A New Approach for Customer Churn Prediction in Telecom Industry

Saumya Saraswat
Department of CSE & IT
Madhav Institute of Technology & Science
Gwalior, M.P., India

Akhilesh Tiwari
Department of CSE & IT,
Madhav Institute of Technology & Science
Gwalior, M.P., India

ABSTRACT
Since its inception, the field of Data Mining and Knowledge Discovery from Databases has been driven by the need to solve many practical problems. With the rapid development of telecommunication industry, the service providers are inclined more towards expansion of the subscriber base. To meet the need of surviving in the competitive environment, the retention of existing customers has become a huge challenge. In the survey done in the Telecom industry, it is stated that the cost of acquiring a new customer is far more that retaining the existing one. Therefore, by collecting knowledge from the telecom industries can help in predicting the association of the customers as whether or not they will leave the company. The required action needs to be undertaken by the telecom industries in order to initiate the acquisition of their associated customers for making their market value stagnant. This paper describe a framework that was proposed to conduct for the churn prediction model using Naïve Bayes algorithm for classification task and then apply Elephant Herding Optimization algorithm for solving optimization task. Elephant Herding Optimization is a metaheuristic algorithm. The proposed methodology thereby generates optimal customers who will leave the service provider which is beneficial for any enterprise in the current scenario for effective decision making and perform appropriate steps to retain those customers.

Keywords
Churn, Naïve Bayesian Classification, Elephant Herding Optimization.

1. INTRODUCTION
Data mining is one of the advanced types of analytical tools at this time which is available. These tools can include statistical models, mathematical algorithms and machine learning methods (algorithms that increase their performance automatically during experience, such as neural networks or decision trees), but data analysis system do not handle large amount of data [1]. Data mining involves data collection and management and includes analysis prediction, but this process has limitations on its capacity. The limitation is that it does not tell the user the value or importance of the models and relationships discovered. A second limitation is a causal relationship identified between the connections between behaviors and / or variables. Because this process requires expert technical and logical specialists who can structure the analysis and explain the result that is produced [2]. In the telecommunications sector, huge amount of data are produced and stored, including: Call data describing the calls that traverse the telecommunication networks, Network data concerning the state of hardware and software components and customer related data. Within such huge amount of critical business data, valuable knowledge can be hidden. In fact, in the earlier phase , it has been shown that it is used as

data extraction. Applications include fraud detection in the telecommunications sector, improved market efficiency and detection and fault location. At the same time, continuous work on customer loyalty and abandonment prevention becomes a necessity, because competition has similar acquisition problems. Retention of existing users is important as it is 5 upto 7 times cheaper to keep a consumer than buying a new one [3].

Churn, is a concern for customers who are more expensive than acquiring new customers, known by the company that is trying to solve the problem. Customer cancellation must be indicated when a customer has stopped using their SIM card as soon as possible [4]. There are two main categories of churners that are:
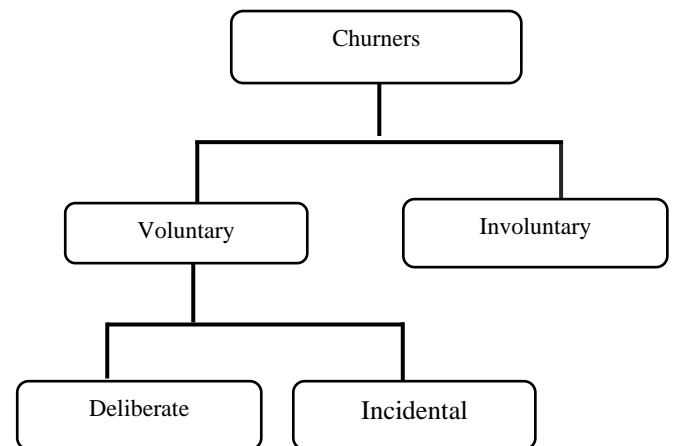

Figure 1: Churn taxonomy

1.1 Involuntary churner:
These are the customers that service providers decide to remove from subscribers list. Therefore' this category includes people that are churned for fraud (customers who cheat), non-payment (customers with credit problem), and underutilization (customers who don't use the phone) [5].

1.2 Voluntary churner:
Voluntary churner is more difficult to determine, it occurs when a customer   makes a decision to terminate his/her service with the provider. When people think about telecom churn it is usually the voluntary kind that comes to mind [5], it can be sub-divided into two main categories:

1.2.1 Incidental churn:
It occurs when something happened in customers lives. For example: change in location churn, change in financial condition, etc.

1.2.2 Deliberate churn:

It happens for reasons of technology (customers wanting newer or better technology), economics (price sensitivity), service quality factors, social or psychological factors, and convenience reasons, and this churn is the problem that most churn management solutions try to solve [5].

This paper presents an supervised learning algorithm for classification which is Naïve Bayes algorithm. This algorithm takes input from the dataset obtained from kaggle for the purpose of scientific research to develop the proposed model and gives output as whether the customer is churner or non churner. Although enough thresholds are applied over the data, the number of churners obtained is still huge. Optimization algorithm are employed so that the resultant are profiting to the service providers. The optimization algorithm used here is EHO (Elephant Herding Optimization).

This paper illustrates: **Section 2**, while presenting the related work that have done in this sector and briefly reviews the naïve bayes and EHO. **Section 3**, While presenting the proposed work for churn prediction using classification and optimization. **Section 4**, while presenting the implementation result and **Section 5**, while presenting the overall result. **Section 6, 7** while presenting future work and conclusion.

2. RELATED WORK

Different data mining techniques have been used by researchers among lots of researches for predicting customer churn in the field of telecommunications. They have explored a number of techniques and applied various algorithms, namely neural network, decision tree, logistic regression, naïve bayes etc. and have also given some suggestions.

1. Essam et al. in [5] have introduced a simple model based on data mining to track customers and their behavior against churn. A dataset of 500 instances with 23 attributes has been used to test and train the model using 3 different techniques i.e., Decision trees, Support Vector Machine(SVM) and Neural networks for classification and k-means algorithm for clustering. Results specify that SVM has been stated as the best apt scheme for predicting churn in telecom

2. KhakAbi et al. [6], provides a brief review of papers from two perspectives: techniques used (model applied) and statistical reports. Through their findings, they have tried to point out the gaps and strengths in the area of churn.

3. Clement et al. [7] used fresh features such as contract-related, call pattern description, and call pattern changes description features resultant from traffic figures and customer profile data. The well-known features were evaluated using Naïve Bayes and Bayesian network and obtained results were compared to results achieved using decision tree. Results have revealed that probabilistic classifiers have shown higher true positive rate than decision tree.

4. Vladislav and Marius in [8] have presented quality measures of churn prediction models including regression analysis, naïve Bayes, decision tree, neural network etc. They have also pointed out the links between churn prediction and customer lifetime value. According to the authors, new prediction facsimiles need to be developed and grouping of proposed techniques can also be used.

5. Saad et al. [9] have applied different machine learning algorithms such as linear and logistic regression, ANN (Artificial Neural Network), K-means clustering, Decision Tree to identify customers who can switch and active customers. The best results were acquired via exhaustive CHAID, a variant of standard decision trees.

6. Ning Lu [10] has proposed a model with an ―Implementation zone where highest probability of churn by customers can be addressed for retentive actions. The author has also proposed a further improvement in performance by analyzing other classification techniques as well or using a hybrid approach for more accurate results

7. Khalida et al. in [11] have used a precise training sample set was used to conduct an tryout on customer churn factor using decision tree. According to the authors, rule information can be easily understood by decision tree. An attempt has been made to implicit various factors responsible for customer churn such as area.

8. John Hadden et al.[12] have shared with the help of a tabular listing (Fig. 7) that the decision tree has come out to be the best technique out of the three may be because of its rule based architecture. And about neural network and regression it is said that these methods required training data and testing data to make calculations and to check for churners and non churners. But for neural and regression the accuracy depends mainly on the weights for the neural network and coefficients for the regression.

9. Amal et al.[13] have reviewed that some techniques generally Decision tree based techniques, neural network trees and Regression techniques are applied in churn prediction. Decision tree based techniques shows better result than all other in terms of accuracy. On the other hand, neural networks outdo other techniques due to size of data sets.

10. Umman Tugba Simsek Gursoy [14] have compared regression techniques with decision tree based techniques. Results have shown that in logistic regression analysis churn prediction accuracy is 66% while in case of decision trees the accuracy measured is 71.76%. Hence decision tree based techniques are superior to predict customer churn in telecom.

11. Subramaniam, Sakthikumar, Arunkumar Thangavelu, and Hemavathy Ramasubbian [15] applied fuzzy multi-criteria classification approach which involves multiple criterion parameter based for identifying the customer churn.

12. Richter, Yossi, Elad Yom-Tov and Noam Slonim [16] have proposed a framework, called Group-First Churn Prediction, which eliminates the priori requirement to know who has recently beaten. The method analyzes customer interactions to predict which groups of subscribers are more prone to turnover, even before a single group member has been defeated.

13. Shyam V.Nath [17] study developed a model with data set that contains of 2000 customers by using Naive Bayes classifier the model was used to predict the future churn for the wireless customers when it tested it showed with 68% accuracy.

14. Sowkarthika B. Akhilesh Tiwari, Uday Pratap Singh[18] proposed algorithm use vague association rule mining and utilizes EHO for generating optimal rule set which is beneficial for effective mining.

15. B.Huang et al., [19] proposed genetic algorithm (NSGA) to find number of features subset in different size and dimension. The experiments were carried out using decision tree C4.5 and results proved that the NSGA algorithm is efficient and successful for churn prediction.

2.1 Naïve Bayes Classification

Naïve Bayes learning produces a probabilistic model of the detected data. Despite its ease, Naïve Bayes has been confirmed to be reasonable with more composite algorithm such as neural network / decision tree in certain domains [20, 21]. Assumed the training set of instances, each is characterised as a vector of features $[x_1, x_2, \dots x_d]$, the assignment is learning from data to be capable to predict most probable class $y_j \in C$ of the new instance whose class is unidentified. Naïve Bayes employs the Bayes's theorem to guess the probabilities of the classes.

$$P(y_j | x_1, x_2, \dots, x_d) = \frac{P(y_j) P(x_1, x_2, \dots x_d | y_j)}{P(x_1, x_2, \dots x_d)} \quad .. \qquad (1)$$

Where $P(y_j)$ is the prior probability of class y_j which is projected as its existence frequency in the training data. $P(y_j | x_1, x_2, \dots x_d)$ is the subsequent probability of class y_j after observing the data. $P(x_1, x_2, \dots x_d | y_j)$ denotes the conditional probability of observing an occurrence with the feature vector $[x_1, x_2, \dots x_d]$ among those having class y_j. And $P(x_1, x_2, \dots x_d)$ is probability of detecting an instance with feature vector $[x_1, x_2, \dots x_d]$ regardless of the class. Since the sum of the subsequent probabilities entirely classes is one $\sum_{y_j \in C} P(y_j | x_1, x_2, \dots x_d) = 1$, denominator on eq. (1)'s right hand side is normalizing factor & can be omitted

$$P(y_j | x_1, x_2, \dots x_d) = P(y_j) P(x_1, x_2, \dots x_d | y_j) \qquad (2)$$

An instance will be labelled as the particular class which has the highest posterior probability y_{MAP}.

$$y_{MAP} = arg\, max_{y_j \in C}\, P(y_j) P(x_1, x_2, \dots x_d | y_j) \qquad (3)$$

In order to estimate the term $P(x_1, x_2, \dots x_d | y_j)$ by counting frequencies, one needs to have a huge training set where every possible combinations $[x_1, x_2, \dots x_d]$ appear many times to obtain reliable estimates [20]. Naïve Bayes solves this problem by its Naïve assumption that features that define instances are conditionally independent given the class. Therefore the probability of observing the combination $[x_1, x_2, \dots x_d]$ is simply the product of the probabilities of observing each individual feature value $P(x_1, x_2, \dots x_d | y_j) = \prod_{i=1}^{d} P(x_i | y_j)$. Substituting this approximation into equation (3) to derive the Naïve Bayes classification rule.

$$y_{MAP} = arg\, max_{y_j \in C}\, P(y_j) \prod_{i=1}^{d} P(x_i | y_j) \qquad (4)$$

As discussed above, for nominal feature, the probability is estimated as the frequency over the training data. For continuous feature, there are two solutions. The first one is to perform discretization on those continuous features, transferring them to nominal ones. The second solution is to assume that they to follow a normal distribution.

The term $P(x_i | y_j)$ is estimated by the fraction $\frac{F(x_i | y_j)}{F(y_j)}$, where $F(y_j)$ is the number of instances in the training set having class y_j, and $F(x_i | y_j)$ is the number of these instances having feature value x_i and class y_j. If the training data doesn't contain any instance with this particular combination of class and feature value, $F(x_i | y_j)$ is zero. The estimate probability according to equation (4) will be zero for every similar cases. To avoid this, a correction called the m-estimate is introduced [22, 23].

$$P(x_i | y_j) = \frac{F(x_i | y_j) + m\, P(x_i)}{F(y_j) + m} \qquad (5)$$

If the prior probability $P(x_i)$ is unknown, uniform distribution is assumed, i.e. if a feature has k possible values, then $P(x_i) = 1/k$. The parameter m can be regarded as the additional m dummy instances appended to the training set.

S. Balaji [24] used Naive Bayesian Classification algorithm for customer classification and to predict churners that are churned on Life Insurance sector. He also used Naïve Bayes classification to classify the customers from larger dataset. It also analyses the issues of using data mining technology for predicting the customer habits. In this analysis, they had tested 10,000 sample of Life Insurance of customers, the unprocessed data can be converted into useful information and then into knowledge for which they had used predictive data mining techniques .Posterior classification process applied for the data in this paper. It proved that the naïve bayes classifier is much better than other classifier for conducting the policy preferences towards the customers. This helps us to raise the income of the organization.

2.2 Elephant Herding Optimization

Elephant herding optimization algorithm is one of the recent swarm intelligence algorithm that was proposed in 2016 by Wang et. al. [25]. Algorithm was inspired by herding behavior of elephants. Complex behavior was simplified for purpose of swarm intelligence algorithm. There are two operations performed in EHO. One is **clan updation** and the other is **clan separation**.

2.2.1 Clan Updation Operation:

The elephant population is comprised of several groups, and each group has finite number of elephants. The elephants in every group or called it clan, stay together under the governance of matriarch. In general, matriarch in every group is the oldest one, and it is considered as the fittest elephant individual in this clan for the optimization problem.

2.2.2 Clan Separation Operation:

In each elephant group, male elephant will leave their family group and stay alone as they mature enough from the leading elephant groups at each generation.

In order to make elephant's grazing behavior solve all kinds of global optimization problems, here preference will be given to simplify it in the following idealized rules:

- The elephant population is composed of some clans, and each clan has fixed number of elephants.
- A fixed number of male elephants will leave their family group and live solitarily far away from the main elephant group at each generation.
- The elephants in each clan live together under the leadership of a matriarch.

2.2.3 EHO Algorithm:

Step 1: Initialize the population.

Step 2: Repeat

Step 3: Sort all the elephants according to their fitness value which is calculated using Fitness Function.

Step 4: Apply clan updation operation. Each elephant j of clan i moves according matriarch where matriarch is the elephant c_i with the best fitness value. The updated position is calculated as

$$x_{new, c_{i,j}} = x_{c_{i,j}} + \alpha * \left(x_{best, c_i} - x_{c_{i,j}} \right) * r \qquad \dots\dots (7)$$

Where $x_{new,c_{i,j}}$ represent new position of elephant j in clan i, $x_{c_{i,j}}$ represent old position of elephant j in clan i, x_{best,c_i} represent position with best fit (matriarch), $\alpha \epsilon [0,1]$ is algorithm's parameter, $r \epsilon [0,1]$ is the random number used to improve diversity of population in the later stages of algorithm. Position update for the best fit in the clan x_{best,c_i} is calculated as:

$$x_{new,c_i} = \beta * x_{center,c_i} \qquad ....(8)$$

Where $\quad x_{center,c_i} = \frac{1}{n_{c_i}} \sum_{j=1}^{nc_i} x_{c_{i,j,d}}$, $\beta \varepsilon [0,1]$ is second parameter of the algorithm, x_{center,c_i} represent the center of the clan c_i, n is the number of elephants in clan c_i.

Step 5: Apply clan separation operation. In each clan i some number of the elephants with the worst value are moved to the new position according to following equation

$$x_{worst,c_i} = x_{min} + (x_{max} - x_{min} + 1) * rand \qquad (9)$$

Where rand $\varepsilon [0,1]$, x_{min} is the lower bound of the search space, x_{max} is the upper bound of the search

space.

Step 6: Evaluate population by the newly updated position.

Step 7: Untill the stop criteria.

Step 8: Return the best solution among the population.

3. PROPOSED WORK

The main objective of this paper is to predict whether the customer will churn or not. And to prevent customers from switching necessary option is that, to extend the packages or offering new incentives or services for those customers who are likely to switch to another service provider from the current service provider [5]. Because these services which is provided by the company mainly generate revenue. Data mining techniques have been broadly used to develop model of churn prediction. In this paper, we use classification algorithm with optimization algorithm to find out the optimized customers.

3.1 Dataset

Telecommunication data faces a large number of strict policies and rules that make access to data extremely difficult. For these reasons, we used the dataset which was obtained from the www.kaggle.com and it contains 21 attributes and 3333 instances. Churn in Telecom dataset attributes are as follows:

Table 1: Schema of Churn Dataset

S.no	Attributes	Type
1.	State	String
2.	Account Length	Numeric
3.	Area Code	Numeric
4.	Phone Number	String
5.	International Plan	String
6.	Voice Mail Plan	String
7.	Number of Voice Mail Messages	Numeric
8.	Total Day Minutes	Numeric
9.	Total Day Calls	Numeric
10.	Total Day Charges	Numeric
11.	Total Evening Minutes	Numeric
12.	Total Evening Calls	Numeric
13.	Total Evening Charges	Numeric
14.	Total Night Minutes	Numeric
15.	Total Night Calls	Numeric
16.	Total Night Charges	Numeric
17.	Total Initial Minutes	Numeric
18.	Total Initial Calls	Numeric
19.	Total Initial Charges	Numeric
20.	Customer Service Calls	Numeric
21.	Churn	Boolean

3.2 Algorithm:

Input: Telecom churn dataset

Output: Optimized churners from the total churners

Step 1: Apply the classification Naïve Bayesian Algorithm.

Step 2: Compare the posterior probability of classes and highest one is choosen as predicted output.

Step 3: Apply Elephant Herding Optimization.

Calculation of fitness function is given by:

$$f(x) = \frac{\alpha * total\ day\ charge\ +\ \beta * total\ day\ minute\ +\ \gamma * customer\ service\ calls}{\alpha + \beta + \gamma}$$

………(6)

Where $\alpha, \beta, \gamma\ \varepsilon\ [0,1]$
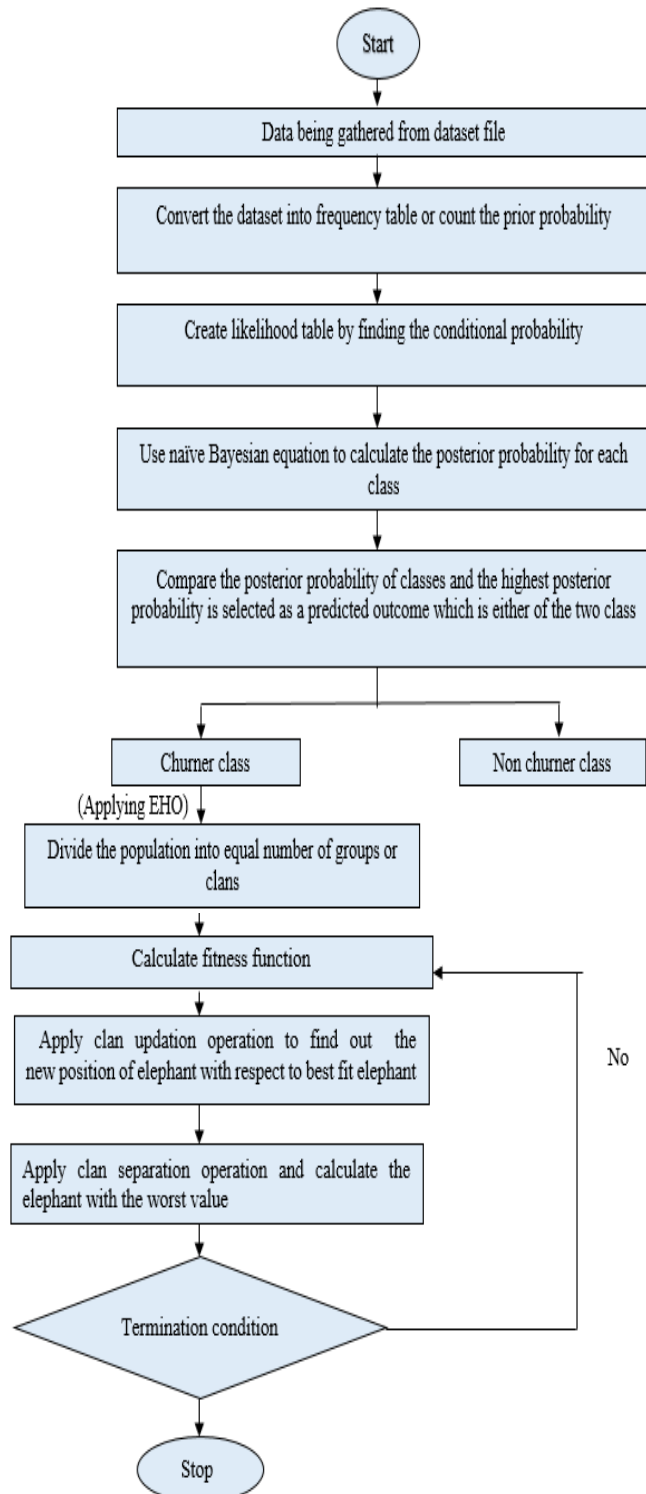
3.3 Churn Prediction Framework

Figure 2: Flow chart of proposed work

4. IMPLEMENTATION RESULT

The model in this paper is implemented on eclipse software and it is implemented using Naïve Bayesian algorithm used for classification and Elephant Herding Optimization for optimization. The snapshots shown below represent the output of prediction and optimized output.

Confusion matrix is useful tool for analyzing how well your classifier can recognize tuples of different classes. True Positive (TP) and True Negative (TN) tell us when the classifier is getting things right, while False Positive (FP) and False   Negative (FN) tell us when the classifier is getting things wrong, P' is the number of tuples that were labeled as positive (TP+FP) and N' is the number of tuples that were labeled as negative (TN+FN) [26].

Figure 3: Depicts implementation of Naïve Bayesian

```
66    System.out.println("kappa = "+eval.kappa());
67    System.out.println("MAE = "+eval.meanAbsoluteError());
68    System.out.println("RMSE = "+eval.rootMeanSquaredError());
69    System.out.println("RAE = "+eval.relativeAbsoluteError());
70    System.out.println("RRSE = "+eval.rootRelativeSquaredError());
71    //System.out.println("Precision = "+eval.precision(1));
72    //System.out.println("Recall = "+eval.recall(1));
73    System.out.println("fMeasure = "+eval.fMeasure(1));
74    System.out.println("Error Rate = "+eval.errorRate());
75    //the confusion matrix
76    System.out.println(eval.toMatrixString("=== Overall Confusion Matrix =
```

Problems @ Javadoc Declaration Console

<terminated> nbcombeho [Java Application] C:\Program Files\Java\jre1.8.0_161\bin\javaw.exe (Jul 28, 2018, 3:07:50 PM)

```
Correct % = 87.81878187818782
Incorrect % = 12.181218121812181
AUC = 0.8237746540263703
kappa = 0.47604809894853256
MAE = 0.1902462317963338
RMSE = 0.3074199676168866
RAE = 76.67173390708953
RRSE = 87.33146152066952
fMeasure = 0.54586129753915
Error Rate = 0.12181218121812182
=== Overall Confusion Matrix ===

    a    b   <-- classified as
 2683  167 |   a = False
  239  244 |   b = True

optimized instances = 199
```

Figure 4: Depicts the result with Naïve Bayesian and EHO

5. RESULT

The result are shown as

Table 2: Resultant confusion matrix

Classes	Non Churn	Churn	Total
Non Churn	2683	167	2850
Churn	239	244	483
Total	2922	411	3333

It describes confusion matrix there are two classes which are Non-churn and Churn, so:

True Positive (TP) instance that classified correctly as **Non-churn** = 2683

True Negative (TN) instance that classified correctly as **Churn** = 244

False Positive (FP) instance that classified incorrectly as **Non-churn** = 239

False Negative (FN) instance that classified incorrectly as **Churn** = 167

The number of tuples labeled as **positive P'** = TP + FP = 2922

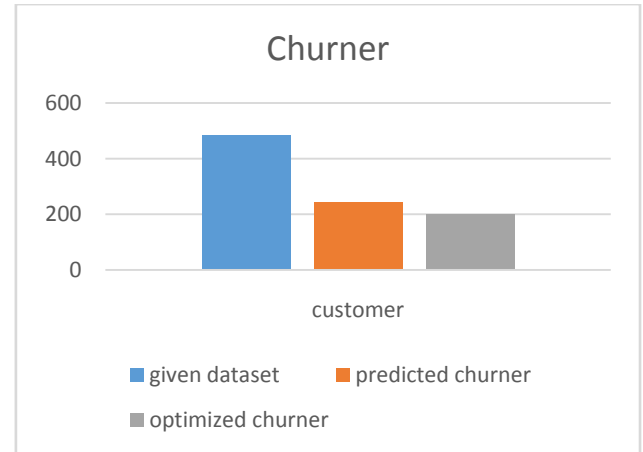The number of tuples labeled as **negative N'** = TN + FN = 411

Figure 5: System Result for Churner

After prediction 244 correctly predicted as churner customer from 483 using naïve equation 1-5. And after applying Elephant Herding Optimization Algorithm 199 churner customer are find as optimized churner from 244 who can definitely churn using equation 6-9.
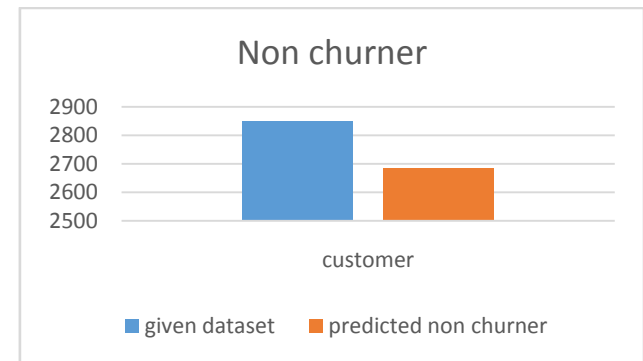

Figure 6: System Result for Non Churner

After prediction we have 2683 correctly predicted as Non-churner customer from 2850. And optimization is only performed to find out churner customers. Because non churner customer are not switching, they are the customer of present service provider and satisfied with the services provided.

With the help of given confusion matrix obtained after experiment accuracy can be calculated as:-

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = 87\%$$

6. FUTURE WORK

In the current framework feature selection algorithm is not used, so in future work will be done on the mechanism which would be able to choose feature in order to improve the churn accuracy and only important features are selected. For this purpose, requirement is to append selection method just above proposed framework. This will help in finding the important features. And the next limitation is that it is a huge success if a model can predict churners beforehand although it is better to know when customers churn. Experimented dataset do not represent current situation since it is retrieved from established repository due to business confidentiality and privacy.

Future research work will also consider combining hybrid data mining approaches such as clustering and classification techniques and try and improve the performance of the

algorithms. And other prediction techniques such as vector machines etc. and optimization technique such as genetic algorithm etc. can be applied. The current abandonment forecasting methodology can be tested for other sectors, such as banks, insurance or medical care, and prediction comparisons can be made and to find the optimized result.

7. CONCLUSION

Telecommunication industry has suffered from high churn rates and immense churning loss. Although the business loss is unavoidable, but still churn can be managed and kept in an acceptable level. Good methods need to be developed and existing methods have to be enhanced to prevent the telecommunication industry to face challenges. In this paper description of a new approach for prediction of churner and among those churner finding the optimized output by using EHO algorithm to provide benefit to the service provider by finding the actual optimized churner.

8. REFERENCES

[1] H. Newton, Newton's Telecom Dictionary, CMP Books,

[2] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996), Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the 2nd ACM international conference on knowledge discovery and data mining (KDD), Portland, OR, pp 82–88

[3] Sheth, Jagdish, and Rajendra Sisodia. - The 4 A's of Marketing: Creating Value for Customer,Company and Society. Routledge, 2012.

[4] L.J.S.M. Alberts, 2006. Churn prediction in the mobile telecommunications industry

[5] Essam Shaaban, Yehia Helmy, Ayman Khedr, Mona Nasr, "A proposed model of prediction of abandonment", International Journal of Engineering and Applications Research (IJERA) ISSN: 2248-9622 Vol. 2, Issue 4, June-July 2012.

[6]S. KhakAbi, M. Namvar, Mohd R. Gholamian,"Data Mining Applications in Customer Churn Management", Proc. Of IEEE international conference on Intelligent Systems, Modelling and Simulation, 2010.

[7] Clement Kirui, Li Hong, Wilson Cheruiyot, Hillary Kirui, Predict the rotation of customers in the mobile sector using probabilistic classifiers in Data Mining, the International Journal of Computer Science.

[8] Vladislav Lazarov, Marius Capota, ―Churn Prediction‖, Technische Universität Munchen.

[9] Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal, proposed a model of forecast of the abandonment of telecommunications subscribers using machine learning, IEEE, 2013.

[10] Ning Lu, Hua Lin, Jie Lu, ―A Customer Churn Prediction Model in Telecom Industry Using Boosting‖, IEEE Transactions on Industrial Informatics.

[11]Khalida, Sunarti, Norazrina, Faizin, ―Data Mining in Churn Analysis Model for Telecommunication Industry Journal of Statistical Modeling and Analytics.

[12] Hadden J., Tiwari A., Roy R., Ruta D. "Churn prediction: Does technology matter?" International Science Index, Engineering and Technology, 2008.

[13] Amal M. Almana, Mehmet Sabih Aksoy, Rasheed Alzaharni, "Survey of Data Mining Techniques in the Analysis of Customer Abandonment for the Telecommunications Industry", International Journal of Research and Engineering Applications.

[14] Umman Tugba Şimsek Gursoy, "Analysis of the abandonment of customers in the telecommunications sector", Journal of Istanbul University School of Business Administration.

[15] Subramaniam, Sakthikumar, Arunkumar Thangavelu and Hemavathy Ramasubbian. "Fact-AnAdaptive method to predict client abandonment rate using the widespread multi-criterion classification method for decision-making". Asian Journal of Science and Technology 4.11 (2013): 227-233.

[16] Richter, Yossi, Elad Yom-Tov, and Noam Slonim. "Predicting Customer Churn in Mobile Networks through Analysis of Social Groups." *SDM*. 2010.

[17] Shyam V.Nath, R.S.B. Customer churn analysis in the wireless industry: A data mining approach.

[18]Sowkarthika B., Akhilesh Tiwari, Uday Pratap Singh - Elephant Herding Optimization based Vague Association Rule Mining Algorithm, International Journal of Computer Applications ,Volume 164 – No 5, April 2017

[19] Bingquan Huang, B.Buckley, T.M.Kechadi, "Selection of multiple features through the use of NSGA-II for prediction of customer abandonment in telecommunications", Expert Systems with Applications 37 (2010) 3638-3646.

[20]Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Science/Engineering/Math.

[21] George, H. J., & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. Proceeding of the Eleventh Conference on Uncertainty in Artificial Intelligence (pp. 338-345). San Mateo: Morgan Kaufmann.

[22] Gutkin, M. (2008). Feature selection methods for classification of gene expression profiles. Tel-Aviv University.

[23] Cetnik, B. (1990). Estimating Probabilities: A crucial task in machine learning. Ninth European Conference on Artificial Intelligence, (pp. 147-149). London.

[24] S. Balaji, S.K. Srinivasta, -used Naïve Bayes Life Insurance Classification Approach Database for the actual prediction of customer preferences in life insurance products, International Journal of Computer Applications, Vol.51, No. 3, 2012.

[25]G.-G. Wang, S. Deb, X.-Z. Gao, and L. D. S. Coelho, "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour," International Journal of Bio-Inspired Computation, vol. 8, no. 6, pp. 394–409, 2016.

[26] MD.Rashid Farooqi, K. R. 2011. A comprehensive study of CRM through data mining techniques. Proceedings of the National Conference.