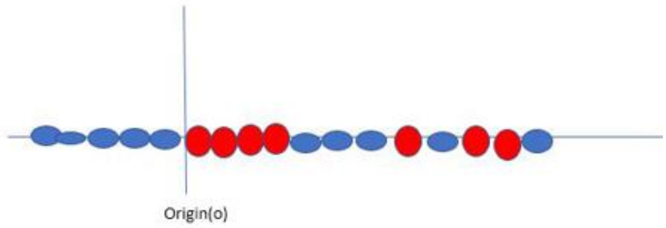


Here we have one blue ball in the boundary of the red ball. So how does SVM classify the data? It's simple! The blue ball in the boundary of red ones is an outlier of blue balls. The SVM algorithm has the characteristics to ignore the outlier and finds the best hyperplane that maximizes the margin. SVM is robust to outliers

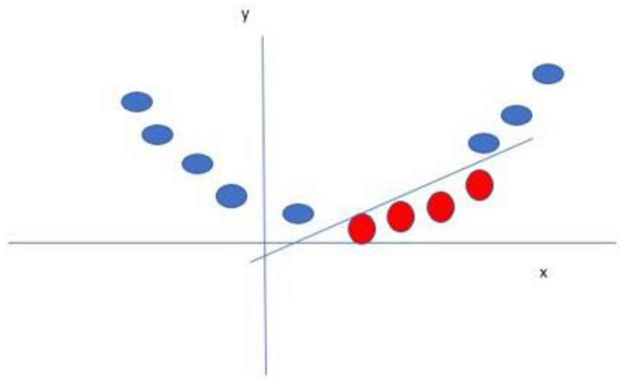


So in this type of data points what SVM does is, it finds maximum margin as done with previous data sets along with that it adds a penalty each time a point crosses the margin. So the margins in these type of cases are called soft margin. When there is a soft margin to the data set, the SVM tries to minimize *(1/margin+Λ(∑penalty))*. Hinge loss is a commonly used penalty. If no violations no hinge loss. If violations hinge loss proportional to the distance of violation.

Till now, we were talking about linearly separable data(the group of blue balls and red balls are separable by a straight line/linear line). What to do if data are not linearly separable?



Say, our data is like shown in the figure above. SVM solves this by creating a new variable using a kernel. We call a point x_i on the line and we create a new variable y_i as a function of distance from origin o.so if we plot this we get something like as shown below



In this case, the new variable y is created as a function of distance from the origin. A non-linear function that creates a new variable is referred to as kernel.

SVM Kernel:

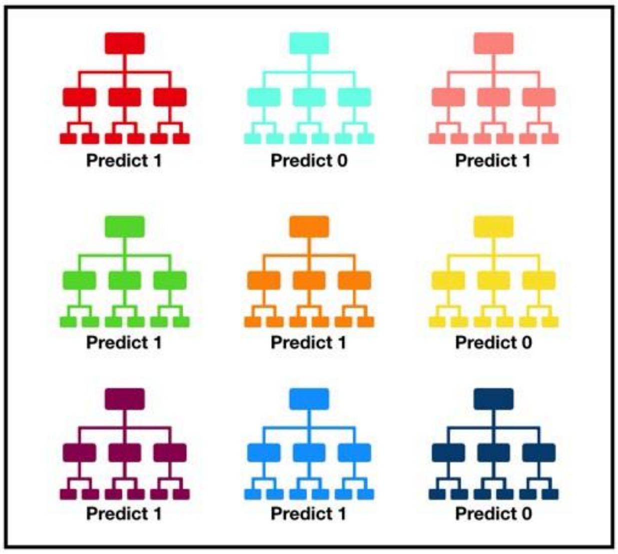
The SVM kernel is a function that takes low dimensional input space and transforms it into higher-dimensional space, ie it converts not separable problem to separable problem. It is mostly useful in non-linear separation problems. Simply put the kernel, it does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined.

Advantages of SVM:

- Effective in high dimensional cases
- Its memory efficient as it uses a subset of training points in the decision function called support vectors
- Different kernel functions can be specified for the decision functions and its possible to specify custom kernels

Random Forest:

Like its name suggests, a random forest is made up of several independent decision trees that work together as an ensemble. Every every tree in the random forest spits out a class forecast, and the classification that receives the most votes becomes the prediction made by our model (see figure below).



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

Making a Prediction Using Visualization of a Random Forest Model

The wisdom of crowds is the basic idea underlying random forest, which is a straightforward but effective idea. The reason the random forest model performs so well, in terms of data science, is:

A large number of generally uncorrelated models (trees) working together will perform better than any of the component models working alone.

The key is the poor correlation between models. Uncorrelated models can provide ensemble forecasts that are more accurate than any of the individual predictions, much like how assets with low correlations (such stocks and bonds) combine to build a portfolio that is more than the sum of its parts.

The trees shield each other from their own mistakes, which results in this amazing effect (So long as they don't all consistently err in the same way). Many trees will be right while some may be wrong, allowing the group of trees to travel in the proper direction. Consequently, in order for random forest to operate successfully:

1. 1. Our features must include some real signal in order for models created utilizing them to outperform guesswork.
2. 2. There must be low correlations between the predictions (and hence the mistakes) produced by each tree.

It's important to understand the amazing consequences of having numerous uncorrelated models, therefore I'll give you an illustration to make the point. Assume we are participating in the following game:

- I create a number using a uniformly distributed random number generator;
- if the result is larger than or equal to 40, you win (you have a 60% probability of winning), and I give you some money.
- I win and you give me the same amount if it is less than 40.
- At this point, I'm giving you the following options. The options are:

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

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

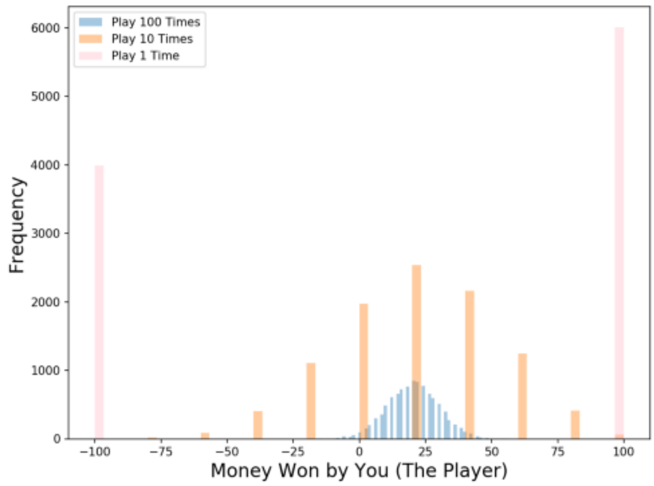
3. Game 3: Place a single $100 wager.

Which one do you prefer? Each game's anticipated value is the same:

Expected Value Game 1 = (0.60*1 + 0.40*-1)*100 = 20

Expected Value Game 2= (0.60*10 + 0.40*-10)*10 = 20

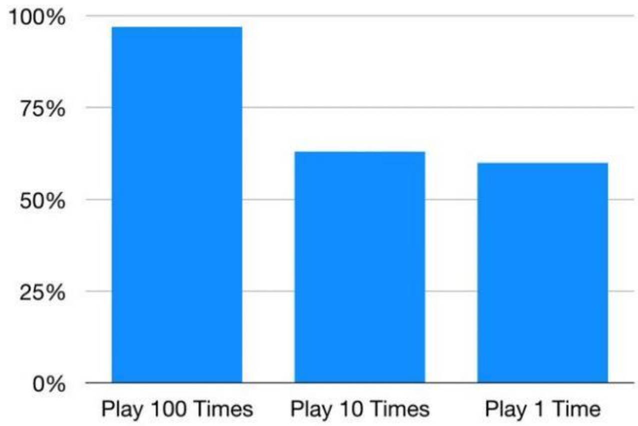
Expected Value Game 3= 0.60*100 + 0.40*-100 = 20



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

The distributions, what about them? Let's use a Monte Carlo simulation to see the outcomes (we will run 10,000 simulations of each game type; for example, we will simulate 10,000 times the 100 plays of Game 1). Which game would you choose based on the chart on the left? the same even when the predicted values, **As the result distributions transition from positive and narrow (blue) to binary, they significantly diverge (pink).**

The best opportunity to gain some money is in Game 1 (where we play 100 times), where you win in 97% of the 10,000 simulations I conducted! You win in 63% of the simulations for Game 2 (where we play ten times), a sharp fall (and a drastic increase in your probability of losing money). And in Game 3, which we just play once, 60% of the simulations result in you making money, as predicted.



Probability of Making Money for Each Game

Since the games have the same expected value, even if their outcome distributions are identical, they are not identical. The more plays we divide our $100 wager into, the more certain we may be that we will win. This works because each play stands alone from the others, as was previously noted.

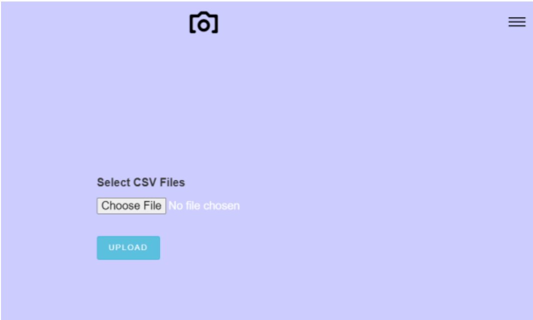
Each tree in a random forest is similar to a play in our earlier game. Just now, we saw how playing more often boosted our likelihood of winning. Similar to a random forest model, the more uncorrelated trees we include in our model, the higher the likelihood that our forecasts would be accurate.

V. RESULTS AND DISCUSSION

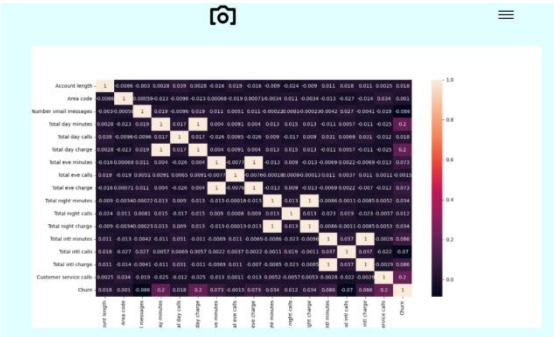
The following screenshots are depicted the flow and working process of project.

Home page: This is the home page of churn prediction in telecom industries. In our project, we are detecting whether the customer will be using the same network or not.

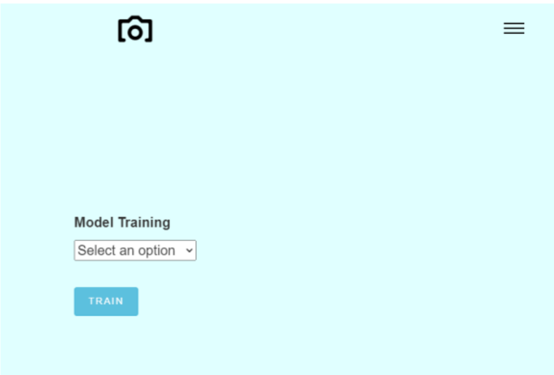
Upload file: Here we are uploading the dataset through which we are working.



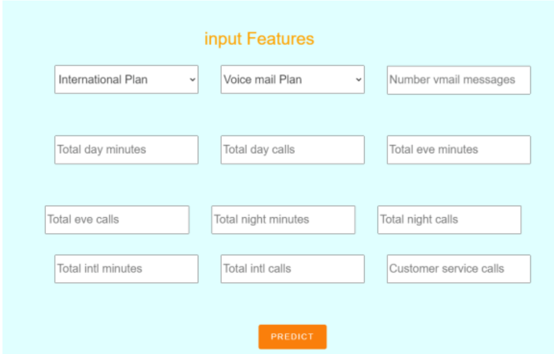
Visualizations: This page displays Visualizations with correlation plot.



Training: In this page we can select the model training and training will be done with selected model.



Prediction: This page displays the input features for prediction.



VI.CONCLUSION

We have successfully developed a system to predict whether the customer is churned or not in this application. This is created in a user-friendly environment with Python programming and Flask. The system is likely to gather data from the user in order to predict whether the customer will continue with the same network or the customer may change the network.

VII. REFERENCES

[1]. Siu NY, Zhang TJ, Yau CY. The roles of justice and customer satisfaction in customer retention: A lesson from service recovery. Journal of business ethics. 2013 Jun 1;114 (4):675-86.

[2]. Hossain MM, Suchy NJ. Influence of customer satisfaction on loyalty: A study on mobile telecommunication industry. Journal of Social Sciences. 2013;9(2):73-80.

[3]. Maldonado S, Flores Á, Verbraken T, Baesens B, Weber R. Profitbased feature selection using support vector machines–General framework and an application for customer retention. Applied Soft Computing. 2015 Oct 1;35:740-8.

[4]. Maga M, Canale P, Bohe A, inventors; Accenture Global Services Ltd, assignee. Churn prediction and management system. United States patent US 8,712,828. 2014 Apr 29

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

[6]. C. P. Wei and I. T. Chiu, "Turning telecommunications call details to churn prediction: a data mining approach," Expert Systems with Applications, Vol. 23, 2002, pp. 103-112.

[7]. J. H. Ahn, S. P. Han, and Y. S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," Telecommunications Policy, Vol. 30, Issues 10–11, 2006, pp. 552-568.

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

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

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

[11]. Jadhav, R. J., & Pawar, U. T. (2011). Churn prediction in telecommunication using data mining technology. International Journal of Advanced Computer Science and Applications, 2(2).

[12]. Phadke, C., Uzunalioglu, H., Mendiratta, V. B., Kushnir, D., & Doran, D. (2013). Prediction of subscriber churn using social network analysis. Bell Labs Technical Journal, 17(4), 63-76.

[13]. Rosenberg, L. J., & Czepiel, J. A. (1984). A marketing approach for customer retention. Journal of consumer marketing, 1(2), 45-51.

[14]. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1-9.

Cite this article as :

G. Harisesha Chandana, K. Lakshmikanth, "Predict Your Customer Through Customer Behavior with Dynamic Churn Prediction Using Machine Learning Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 112-121, September-October 2022.
Journal URL : https://ijsrcseit.com/CSEIT228520