**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**
**IMPACT FACTOR: 6.199**

***IJCSMC, Vol. 8, Issue. 4, April 2019, pg.94 – 102***

# Next Best Action Using Prediction Analysis

**Rohan Naik[1]; Saurabh Mane[2]; Chirag Mali[3]; Nidhi Manglore[4]; Sankusu Sharma[5]**

[1]Department of Computer Engineering, St. John College of Engineering and Management, India
[2]Department of Computer Engineering, St. John College of Engineering and Management, India
[3]Department of Computer Engineering, St. John College of Engineering and Management, India
[4]Department of Computer Engineering, St. John College of Engineering and Management, India
[5]Department of Computer Engineering, St. John College of Engineering and Management, India
[1] naikrohan73@gmail.com; [2] saurabhmane1997@gmail.com; [3] chiragmali98@gmail.com;
[4] mangalorenidhi7@gmail.com; [5] sankusus@sjcet.co.in

***Abstract— As the market has become increasingly saturated, churn prediction and management has become of great concern to many industries. A company wishing to retain its customer needs to be able to predict those who are likely to churn and will make those customers the focus of customer retention efforts. Today Customer data has properties of large samples, high dimensions, and more noises. In response to the limitations of existing feature selection in churn-prediction, we introduce and experimentally evaluate Support vector machine-recursive feature elimination attribute selection algorithm. It can identify key attributes of customer churn, rule out the related and redundant attributes, and reduce the dimensions of data. It is more important that this algorithm is related to the followed classification learning algorithm so that it can be better integrated into churn prediction. The empirical evaluation results suggest that the proposed feature selection algorithm extracts less key attributes and exhibits better satisfactory predictive effectiveness than the other three comparable attribute selection algorithms. Using this algorithm, a company can predict which customer has more probability to churn and can provide a scheme or take appropriate actions so that the customer continues to use the services of the company.***
***Keywords— Churn, Telecom, SVM, Prediction, Machine Learning, Supervised Learning***

## I. Introduction

Recent successes in applying churn prediction algorithms have inspired Machine Learning researches to explore new research opportunity at the intersection of these previously separated domains. "Churn" refers both to customers' migration and to their loss of value. So, "churn rate" refers either to the percentage of customers who end their relationship with the organization, or to the customers who still receive their services, but not as much or not as often as they used to. There are various algorithms used to predict and analyze the churn value of the customer. There is no one strategy for anticipating which customers are likely to churn in the next month, Even the expression "churn modeling" has numerous implications: It can refer to calculating the proportion of customers who are churning, forecasting a future churn rate, or predicting the risk of churn for particular individuals. Current algorithms that are under use are not efficient as they utilize a large amount of space and produce poor results. In this thesis, we are exploring SVM and comparing it with existing algorithms to corroborate how more efficient results can be produced and churn prediction can be widely implemented. Space Vector Machine is a supervised learning model that analyses data used for classification and regression analysis. SVM should find out all the factors essential to predict the churn and then reduce the amount of data. Moreover, after churning the algorithm should also provide a solution for the same to avoid churning.

 

## II. Literature Review

In [1] we explore the implication of Synthetic Minority Over-sampling Technique (SMOTE) to reduce the imbalance in data in collaboration with different feature reduction techniques such as Co-relation feature extraction, Gain ratio, Information gain and feature evaluation method. Classification and Regression Trees (CART), Bagged CART and Partial Decision Trees (PART) classifiers are trained to analyse the performance on balanced and reduced feature space dataset. Prediction performance of the classifiers is evaluated through measures such as Area under the Curve (AUC), sensitivity and specificity. Finally, it is concluded through simulations that the proposed method based on SMOTE, co-relation, and ensemble performs well for predicting churners as against simply applying learners on the unrefined dataset. Therefore, this methodology can be helpful for the telecommunication industry to predict churn.

In [2] author's viz. Y. Li, G. Xia have used decision tree and neural network for churn prediction. They have delineated the process of churn prediction right from data acquisition to churn analysis. They have also focused on pre-processing where data cleaning, feature abstraction is used before giving it as input to the algorithms. In their study, they observed that the decision tree model surpasses the neural network model in the prediction of churn and it is also easy to construct. They have proposed that selecting the right combination of attributes and fixing the proper threshold values may produce more accurate results. Their study limits itself with the prediction of churn, and no steps were analyzed to include retention policies. This can not only improve the prediction accuracy of the model but also make it explainable itself, which is of profound researching value. Logistic Regression helps to understand the degree to which each feature affects the decision of churn and decision tree provides a graphical overview of the available data from which rules can be generated, and strategies can be built for customer retention.

In [3] the main aim was to improve IBM's ROC area score. Some of the teams in KDD 2009 competition have decided to use the ensemble methods. They observed that ensemble methods outperformed the individual classifiers for the dataset they created. Orange Telecom also posted the small version of the dataset which included 50,000 examples and 230 variables. They performed some experiments with this small dataset and planned to test the best performing set of methods to the larger dataset after improving the result from the small one. Because the dimension of the feature set was significantly large, to improve the overall classification performance they decided to implement feature selection. They experimented on the effects of different feature subsets on the performance of a classifier. They excluded the features that had single or no value from the dataset. In the output huge percentage of values of variable 9 gathered around 0.

In [4] they used improved balanced random forests (IBRF) to predict the customer churn while integrating a sampling technique and cost-sensitive learning into the standard random forests to achieve a better performance than most existing algorithms. The nature of IBRF is that the best features are iteratively learned by altering the class distribution and by putting higher penalties on misclassification of the minority class. Applied to a credit debt customer database of an anonymous commercial bank in China, they were proven to significantly improve prediction accuracy compared with other algorithms, such as artificial neural networks, decision trees, and class-weighted core support vector machines (CWC-SVM).

In [5] the authors proposed that the system provides a statistical survival analysis tool to predict customer churn based on the comparison between decision trees and logistic regression. Selecting the right combination of attributes and fixing the proper threshold values may produce more accurate results of predicting churn customers. The proposed model suggests that data mining techniques can be a promising solution for customer churn management. Telecom companies can predict in advance which customers are at risk of leaving, and those customers consequently saving a lot of revenues namely the ones which are used for replacing the also lost the ones that are wasted for retaining already loyal customers.

Churn model is not the one with the best statistical precision but the one that provides the best insights to prevent churn behavior further. Using the results obtained using decision trees and logistic regression, it will be easy to design retention policies and tactics to help maintain the customers as these methods provide easily deducible explanations about the reasons behind the decision of churn along with the list of customers with high probability to churn.

In [6] the authors have categorized churn in two types; (1) voluntary, those who leave for their reasons, and (2) involuntary, those who are released from their services by the organization. Usually, companies focus on voluntary churn, where an employee would either leave for a better opportunity in terms of pay, benefits, work environment, etc. or due to negative reasons at an organization such as conflict with the supervisors, lack of opportunities for promotion, lack of interesting work and many more. According to them, employee churn results in financial, time and effort loss for organizations. It is a big issue since trained and experienced hard to replace and costly.

They analyzed current employee data to predict future churners and learn the causes of employee turnover. The results of their study demonstrated that data mining algorithms could be used to build reliable and accurate predictive models for employee churn. The problem of churn prediction is not just to identify churners from no churners. By using exploratory data analysis and data mining techniques, they predicted the churn probability for each employee and gave them a score to make the retention strategies.

## III. METHODOLOGY

*A. Architecture Diagram:*

The System consists of the following modules:

1. Data Exploration
2. Feature Selection
3. Feature Pre-Processing
4. Model Training and Testing
5. Result Evaluation

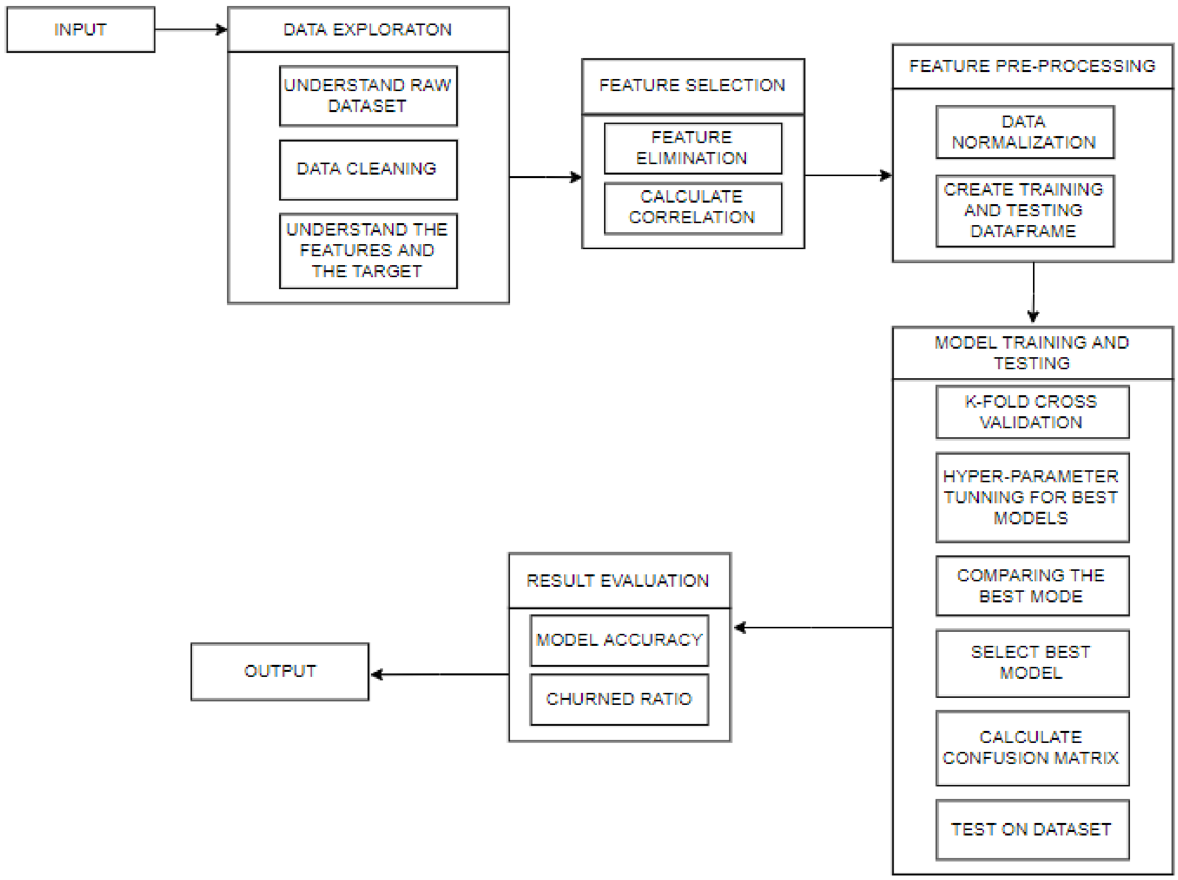


Fig. 1. Architecture Diagram

*B. Implementation:*

*1. Data Exploration:*

Data exploration is the initial step in data analysis and generally includes summarizing the main characteristics of a data set, including its size, precision, original patterns in the data, cleaning the data and other attributes. Cleaning of data includes removing white spaces, checking if there are any vacant values in the dataset. If white spaces are detected then they are removed, if there are any missing values, then those values are appended by a garbage value, i.e., 0.

 

Our study is performed on a database that has nearly 5,000 customers described by 21 variables. The variables that have nothing to do with prediction are excluded. Then, 7 variables remain. The distribution of data used in training and testing are given in Table 1.

Table 1. Distribution of data used in Training & testing

| **Training dataset** | Number of normal examples | Number of churn examples |
|---|---|---|
| | 4300/5000 | 700/5000 |

The data is unbalanced and thus needs to be balanced to differentiate reliably between defectors and nondefectors. This is done using cross-validation.

Further K-S test and Z-test are performed on each feature to check whether the distributions of each feature of churn or not are drawn from the same distribution and whether the difference in mean values is statistically different for features with 0/1 values.

As **account_length**, **total_day_calls**, **total_eve_calls**, and **total_night_calls** return high p-values (>0.23) so null hypothesis can't be rejected. The difference between the two samples are not significant, these features are probably irrelevant to our target variable.

2. *Feature Selection:*

Feature Selection is the process where features that contribute most to the output or prediction variable are automatically or manually selected. Having irrelevant features in data can decrease the accuracy of the models and make the model learn based on unnecessary features.

Feature selection methods are used for four reasons:

- Simplification of models to make them simpler to decipher.
- shorter training times
- to avoid the curse of dimensionality
- enhanced generalization by reducing overfitting

We perform feature selection using Correlation Matrix with Heatmap. Refer to fig.2.

From Data Exploration and Feature Selection steps, the redundant features are removed. Thus the seven important features are **intl_plan, number_vmail_messages, total_day_charge, total_eve_charge, total_night_charge, total_intl_charge** and **number_customer_service_calls**.

3. *Feature Pre-Processing:*

It is a step in machine learning where numerical and categorical data is normalized or standardized. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. For machine learning, every dataset does not require normalization, it is required only when features have different ranges.

4. *Model Training and Testing:*

This is where various Machine learning models are trained and compared to find out which performs the best. The four specific models being compared are

- Support Vector Machine (**SVM**)
- Random Forest
- K Nearest Neighbors (**kNN**)
- Logistic Regression

Cross-validation is a statistical method used to estimate the versatility of machine learning models. It is normally utilized in applied machine learning to think about and select a model for a given predictive modeling problem since it is straightforward, and results in skill estimates that generally have a lower bias than other methods. Refer to fig.3.

*Support Vector Machine (**SVM**)*

Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while consequently abstaining from overfitting the data. Support Vector machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. Given a set of training examples, each marked as belonging to one or the other of the two categories, An SVM training algorithm builds a model that assigns new examples to either of the categories, making it a non-probabilistic binary linear classifier. SVM deals with separation of categories that are represented as points in space. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can proficiently perform a non-linear classification utilizing kernel trick which involves implicitly mapping their inputs into high-dimensional feature spaces. On prediction, if the customer is on the churning side, then we analyze the factors that the customer use the most and accordingly give him another scheme so that the telecom company retains its customers. Refer to fig.4.

*Random Forest*

Random forest is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

The first algorithm for random decision forests was created by Tin Kam Ho using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. Random Forest has similar hyperparameters as a decision tree or a bagging classifier Random Forest adds additional randomness to the model while growing the trees. Rather than scanning for the fundamental component while splitting a node, it searches for the best feature among arbitrary subset of features. This results in a wide diversity that generally results in a better model.

Therefore, in Random Forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node.

Each tree is developed as follows:

If the number of cases in the training set is N, sample N cases at random **-** but *with replacement*, from the original data. This sample will be the training set for growing the tree.

1. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is kept constant during growing the forest.
2. Each tree is developed to the highest degree conceivable. There is no pruning.

Refer to fig.5.

*K Nearest Neighbors (**kNN**)*

KNN can be utilized for both classification and regression predictive problems. Be that as it may, it is generally utilized in classification problems in the industry. To evaluate any technique, we usually look at three critical aspects:

1. Ease to interpret the output
2. Calculation time
3. Predictive Power

We can implement a KNN model using following steps:

1. Load the data
2. Initialize the value of k
3. To get the predicted class, iterate from 1 to the total number of training data points.
    1. Calculate the distance between test data and every row of training data. Use Euclidean distance, Chebyshev or cosine as distance metric.
    2. Sort the calculated distances in ascending order based on distance values.
    3. Get top k rows from the sorted array.
    4. Get the most frequent class of these rows.
    5. Return the predicted class.

Refer to fig.6.

*Logistic Regression*

Logistic Regression is used when the target variable is categorical.

For example,

To predict whether an email is a spam (1) or not(0).

Whether the tumor is malicious (1) or not (0).

Consider a situation where we need to classify whether an email is a spam or not. In the event that we utilize linear regression for this issue, there is a requirement for setting up a threshold depending on which classification should be possible.. Say if the actual class is malicious, where predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malicious which can lead to serious consequences in real time.

From this example, it can be induced that linear regression is not appropriate for the classification problem. Linear regression is unbounded, and thus logistic regression is brought into the picture, whose value strictly ranges from 0 to 1.

*5. Result Evaluation:*

The accuracy of Logistic regression is less and is unable to generalize the prediction accuracy. KNN, Random Forest tend to overfit the training set. SVC classifier better generalizes the prediction since the training and cross-validation curves are close together.

Thus, the results of K-fold cross-validation and Learning curves show that the accuracy of Support Vector Machine (SVM) is the highest and is chosen for churn prediction. Refer to fig.8.

In the testing phase, we tested dataset consisting of 1000 customer values to predict the possibility of whether the customer that will churn or not.

After testing, SVM was found to be the most accurate for the given dataset with the accuracy of 94.3%. Fig.9. Shows the ratio of churned customers to ratio of unchurned customers. Fig.10. shows some churned values from the output list.

## IV. RESULTS

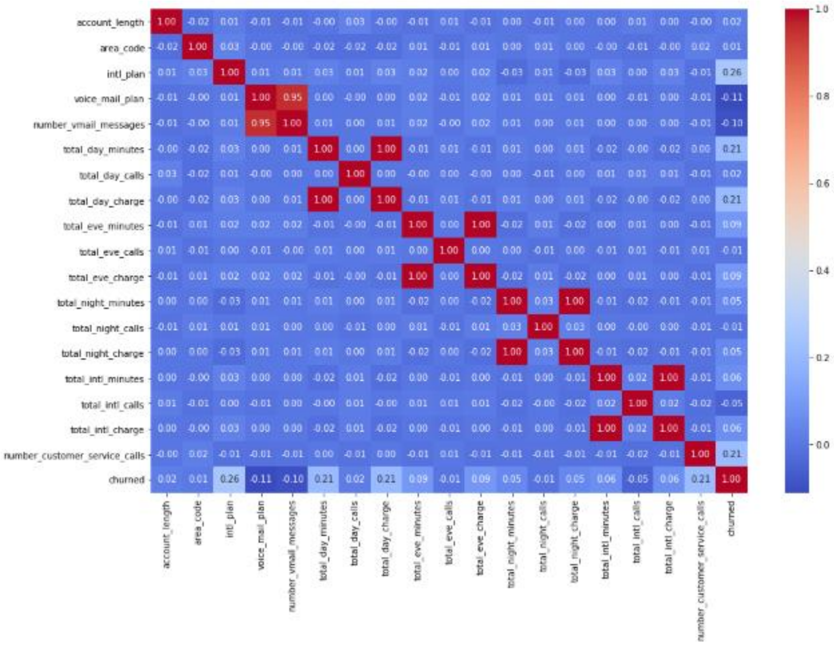


Fig. 2. Heatmap of Correlation Matrix

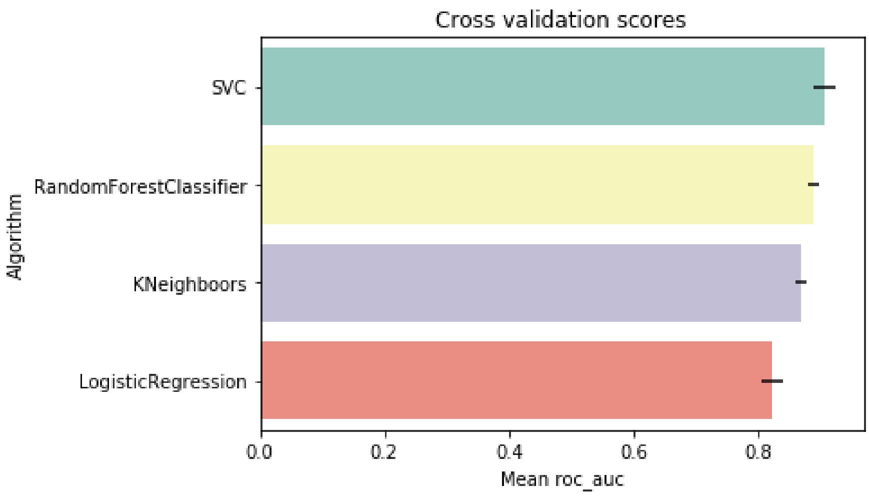


Fig. 3. Cross Validation

 

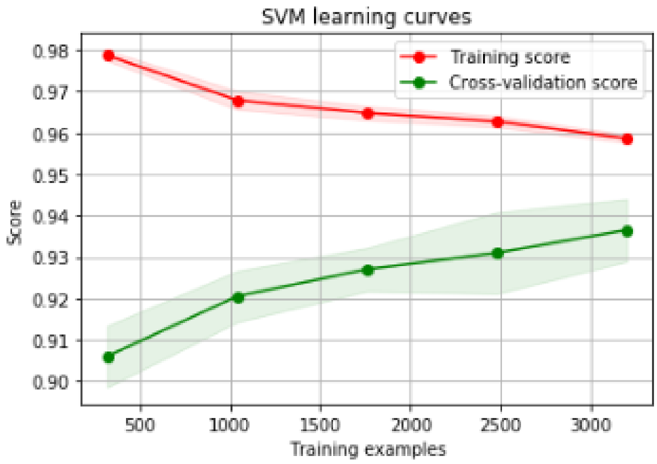


Fig. 4. Learning Curve for SVM

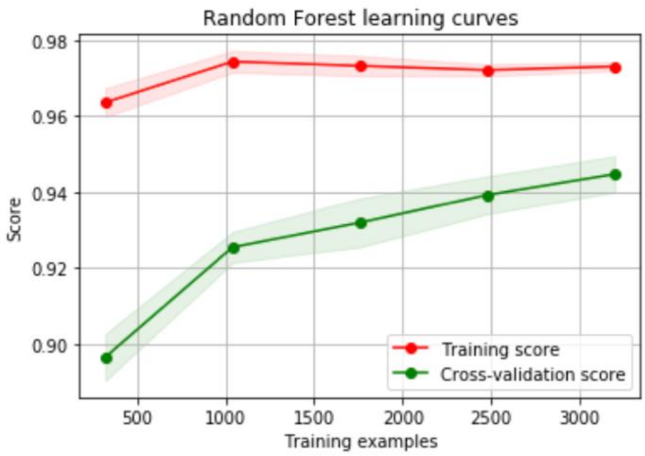


Fig. 5. Learning Curve for Random Forest

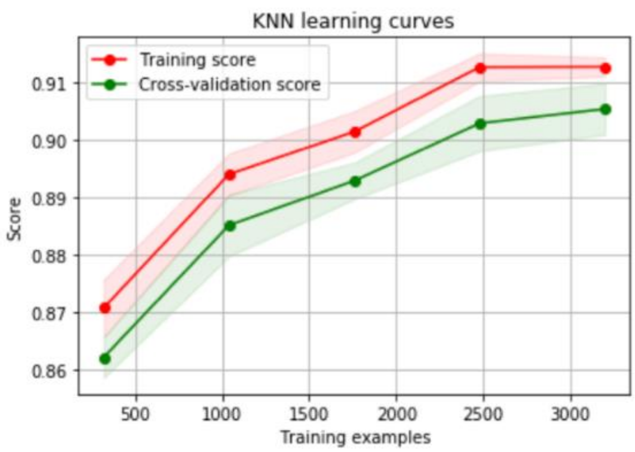


Fig. 6. Learning Curve for kNN

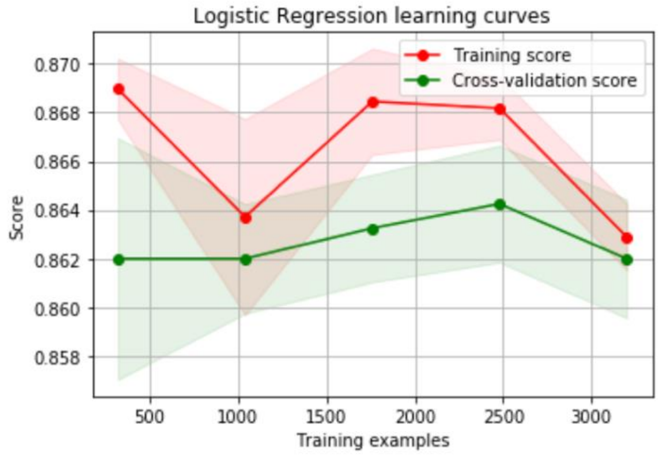


Fig. 7. Learning Curve for Logistic Regression

```
Support Vector Machine
Accuracy is 0.943
Precision is 0.8867924528301887
Recall is 0.6762589928057554
F1 score is 0.7673469387755103
ROC AUC is 0.9090817937984108
```

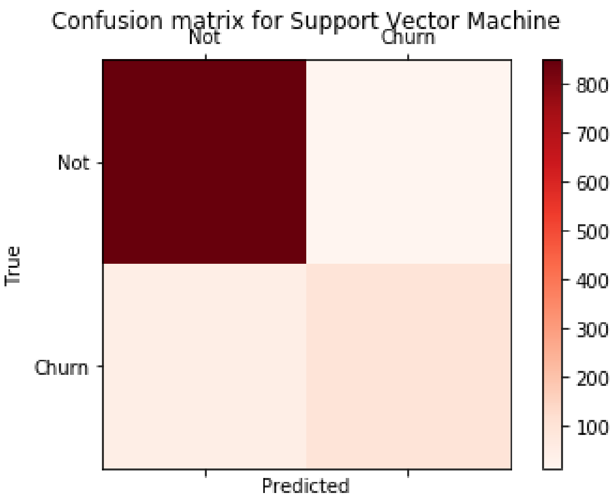


Fig. 8. Confusion Matrix for SVM

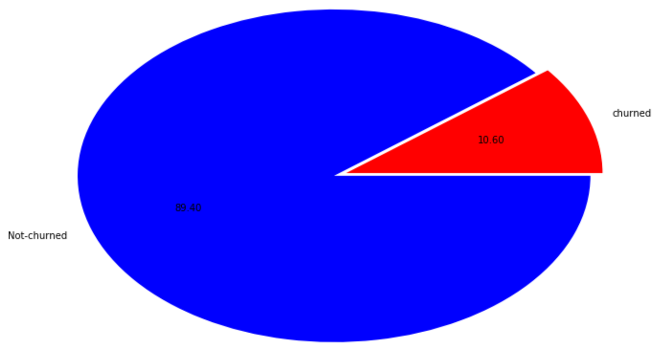


Fig. 9. Ratio of Churned and Not-Churned

| | Prediction |
|---|---|
| Phone no | |
| 329-5112 | 1 |
| 338-2768 | 1 |
| 388-5162 | 1 |
| 360-1292 | 1 |
| 345-8273 | 1 |
| 351-3724 | 1 |
| 384-1980 | 1 |
| 397-2769 | 1 |
| 378-7878 | 1 |

Fig. 10. Predicted churn values

## V. Conclusion

The idea is useful for predicting the churn of the user with high accuracy. Customer churn in product-based companies will affect the company's status as well as finance. Predicting the possibility of the customer to churn is therefore considered as an essential aspect in the companies. This will prevent the customers from leaving the company, thus maximizing the profit and improving the status of the company. Here Support Vector Machine is used to predict the possibility of the customer to churn from the company since SVM is more accurate than the other models. First, the data is cleaned and then it is trained. After that, the data is tested, and the accuracy of the machine is calculated. As a future scope of work, our system can be implemented in companies with different dataset to retain the customers for leaving the company.

## Acknowledgement

We thank our guide, Mr. Sankusu Sharma who has extended all valuable guidance and help through various stages for the development of the project. His Valuable suggestions were of immense help throughout the project work.

We convey our sincere regards to our respected principal Dr. G.V. Mulgund and Head of Department Dr. G.A. Walikar for their valuable support.

# References

[1] K Mishra, R Rani. 2017. "Customer churn prediction in telecommunication using machine learning." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), August 2017.

[2] Y. Li, G. Xia, 2010. "The explanation of support vector machine in churn prediction." In 2012, International conference on E-Product, E-Service and E-Entertainment, 2010.

[3] U. Yabas, H. Cankaya, T. Ince. 2012. "Customer churn prediction for telecom services." 2012 IEEE 36th Annual computer software and application conference, Nov.2012.

[4] Karthik Revanuru, Kaushik Turlapaty, and Shrisha Rao. 2017. "Neural Machine Translation of Indian Languages". In Compute "17:10th Annual ACM India Compute Conference, November 16–18-2017, Bhopal, India.

[5] Brenda Reyes Ayala, Jiangping Chen,2017. "A Machine Learning Approach to Evaluating Translation Quality", IEEE 2017.

[6] Ibrahim Onuralp Yigit, Hamed Shourabizadeh. "An Approach for Predicting Employee Churn by using data mining", International Artificial Intelligence and data processing Symposium'17, Turkey, IEEE 2017.

 