Knowledge Based Customer Churan Prediction For Telecom Services Using Python

A project report submitted for the partial fulfillment of the

Bachelor of Technology Degree

in

Computer Science & Engineering

under

Maulana Abul Kalam Azad University of Technology

by

Pranjal Chowdhury

Roll No: 10400116145, Registration Number: 161040110102

Partho Protim Sarkar

Roll No: 10400116150, Registration Number: 161040110097

Academic Session: 2016-2020

Under the Supervision of

Prof. Anupam Mondal

Department of Computer Science and Engineering
Institute of Engineering & Management

Y-12, Salt Lake, Sector 5, Kolkata, Pin 700091, West Bengal, India Affiliated To

Maulana Abul Kalam Azad University of Technology

BF 142, BF Block, Sector 1, Kolkata, West Bengal 700064

May 2020

INSTITUTE
OF ENGINEERING & MANAGEMENT
श्रद्धावान लभते ज्ञानम् Salt Lake Electronics Complex, Kolkata - 700091, WB, INDIA

Phone : (033) 2357 2969/2059/2995
(033) 2357 8189/8908/5389
Fax : 91 33 2357 8302
E-mail : director@iemcal.com
Website : www.iemcal.com

CERTIFICATE

TO WHOM IT MAY CONCERN

This is to certify that the project report titled **"Knowledge Based Customer Churan Prediction For Telecom Services Using Python"**, submitted by **Pranjal Chowdhury, Roll No: 10400116145, Registration Number: 161040110102**, **Partho Protim Sarkar, Roll No: 10400116150, Registration Number: 161040110097** students of **Institute of Engineering & Management** in partial fulfillment of requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, is a bona fide work carried out under the supervision of **Prof. Anupam Mondal** during the final year of the academic session of 2016-2020. The content of this report has not been submitted to any other university or institute for the award of any other degree.

It is further certified that the work is entirely original and the performance has been found to be satisfactory.

Prof. Anupam Mondal

Assistant Professor

Department of Computer Science and Engineering
Institute of Engineering & Management

Prof.(Dr.) Sourav Saha

H.O.D.

Department of Computer Science and Engineering
Institute of Engineering & Management

Prof.(Dr.) Amlan Kusum Nayak

Principal

Institute of Engineering & Management

Gurukul Campus: Y-12, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, Phone: (033) 2357 2969
Management House: D-1, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, Phone: (033) 2357 8908
Ashram Building: GN-34/2, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, Phone: (033) 2357 2059/2995

INSTITUTE OF ENGINEERING & MANAGEMENT

DECLARATION

FOR NON-COMMITMENT OF PLAGIARISM

We, **Pranjal Chowdhury**, **Partho Protim Sarkar**, students of B.Tech. in the Department of Computer Science and Engineering, Institute of Engineering & Management have submitted the project report in partial fulfillment of the requirements to obtain the above noted degree. We declare that we have not committed plagiarism in any form or violated copyright while writing the report and have acknowledged the sources and/or the credit of other authors wherever applicable. If subsequently it is found that we have committed plagiarism or violated copyright, then the authority has full right to cancel/reject/revoke our degree.

Name of the Student: PRANJAL CHOWDHURY

Full Signature: ____________________

Name of the Student:PARTHO PROTIM SARKAR

Full Signature:____________________

Date:____________________________

Abstract:

With the rapid growth of digital systems and associated information technologies, there is an emerging trend in the global economy to build digital customer relationship management (CRM) systems. Customer churn prediction is a main feature of in modern telecom communication CRM systems. The eight prediction techniques (Logistic Regression, Support Vector Machine, Random Forest, Perceptron, ANN, KNN, Decision Tree & Linear Regression) are applied in customer churn as predictors, based on the new features. The experimental results show that the new features with the seven modelling techniques are more effective than the existing ones for customer churn prediction in the telecommunication service field.

Acknowledgements

We must not forget to acknowledge everyone who has provided constant support to us during our B.Tech course. First and foremost, we would like to express sincere gratitude to our supervisor **Prof. Anupam Mondal** for his continuous support and motivation in fueling the pursuance of carrying out this project endeavor. Without his guidance and persistent encouragement, this project work would not have been possible. He has been a tremendous mentor for us throughout this academic journey. Many of his academic advises about our career growth have been priceless.

We would like to convey sincere gratitude to **Prof. Himadri Nath Saha** for providing us constant inspiration to stand firm against several setbacks throughout the course. Additionally, we would like to thank all the technical, non-technical and office staffs of our department for extending facilitating cooperation wherever required. We also express gratitude to all of our friends in the department for providing the friendly environment to work on the project work.

We would also like to thank our Director **Prof. Satyajit Chakraborti** for providing us an outstanding platform in order to develop our academic career. In addition, we also preserve a very special thankful feeling about our Principal **Prof. Amlan Kusum Nayak** for being a constant source of inspiration.

A special thank is due to our family. Words cannot express how grateful we are to our parents for all the sacrifices that they have made while giving us necessary strength to stand on our own feet.

Finally, we would like to thank everybody who has provided assistance, in whatever little form, towards successful realization of this project but with an apology that we could not mention everybody’s name individually.

Contents

1. Introduction
   I. Motivation
   II. Related Work
2. Methodology
   I. Logistic Regressions (LR)
   II. Decision Trees
   III. Random forests
   IV. k Nearest Neighbor Classifier
   V. Linear Classifiers (LC)
   VI. Artificial neural networks
   VII. Support Vector Machines (SVM)
   VIII. Perceptron
3. Experimental Results and Discussion
   I. ExperimentalSetup
   II. ExperimentalResults
   III. Performance Evaluation of Experiment

4. Conclusion

5. Bibliography

1. Introduction

Motivation

The service companies of telecommunication service businesses in particular suffer from a loss of valuable customers to competitors; this is known as customer churn. In the last few years, there have been many changes in the telecommunications industry, such as, the liberalisation of the market opening up competition in the market, new services and new technologies. The churn of customers causes a huge loss of telecommunication services and it becomes a very serious problem. Recently, data mining techniques have emerged to tackle the challenging problems of customer churn in telecommunication service field. As one of the important measures to retain customers, churn prediction has been a concern in the telecommunication industry and. Over the last decade, the majority of churn prediction has been focused on voice services available over mobile and fixed-line networks. As mentioned in Luo et al and Zhang et al, in contrast to the mobiles services, there are less researchers to investigate the churn prediction for the land-line telecommunication services. Generally, the features used for churn prediction in mobile telecommunication industry includes customer demographics, contractual data, customer service logs, call details, complaint data, bill and payment information. In contrast to the mobiles services, there are less amounts of qualified information for land-line services providers. The data of land-line communication services is different to mobile services. Some of this data is missing, less reliable or incomplete in land-line communication service providers. For instances, customer ages and complaint data, fault reports are unavailable and only the call details of a few months are available. Due to business confidentiality and privacy, there are no public datasets for churn prediction. For churn prediction in the land-line telecommunication service field, Luo et al presented a set of features, which are the duration of service use, payment type, the amount and structure of monthly service fees, Proportions variables, consumption level rates variables and the growth rates of the second three months. Recently, Huang, Kechadi, and Buckley presented a set of features, including one sixmonth Henley segmentation, line-information, bill and payment information, account information, call details and service log data, etc. In addition, most of the literature shows the features that are the aggregated call-details are important for the customer churn prediction. These features are obtained by aggregating the duration, fees and the number of calls for any types of calls for each period. However, the call details can be further divided into more precise information, according to different types of calls. This more precise information might be more useful than the existing features of call details for churn prediction. In order to improve the accuracy of customer churn prediction in telecommunication service field, we present a new set of features with seven modelling techniques in this paper. The new features are the 2 six-month Henley segmentation, precise 4- month call details, information of grants, line information, bill and payment information, account information, Demographic profiles and service orders that are extracted from existing limited information. The modelling techniques are Logict Regression, Naive Bayes, Linear classifiers, Decision Tree , Multilayer perceptrons artificial neural networks, Support Vector Machines and the Evolutionary Data Mining Algorithm. Finally, the comparative experiments are carried out. The experimental results show that the presented features and eight modelling techniques are more effective than the existing features for the customer churn prediction in land-line communicational services. The rest of this paper is organised as following: next section introduces the evaluation criteria of churn prediction systems. our methodology which includes the techniques of feature extraction, normalisation and prediction.

Related Work

AUTHOR NAME:

Nabgha Hashmi, Naveed Anwer Butt and Dr.Muddesar Iqbal

PAPER NAME:

Customer Churn Prediction in Telecommunication :

A Decade Review and Classification

CONTRIBUTION:

Research make a contribution in the field of customer churn predictive modeling in telecommunication. Thus the paper draws a sketch line for the researchers for reviewing and accumulation of the trends about data mining applications in the field of telecommunication.

CHALLENGES:

The classification techniques are good for analyzing qualitative and continuous data and afterwards interpreting results but these techniques do not guarantee the appropriable accuracy of prediction model for large enough, highly dimensional, non linear or time series datasets.

AUTHOR NAME:

IRFAN ULLAH , BASIT RAZA , AHMAD KAMRAN MALIK , MUHAMMAD IMRAN , SAIF UL ISLAM , AND SUNG WON KIM

PAPER NAME:

A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector

CONTRIBUTION:

A customer churn model is provided for data analytics and validated through standard evaluation metrics. The obtained results show that our proposed churn model performed better by using machine learning techniques

CHALLENGES:

Investigate eager leaning and lazy learning approaches for better churn prediction. The study can be further extended to explore the changing behavior patterns of churn customers by applying Artificial Intelligence techniques for predictions and trend analysis.

AUTHOR NAME:

Adnan Amin, Sajid Anwar, Awais Adnan,Muhammad Nawaz , Khalid Alwaif, Amir Hussain, Kaizhu Huang

PAPER NAME:

Customer churn prediction in the telecommunication sector using a rough set approach

CONTRIBUTION:

Proposed an intelligent rule-based decision-making technique, based on rough set theory (RST), to extract important decision rules related to customer churn and non-churn. The proposed approach performed classification of churn from non-churn customers, along with prediction of those customers who will churn or

may possibly churn in the near future.

CHALLENGES:

In this study the profiles of predicted customer churns

were not considered. churn datasets exhibit the class imbalance problem; whereby, the churn class contains fewer number of samples as compared to the non-churn class and eliminating and detecting of outliers would greatly contribute to providing better results.

AUTHOR NAME:

Georges D. Olle Olle and Shuqin Cai

PAPER NAME:

A Hybrid Churn Prediction Model in Mobile Telecommunication Industry

CONTRIBUTION:

Presented a new hybrid model for Churn prediction that predict customers with high propensity to churn, profiling the reason of churn and examining the gap between the churn decision and the deactivation time. In theory it contributes to the problem of increasing TP and decrease FP for more accuracy on the prediction by using a hybrid model.

CHALLENGES:

The evaluation of the model shows that its accuracy is higher than when using single model and that the results could be ameliorate when the data distribution are less skewed.

2.Methodology

Logistic Regressions (LR)

Logistic regression (Hosmer & Lemeshow, 1989) is a widely used statistical modelling technique for discriminative probabilistic classification. Logistic regression estimates the probability of a certain event taking places. The model can be written as:

$$prob(y=1) = \frac{e^{\beta_0 + \sum_{k=1}^{K} \beta_k x_k}}{1 - e^{\beta_0 + \sum_{k=1}^{K} \beta_k x_k}}$$

where Y is a binary dependent variable which presents whether the event occurs (e.g. y = 1 if event takes place, y = 0 otherwise), x1,x2,...,xK are the independent inputs. b0,b1,...,bK are the regression coefficients that can be estimated by the maximum likelihood method, based on the provided training data. The details of the logistic regression models can be found in Hosmer and Lemeshow (1989)

Decision Trees

A method known as ‘‘divide and conquer’’ is applied to construct a binary tree. Initially, the method starts to search an attribute with best information gain at root node and divide the tree into sub-trees. Similarly, the sub-tree is further separated recursively following the same rule. The partitioning stops if the leaf node is reached or there is no information gain. Once the tree is created, rules can be obtained by traversing each branch of the tree. The details of Decision Trees based on C4.5 algorithm are in literature (Quinlan, 1993, 1996).

Random forests

With regard to binary classification tasks, decision trees (DT) have become very popular, thanks to their ease of use and interpretability as well as their ability to deal with covariates measured at different measurement levels (including nominal variables). Nevertheless, conventional decision trees techniques also have their disadvantages. For instance, Dudoit, Fridlyand, and Speed mention their lack of robustness and the suboptimal performance. Fortunately, many of these disadvantages have been dealt with by some researchers who optimized the DT technique. More specifically, the creation of an ensemble of trees followed by a vote for the most popular class, labeled forests, is the result of such a DT optimization. In this paper, we also use the more advanced DT technique. We select the random forests as proposed by Breiman, which uses the strategy of a random selection of a subset of m predictors to grow each tree, where each tree is grown on a bootstrap sample of the training set. This number, m, is used to split the nodes and is much smaller than the total number of variables available for analysis. Since its introduction, random forests have been enjoying increased popularity. The number of applications in fields with large datasets is growing: e.g. in bioinformatics. On the other hand, the number of applications in economics, and, more specifically in marketing related issues are rather scarce. The available applications using random forests reveal that the predictive performance is among the best of available techniques. Furthermore, an interesting by-product of the technique are the produced importance measures for each variable that indicate which variables have the strongest impact on the dependent variables of investigation. Another advantage of the technique concerns the consistent high and robust performance results. Finally, the random forests as proposed by Breiman have reasonable computing times and are easy to use; the only two parameters a user of the technique has to determine are the number of trees to be used and the number of variables (m) to be randomly selected from the available set of variables. In both cases, we follow Breiman's recommendation to pick a large number for the number of trees to be used, as well as the square root of the number of variables for the latter parameter. Since the number of explanatory variables equals to 30 in this study, we fix the number of variables to six.

k Nearest Neighbor

k Nearest Neighbor (or kNN) is a supervised machine learning algorithm useful for classification problems. It calculates the distance between the test data and the input and gives the prediction according.

kNN calculates the distance between data points

$$\begin{aligned}\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) &= \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} \\ &= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.\end{aligned}$$

The above formula takes in n number of dimensions or here we can say them as our features in machine learning. The data point which is located at the minimum distance from the test point is assumed to belong to the same class.

The above formula works the same in n number of dimensions and therefore it can be used with n number of features.

Linear Classifiers (LC)

A linear classifier maps a feature space X into a set of class labels Y by a linear combination function. Usually, a liner classifier f(x) can be written as follows:

$$f(\vec{x}) = sgn\left(\sum_i w_i x_i + b\right)$$

where wi 2 R are the weights of the classifiers and b 2 R is constant. The value of fð Þ ~x for input vector ~x determines the predicted class label. For example, in binary classifications, the class label is +1 if fð Þ ~x P 0. Otherwise, the class label is 1. The weights wi and constant b can be learned from a set of labelled training samples. The details of linear classifiers can be found in literature (Vapnik, 1998).

Artificial neural networks

A Multilayer Perceptron Neural Networks (MLP) is a supervised feed-forward neural network and usually consists of input, hidden and output layers. Normally, the activation function of MLP is a sigmoid function. If an example of MLPs with one hidden layer, the network outputs can be obtained by transforming the activation functions of the hidden unit using a second layer of processing elements, written as follows:

$$Output_{net}(j) = f\left(\sum_{l=1}^{L} w_{jl} f\left(\sum_{i}^{D} w_{li} x_i\right)\right), \quad j = 1, \ldots, J$$

where D, L and J are total number of units in input, hidden and output layer, respectively, and f is a activation function. The back-propagation (BP) or quick back-propagation learning algorithms would be used to train MLP. More details of MLP can be found on Rumelhart, Hinton, and Williams (1986).

Support Vector Machines (SVM)

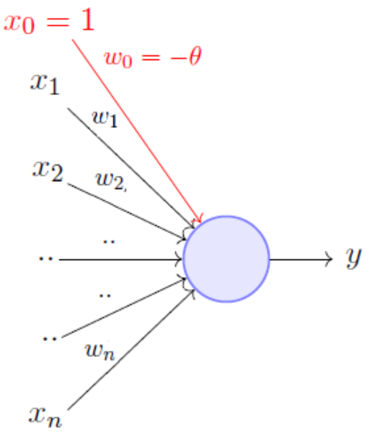
A SVM classifier can be trained by finding a maximal margin hyper-plane in terms of a linear combination of subsets (support vectors) of the training set. If the input feature vectors are nonlinearly separable, SVM firstly maps the data into a high (possibly infinite) dimensional feature space by using the kernel trick (Boser, Guyon, & Vapnik, 1992), and then classifies the data by the maximal margin hyper-plane as following:

$$f(\vec{x}) = sgn\left(\sum_{i}^{M} y_i \alpha_i \phi(\vec{x_i}, \vec{x}) + \delta\right)$$

where M is the number of samples in the training set, x ~i is a support vector with ai > 0, / is a kernel function, ~x is an unknown sample feature vector, and d is a threshold. The parameters {ai} can be obtained by solving a convex quadratic programming problem subject to linear constraints (Burges, 1998). Polynomial kernels and Gaussian radial basis functions (RBF) are usually applied in practise for kernel functions. d can be obtained by taking into account the Karush–Kuhn–Tucker condition (Burges, 1998), and choosing any i for which ai > 0 (i.e. support vectors). However, it is safer in practise to take the average value of d over all support vectors.

Perceptron

The perceptron model is a more general computational model than McCulloch-Pitts neuron. It takes an input, aggregates it (weighted sum) and returns 1 only if the aggregated sum is more than some threshold else returns 0. The threshold as shown above and making it a constant input with a variable weight



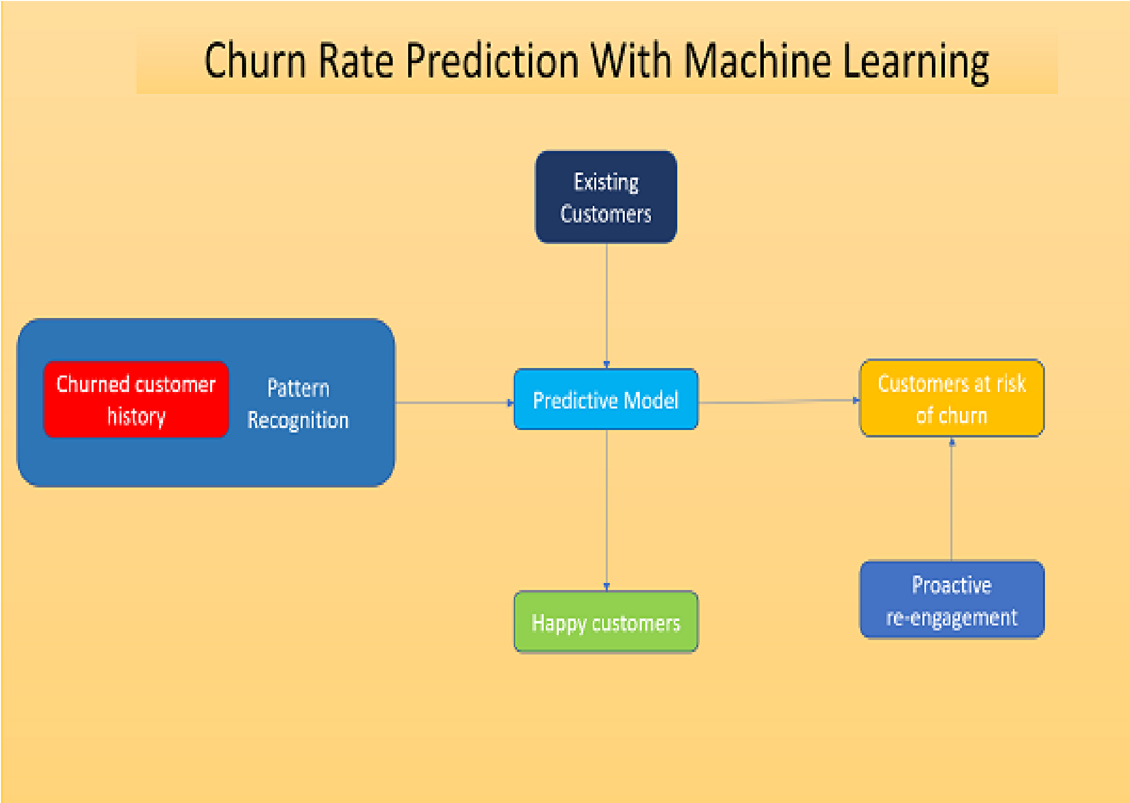
A more accepted convention,

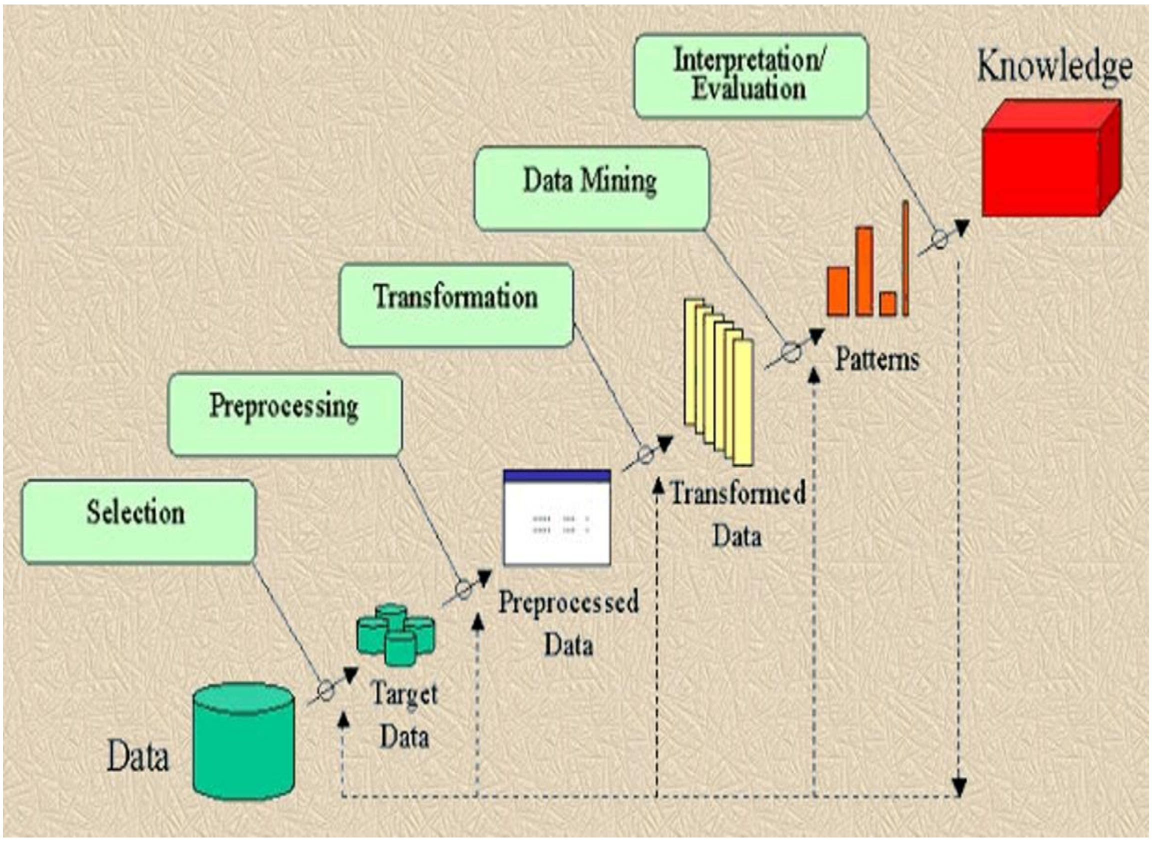
$$y = 1 \quad if \sum_{i=0}^{n} w_i * x_i \geq 0$$

$$= 0 \quad if \sum_{i=0}^{n} w_i * x_i < 0$$

$$where, \quad x_0 = 1 \quad and \quad w_0 = -\theta$$

A single perceptron can only be used to implement linearly separable functions. It takes both real and boolean inputs and associates a set of weights to them, along with a bias (the threshold thing I mentioned above).


Churn Rate Prediction With Machine Learning
Existing Customers
Churned customer history
Pattern Recognition
Predictive Model
Customers at risk of churn
Happy customers
Proactive re-engagement


Interpretation/
Evaluation
Knowledge
Data Mining
Transformation
Patterns
Preprocessing
Transformed
Data
Selection
Preprocessed
Data
Target
Data
Data

Flowchart

3.Experimental Results and Discussion:

ExperimentalSetup:

In order to evaluate the performance of the proposed work for the particular perpose and the performance of prediction , experiments have been conducted based on the database . The dataset comprises the 7034 rows and 21colum which has some integer value, some float and string value.Take the dataset as csv format for the importing the data for performing the experiment . For the first and the second sets of experiments, eight modelling techniques were used to make prediction for each set of features. However, for the third set of experiments, the above seven modelling techniques were used to predict or classify the behaviours of customers. However, the feature sets used in the third set of experiments is high dimensional and might cause this problem.

ExperimentalResults:

Presents the retrieval result of the proposed work based on a set of sample data from the test dataset. It is evident from the observation that the proposed work is able to retrieve most relevant data against a test data with in the top most retrievals based on the illustrated similarity score ranking applied on respective feature code so graph ages. Respective feature code so ages. presents the result. There the data prediction perform with a big data and this is able to predict new data .it able to predict the most probability of churn of the customer.

Performance Evaluation of Experiment:

The performance of the experiment is indicate by the graph s and the model result value. the model value means the prediction and the accuracy of the particular models.it can be shown with in the graph or the value based on the data set. Here we also used several models for the prediction the data most of the models are performed well but some model are not worked well for the resion of big dataset, like linear regression,

In fig 1, the data is correlated we found the correlation of every column of data. In fig 2 express the customer churn trough a pie chart. fig 3 execute the bar chart of the churn rate by gender.fig4 is same like tech support and so one and then the there is a scatter chart with tenure and churn rate . then applying the machine learning predicate the data and get accuracy of the the model. Here is shown that the data Frame of the score and next data frame is prediction of customer

.here is the list of that customer who have the most probability of churn with the percentage value.

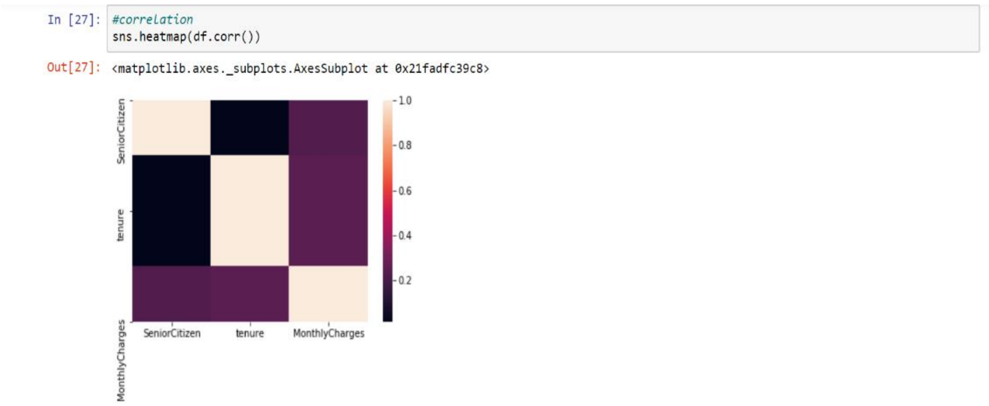


Fig: correlation

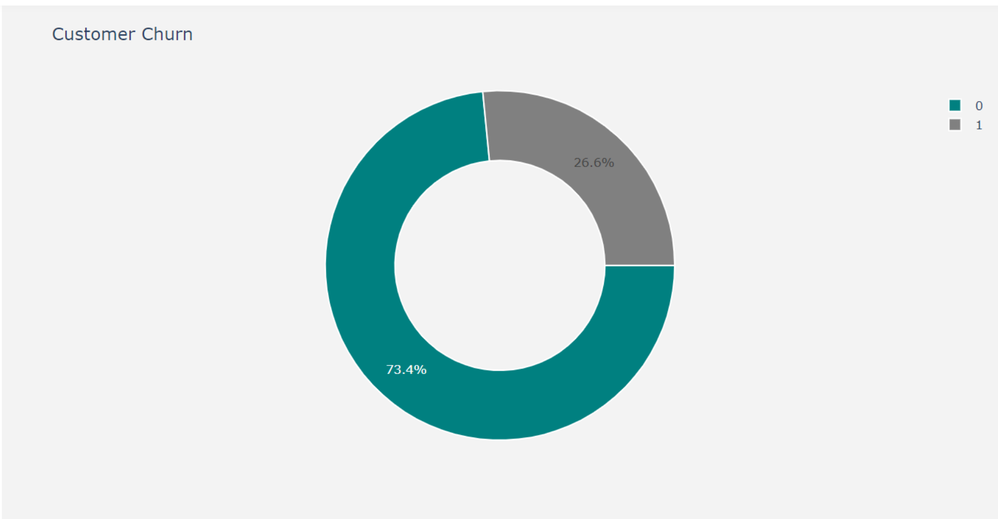


Fig:Customer Churn

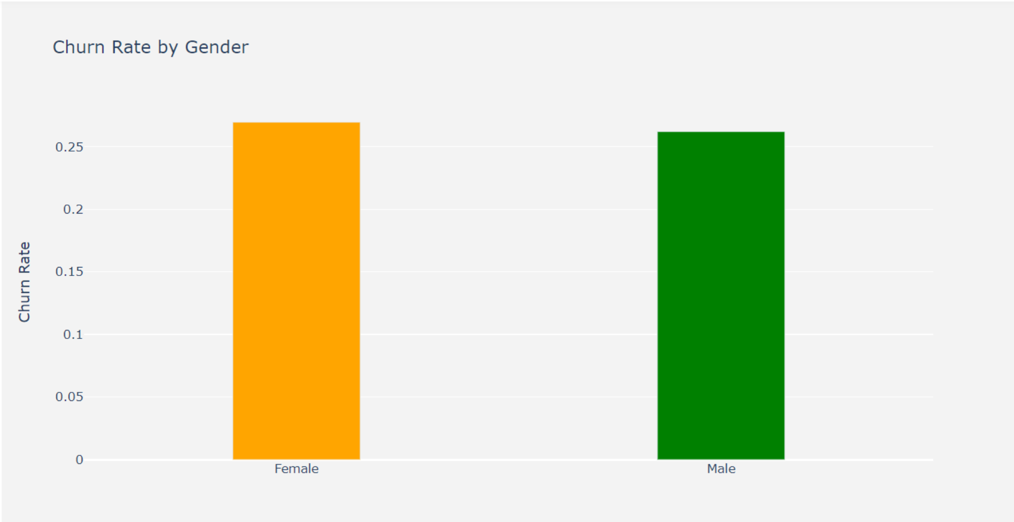


Fig:Churn Rate by Gender

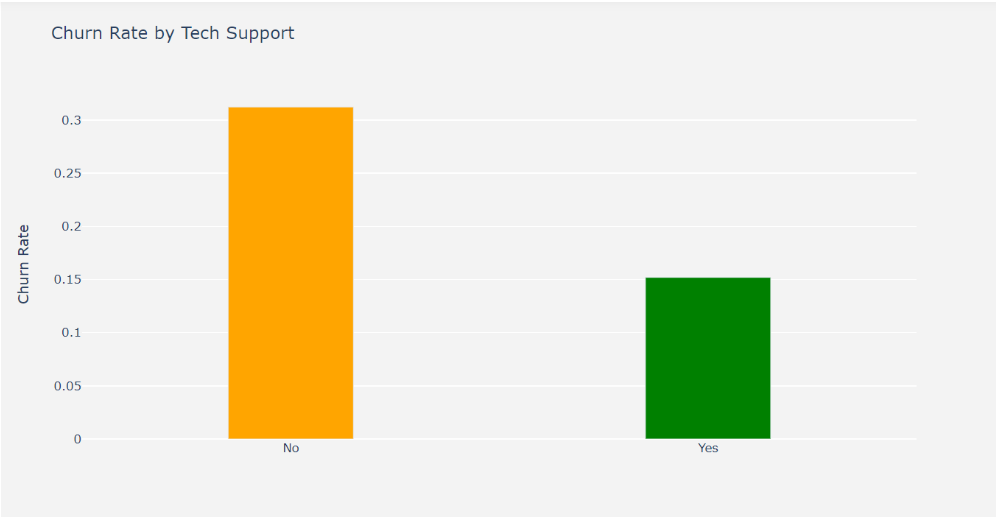


Fig:Churn Rate by Tech Support

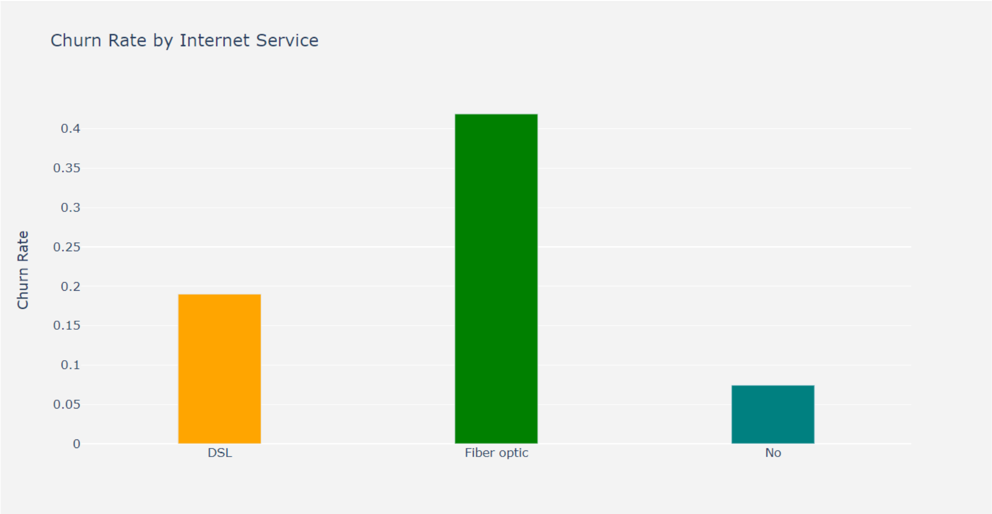


Fig Churn by Contract Duration

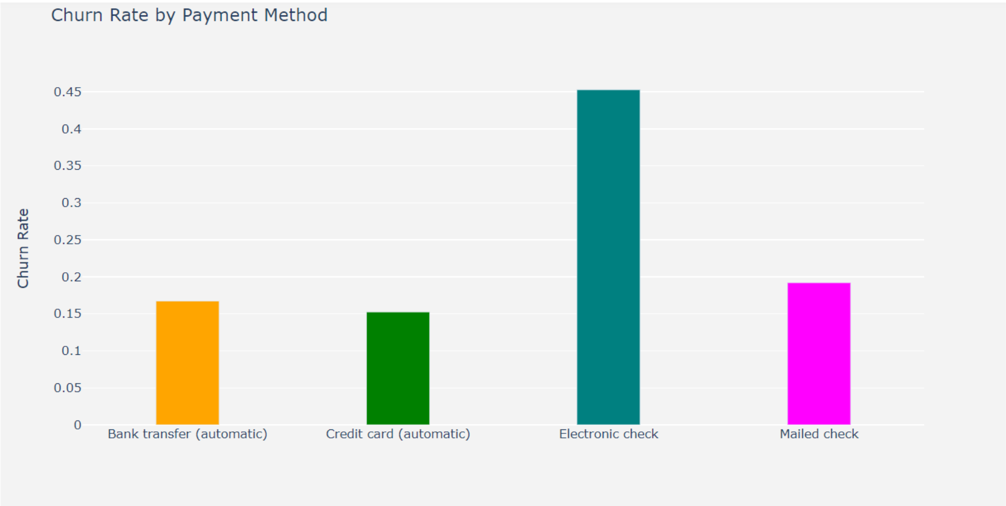


Fig:Churn by Internet Service

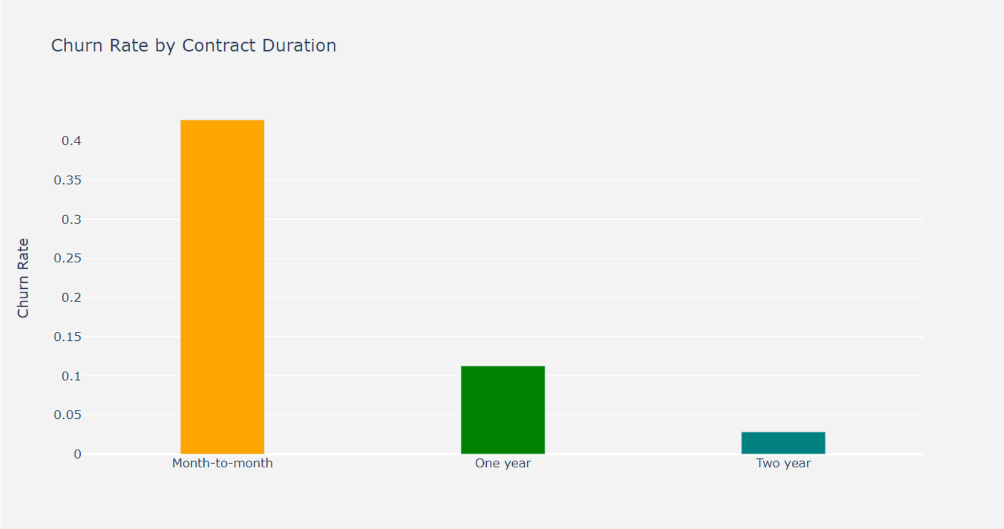


Fig:Churn by Payment Method

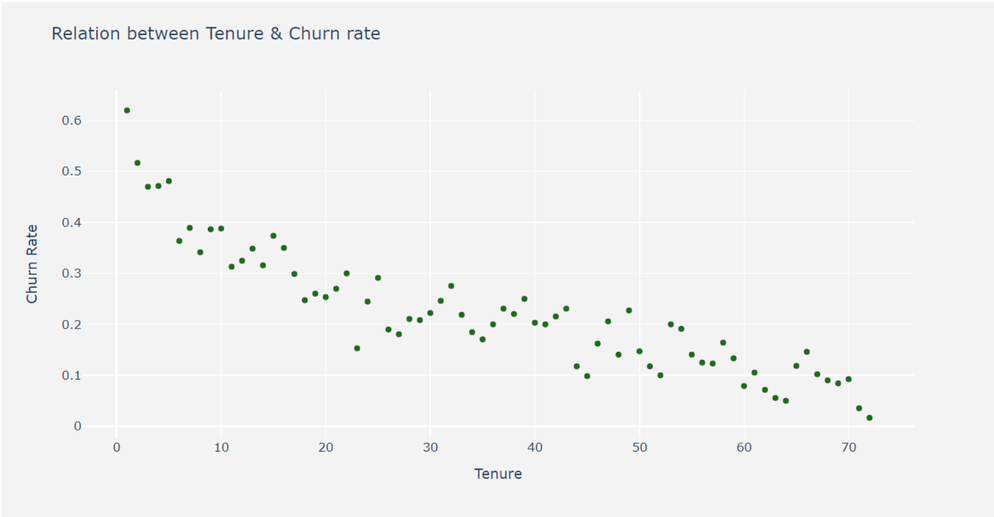


Fig: Relation between Tenure & churn rate

```
In [60]: # Compare Several models according to their Accuracies
Model_Comparison = pd.DataFrame({
    'Model': ['Perceptron','Danse neural network','Logistic Regression', 'Support Vector Machine', 'K-Nearest Neighbor',
              'Decision Tree', 'Random Forest'],
    'Score': [per_accuracy,DNN_accuracy,logmodel_accuracy, svc_accuracy, knn_accuracy,
              dt_accuracy, rf_accuracy]})
Model_Comparison_df = Model_Comparison.sort_values(by='Score', ascending=False)
Model_Comparison_df = Model_Comparison_df.set_index('Score')
Model_Comparison_df.reset_index()
```

Out[60]:

	Score	Model
0	81.14	Logistic Regression
1	80.66	Support Vector Machine
2	79.38	Random Forest
3	78.10	Perceptron
4	78.10	Danse neural network
5	76.87	K-Nearest Neighbor
6	73.27	Decision Tree

Fig: Score of Models

```
In [63]: # Create a Dataframe showcasing probability of Churn of each customer
df[['customerID','Probability_of_Churn']].head(10)
```

Out[63]:

	customerID	Probability_of_Churn
0	7590-VHVEG	0.649626
1	5575-GNVDE	0.044389
2	3668-QPYBK	0.338442
3	7795-CFOCW	0.027125
4	9237-HQITU	0.696231
5	9305-CDSKC	0.781197
6	1452-KIOVK	0.488613
7	6713-OKOMC	0.291929
8	7892-POOKP	0.593652
9	6388-TABGU	0.012203

Fig :customer chart of predicted customer

In [59]:
```
# Compare Several models according to their Accuracies
Model_Comparison = pd.DataFrame({
    'Model': ['Linear Regression','Logistic Regression', 'Support Vector Machine', 'K-Nearest Neighbor',
              'Decision Tree', 'Random Forest'],
    'Score': [LinReg_accuracy,logmodel_accuracy, svc_accuracy, knn_accuracy,
              dt_accuracy, rf_accuracy]})
Model_Comparison_df = Model_Comparison.sort_values(by='Score', ascending=False)
Model_Comparison_df = Model_Comparison_df.set_index('Score')
Model_Comparison_df.reset_index()
```

Out[59]:

	Score	Model
0	81.14	Logistic Regression
1	80.66	Support Vector Machine
2	79.38	Random Forest
3	76.87	K-Nearest Neighbor
4	73.27	Decision Tree
5	28.78	Linear Regression

Fig: Score of Models

4.Conclusion:

This project presented a new set of features for the customer churn prediction in the telecommunication, including the aggregated call details, Henley segmentation, account information, bill information, dial types, line-information, payment information, complain information, service information, and so on. Then eight modelling techniques (LR, LC, DT, RF,PERCEPTRON,ANN, SVM and KNN) were used as predictors in this paper. Finally, based on the new feature set, the existing feature sets and the eight modelling techniques, the comparative experiments were carried out. In the experiments, each subset of the new feature were evaluated and analysed. Here we used logistic regression, linear regression, random forest, decision tree, perceptron, SVM classifier, KNN,ANN all this models. Most all of the models are give much accuracy. Here logistic regression give the most accuracy all over the data and others models are worked efficiently .But the linear regression model are not worked properly for reason of big data and data details. However, there are some limitations with our proposed techniques. In the future, other information should be added into the new feature set in such a way to improve features. The dimensions of input features also should be reduced by using feature selection and extraction techniques which will be studied in the future. In addition, the imbalance classification problem takes place in this application and we only used the sampling technique to attempt to solve the problem. Therefore, more methods for imbalance classifications also should be focused in the future.

5.Bibliography:

1. Au, W., Chan, C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation, 7, 532–545.

2. Boser, B., Guyon, I., & Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In Proceedings the 5th annual ACM workshop on computational learning theory (pp. 144–152). Pittsburgh, PA: ACM Press.

3. Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition, 30, 1145–1159.

4. Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), 121–167

5. Coussement, K., & den Poe, D. V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameterselection techniques. Expert Systems with Applications, 34, 313–327.

6. Domingos, P., & Pazzani, M. J. (1997). On the optimality of the simple bayesian classifier under zero-one loss. Machine Learning, 29(2–3), 103–130.

7. Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2006). Churn prediction: Does technology matter? International Journal of Intelligent Technology, 1(2).

8. Henley, 2009. <http://www.henleymc.ac.uk/>.

9. Hosmer, D., & Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

Huang, B., Kechadi, M.-T., & Buckley, B. (2010). A new feature set with new window techniques for customer churn prediction in land-line telecommunication.

10. Expert Systems with Applications, 37(5), 3657–3665.

11. Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. Expert Systems with Applications, 31, 515–524.

12. Japkowicz, N. (2000). Learning from imbalanced data sets: A comparison of various strategies (pp. 10–15). AAAI Press

13. Japkowicz, N. (2006). Why question machine learning evaluation methods? In AAAI Workshop. Boston.

14. John, H., Ashutosh, T., Rajkumar, R., Dymitr, R. (2007). Computer assisted customer churn management: State-of-the-art and future trends.

15. Jolliffe, I. T. (1986). Principal component analysis. New York: Springer.

16. Langley, P., Iba, W., & Thompson, K. (1992). An analysis of Bayesian classifiers. In Proceedings of the 10th national conference on ARTI CIAL intelligence (pp. 223–228). MIT Press.

17. Luo, B., Shao, P., Liu, J. 2007. Customer churn prediction based on the decision tree in personal handyphone system service. In International conference on service systems and service management (pp. 1–5).

18. Quinlan, J. R. (1993). C4.5: Programs for machine learning. Morgan Kaufman Publishers.

19. Quinlan, J. R. (1996). Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research, 4, 77–90.

20. Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation (Vol. 1). MA: MIT Press.

21. Vapnik, V.N. (1998). The nature of statistical learning theory (2nd ed., pp. 23–57).

Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: A data mining approach. Expert Systems with Applications, 23, 103–112.

22. Yan, L., Wolniewicz, R., & Dodier, R. (2004). Customer behavior prediction – it's all in

the timing. Potentials IEEE, 23(4), 20–25.

23. Zhang, Y., Qi, J., Shu, H., Li, Y. (2006). Case study on crm: Detecting likely churners

with limited information of fixed-line subscriber. In 2006 International

conference on service systems and service management (Vol. 2, pp. 1495–1500).

24. Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39.1 (2012): 1414-1425.

25. Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), 1414-1425.

26. Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39, no. 1 (2012): 1414-1425.

27. Huang, B., Kechadi, M.T. and Buckley, B., 2012. Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), pp.1414-1425.

28. Huang B, Kechadi MT, Buckley B. Customer churn prediction in telecommunications. Expert Systems with Applications. 2012 Jan 1;39(1):1414-25.

29.https://www.sciencedirect.com/topics/engineering/deep-neural-network?fbclid=IwAR1Zis8dPm-lM10K4mEQLsTmTH5usd0l0EI7yEDWSuTqhey7_QziHV1ESso

30.https://medium.com/@sprhlabs/understanding-deep-learning-dnn-rnn-lstm-cnn-and-r-cnn-6602ed94dbff

31. https://towardsdatascience.com/perceptron-learning-algorithm-d5db0deab975

32. Lu, Ning, et al. "A customer churn prediction model in telecom industry using boosting." IEEE Transactions on Industrial Informatics 10.2 (2012): 1659-1665.

33. Lu, N., Lin, H., Lu, J., & Zhang, G. (2012). A customer churn prediction model in telecom industry using boosting. IEEE Transactions on Industrial Informatics, 10(2), 1659-1665.

34. Lu, Ning, Hua Lin, Jie Lu, and Guangquan Zhang. "A customer churn prediction model in telecom industry using boosting." IEEE Transactions on Industrial Informatics 10, no. 2 (2012): 1659-1665.

35. Lu, N., Lin, H., Lu, J. and Zhang, G., 2012. A customer churn prediction model in telecom industry using boosting. IEEE Transactions on Industrial Informatics, 10(2), pp.1659-1665.

36. Lu N, Lin H, Lu J, Zhang G. A customer churn prediction model in telecom industry using boosting. IEEE Transactions on Industrial Informatics. 2012 Oct 12;10(2):1659-65.