Churn Prediction in Telecom Industry Using R

Manpreet Kaur, Dr. Prerna Mahajan

Abstract— **Telecommunication market is expanding day by day. Companies are facing a severe loss of revenue due to increasing competition hence the loss of customers. They are trying to find the reasons of losing customers by measuring customer loyalty to regain the lost customers. The customers leaving the current company and moving to another telecom company are called churn.**

The research paper is using data mining technique and R package to predict the results of churn customers on the benchmark Churn dataset available from (http://www.dataminingconsultant.com/data/churn.txt). The R tool has represented the large dataset churn in form of graphs which depicts the outcomes in various unique pattern visualizations. The Churn Factor is used in many functions to depict the various areas or scenarios where churners can be distinguished. The paper is considering churn factor in account to depict various patterns for churners. R is a powerful statistical programming tool which can represent the dataset graphically with respect to different parameters and it also uses different packages available.

Churns can be reduced by analyzing the past history of the potential customers systematically. In the past few years, the fast emerging requirements from both academia and industry has helped R programming language to emerge as one of the necessary tool for visualization, computational statistics and data science

***Index Terms*—Churn, R Tool, Telecommunication, Data mining.**

I. INTRODUCTION

Numerous telecom companies are present all over the world. Telecommunication market is facing a severe loss of revenue due to increasing competition among them and loss of potential customers. Many companies are finding the reasons of losing customers by measuring customer loyalty to regain the lost customers. To keep up in the competition and to acquire as many customers, most operators invest a huge amount of revenue to expand their business in the beginning. Therefore, it has become important for the operators to earn back the amount they invested along with at least the minimum profit within a very short period of time.

1.1 Churn Prediction

Churn in the terms of telecommunication industry are the customers leaving the current company and moving to another telecom company. With the increasing number of churns, it becomes the operator's process to retain the profitable customers known as churn management. In telecommunication industry each company provides the customers with huge incentives to lure them to switch to their services, it is one of the reasons that customer churn is a big problem in the industry nowadays. To prevent this, the company should know the reasons for which the customer decides to move on to another telecom company. It is very difficult to keep customers intact for long duration as they move to the service that suits most of their needs.

Ms.Manpreet Kaur, Student, Guru Gobind Singh Indraprastha University, Institute Of Information Technology & Management, MCA, Trainee at NIIT TECHNOLOGIES, DELHI , INDIA.

Dr. Prerna Mahajan, Professor, Department of Computer Science,Guru Gobind Singh Indraprastha University/ Institute of Information Technology & Management, Delhi/India

1.2 Types

Telecom Churns can be classified in two main categories: Involuntary and Voluntary. Of the two, Involuntary are easier to identify. Involuntary churn are those customers whom the Telecom industry decides to remove as a subscriber. They are churned for fraud, non-payment and those who don't use the service. On the other hand, Voluntary churn are difficult to determine, here it is the decision of the customer to unsubscribe from the service provider. Voluntary churn can further be classified as incidental and deliberate churn. The former occurs without any prior planning by the churn but due to change in the financial condition, location, etc. Whereas, the latter happens for technological advancement, economics, quality factors and convenience reasons. Most operators are trying to deal with these type of churns mainly.

1.3 Managing Churns

Churn management is very important for reducing churns as acquiring a new customer is more expensive than retaining the existing ones. Churn rate is the measurement for the number of customers moving out and in during a specific period of time. If the reason for churning is known, the providers can then improve their services to fulfill the needs of the customers.

Churns can be reduced by analyzing the past history of the potential customers systematically. Large amount of information is maintained by telecom companies for each of their customers that keeps on changing rapidly due to competitive environment. This information includes the details about billing, calls and network data. The huge availability of information arises the scope of using Data mining techniques in the telecom database. The information available can be analyzed in different perspectives to provide various ways to the operators to predict and reduce churning. Only the relevant details are used in analysis which contribute to the study from the information given.

Data mining techniques are used for discovering the interesting patterns within data. One of the most common data mining technique is Classification, its aim is to classify unknown cases based on the set of known examples into one of the possible classes. Here, in case of telecom churn, Classification helps learn to predict whether a customer will churn or not based on customer's data stored in database.

II. BACKGROUND

2.1. Data Mining Techniques

The process of reducing, analyzing the patterns, predicting the hidden and useful required information from large Database is known as Data Mining. Association rule mining, clustering, classification and regression forms the four techniques used by data mining.

In Data mining new rules and patterns can be discovered by the system known as discovery oriented and system can also check the user's hypothesis called verification oriented. It helps in taking knowledge-driven decisions and for predicting the future trends of the business.

2.2. J48 Decision Tree Technique

J48 construction is like a flow- chart. A test applied on an attribute is denoted by internal node, its effect is denoted by a branch and class labels are presented by leaf- nodes. Process divided in two levels, one is Division of root is recursively based on selection of attribute for all training examples at the tree construction and second is that the noise or outliers branches are identified and removed by Tree pruning. Rules can be classified from the tree. If-then statement is used to represent the knowledge. For each path from root to a leaf one rule is created.

Here we use J48 for churn dataset. The attribute whose value has to be predicted is known as dependent variable. Its value is decided by value of other attributes. These attributes that predict the value of the dependent variable are known as independent variables.

2.3. Tool Used: A Revolution Analytics Tool - R

In the past few years, the fast emerging requirements from both academia and industry has helped R programming language to emerge as one of the necessary tool for visualization, computational statistics and data science. R is most popular in field of data science and important in Finance and analytics- driven companies.
R virtually consists all the possible statistical models, data manipulation and charts that could ever be required by a modern day scientist. One can easily use the best reviewed methods from leading researchers in field of Data Science without any cost. It provides a large collection of graphical and statistical techniques, consisting of modelling (linear and non-linear), statistical tests, time-series, classification, clustering, etc.
R helps in representing complex data as beautiful and unique data visualizations. Evaluation of result in R is very much easier as we do not have to remember any clicks or steps, it is simply a programming language designed specifically for data analysis that also has the capability to use mix and match models for best results.
As R is supported by a large community worldwide, solution to the errors and code is available freely. Its source code is written in C, Fortran and R. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages. R is an open source and can be extended easily as individuals using it can contribute in its growth. Dynamic and static graphics are available through additional packages. R can easily deal with complex and large datasets.
The libraries and packages of R that are being used in this paper are: RWeka, ggplot2, rpart, rJava, class

2.4. Related Literature

"Churn customer is one who leaves the existing company and become a customer of another competitor company. The management that was assumed to determine the customer turnover is called as Churn management." (Hadden, Tiwari, Roy and Ruta, 2007). "Customer movement from one provider to another in telecommunication industry is called customer churn and the operator's process to retain profitable customers counted as churn management" (Berson, Smith & Thearling, 2000) [13].

2.5 Data Set Used

The attributes in our data are taken from Orange Database.

Table I: Orange Dataset Attributes

S.No.	Attribute name
1	State
2	Account. Length
3	Area. Code
4	Phone
5	Int .l .Plan
6	VMail.Plan
7	VMail.Message
8	Day.Mins
9	Day.Calls
10	Day.Charge
11	Eve.Mins
12	Eve.Calls
13	Eve.Charge
14	Night.Mins
15	Night.Calls
16	Night.Charge
17	Intl.Mins
18	Intl.Calls
19	Intl.Charge
20	CustServ.Calls
21	Churn.

III. Algorithm and Libraries Used

3.1. J48 Algorithm

J48 (formula, data, subset, control= Weka_control ())
`Predict` is a generic function for predictions from the results of model fitting functions.

3.1.1. Steps:

Step 1. A flow-chart-like tree structure. Internal node denotes a test on an attribute. Outcome of the test is represented by Branch. Class labels are represented by Leaf nodes.
Step 2. Decision tree generation comprised of two phases.
Tree construction: At start, root contains all the training examples. Tree pruning: Branches that reflect noise and outliers are identified and removed.
Step 3. Decision tree is used to classify an Unknown sample. Attribute values of the sample are tested against the decision tree.
Step 4. When all samples for a given node belong to the same class, or there are no remaining attributes for further partitioning then the partitioning is stopped.

International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-3, Issue-5, May 2015

3.1.2. Extracting Classification Rules from Trees

1. IF-THEN rules are used.
2. From root to leaf one rule is created for each path.
3. Each attribute-value pair along a path forms a conjunction.
4. The leaf node holds the class prediction.
5. Rules are easier for humans to understand.

3.2. Using rpart package

rpart (formula, data, method)
f<-rpart(Churn.~CustServ.Calls+Eve.Charge+Intl.Charge+Night.Charge+Day.Charge, method="class", data=churn)
Package rpart is used in plotting the graphs. The functions within rpart that are used are as follows:

3.2.1.1. Using plotcp function

Applied on the set of possible cost- complexity pruning of a tree from a nested set. A cross- validation is already performed by rpart on the geometric means of the Interval values of cp where pruning is optimal. The mean and standard deviation of errors in cross-validated prediction against each of the geometric means is stored in cptable in 'f' are plotted by this function. A good choice of cp for pruning is often the leftmost value for which the mean lies below the horizontal line.

3.3. Using Plot function

plot(Churn. ~., data = churn, type = "c")
lines(Churn.~ Day.Charge,type="l")
In plot function, x and y axis are mentioned along with the data source and the type of graph that is, curve, line etc.

3.4. Using ggplot2 package

`qplot` is the basic plotting function in the ggplot2 package. It is familiar with plot function. It is quick plot as it produces complex plots in mere one line, which often require several lines of code using other plotting systems. It helps depicting more than 2 variables in a single graph with help of colors and geometrical shapes and a lot more.

IV. Experimental Results And Observations

4.1 Reading Data Set Churn from CSV file

churn<-read.csv("C:\\Users\\Documents\\R\\win-library\\3.1\\RWeka\\R\\churn.csv", header=T)

4.2. Names of all the attributes

```
> names(churn)
 [1] "State"          "Account.Length" "Area.Code"      "Phone"
 [5] "Int.l.Plan"     "VMail.Plan"     "VMail.Message"  "Day.Mins"
 [9] "Day.Calls"      "Day.Charge"     "Eve.Mins"       "Eve.Calls"
[13] "Eve.Charge"     "Night.Mins"     "Night.Calls"    "Night.Charge"
[17] "Intl.Mins"      "Intl.Calls"     "Intl.Charge"    "CustServ.Calls"
[21] "Churn."
```

4.3. Description of complete data Set

Table II: Description of Data Set Attribute

```
> str(churn)
'data.frame':   3333 obs. of  21 variables:
 $ State         : Factor w/ 51 levels "AK","AL","AR",..: 17 36 32 36 37 2 20 2$
 $ Account.Length: int  128 107 137 84 75 118 121 147 117 141 ...
 $ Area.Code     : int  415 415 415 408 415 510 510 415 408 415 ...
 $ Phone         : Factor w/ 3333 levels "327-1058","327-1319",..: 1927 1576 11$
 $ Int.l.Plan    : Factor w/ 2 levels "no","yes": 1 1 1 2 2 2 1 2 1 2 ...
 $ VMail.Plan    : Factor w/ 2 levels "no","yes": 2 2 1 1 1 1 2 1 1 2 ...
 $ VMail.Message : int  25 26 0 0 0 0 24 0 0 37 ...
 $ Day.Mins      : num  265 162 243 299 167 ...
 $ Day.Calls     : int  110 123 114 71 113 98 88 79 97 84 ...
 $ Day.Charge    : num  45.1 27.5 41.4 50.9 28.3 ...
 $ Eve.Mins      : num  197.4 195.5 121.2 61.9 148.3 ...
 $ Eve.Calls     : int  99 103 110 88 122 101 108 94 80 111 ...
 $ Eve.Charge    : num  16.78 16.62 10.3 5.26 12.61 ...
 $ Night.Mins    : num  245 254 163 197 187 ...
 $ Night.Calls   : int  91 103 104 89 121 118 118 96 90 97 ...
 $ Night.Charge  : num  11.01 11.45 7.32 8.86 8.41 ...
 $ Intl.Mins     : num  10 13.7 12.2 6.6 10.1 6.3 7.5 7.1 8.7 11.2 ...
 $ Intl.Calls    : int  3 3 5 7 3 6 7 6 4 5 ...
 $ Intl.Charge   : num  2.7 3.7 3.29 1.78 2.73 1.7 2.03 1.92 2.35 3.02 ...
 $ CustServ.Calls: int  1 1 0 2 3 0 3 0 1 0 ...
 $ Churn.        : Factor w/ 2 levels "False.","True.": 1 1 1 1 1 1 1 1 1 1 ...
```

4.4. Summary of Data set

Table III: Summary of Dataset

```
> summary(churn)
     State      Account.Length    Area.Code           Phone        Int.l.Plan
 WV     : 106   Min.   :  1.0   Min.   :408.0   327-1058:   1   no :3010
 MN     :  84   1st Qu.: 74.0   1st Qu.:408.0   327-1319:   1   yes: 323
 NY     :  83   Median :101.0   Median :415.0   327-3053:   1
 AL     :  80   Mean   :101.1   Mean   :437.2   327-3587:   1
 OH     :  78   3rd Qu.:127.0   3rd Qu.:510.0   327-3850:   1
 OR     :  78   Max.   :243.0   Max.   :510.0   327-3954:   1
 (Other):2824                                   (Other) :3327
 VMail.Plan VMail.Message      Day.Mins        Day.Calls       Day.Charge
 no :2411   Min.   : 0.000   Min.   :  0.0   Min.   :  0.0   Min.   : 0.00
 yes: 922   1st Qu.: 0.000   1st Qu.:143.7   1st Qu.: 87.0   1st Qu.:24.43
            Median : 0.000   Median :179.4   Median :101.0   Median :30.50
            Mean   : 8.099   Mean   :179.8   Mean   :100.4   Mean   :30.56
            3rd Qu.:20.000   3rd Qu.:216.4   3rd Qu.:114.0   3rd Qu.:36.79
            Max.   :51.000   Max.   :350.8   Max.   :165.0   Max.   :59.64

    Eve.Mins       Eve.Calls       Eve.Charge      Night.Mins
 Min.   :  0.0   Min.   :  0.0   Min.   : 0.00   Min.   : 23.2
 1st Qu.:166.6   1st Qu.: 87.0   1st Qu.:14.16   1st Qu.:167.0
 Median :201.4   Median :100.0   Median :17.12   Median :201.2
 Mean   :201.0   Mean   :100.1   Mean   :17.08   Mean   :200.9
 3rd Qu.:235.3   3rd Qu.:114.0   3rd Qu.:20.00   3rd Qu.:235.3
 Max.   :363.7   Max.   :170.0   Max.   :30.91   Max.   :395.0

  Night.Calls     Night.Charge      Intl.Mins        Intl.Calls
 Min.   : 33.0   Min.   : 1.040   Min.   : 0.00   Min.   : 0.000
 1st Qu.: 87.0   1st Qu.: 7.520   1st Qu.: 8.50   1st Qu.: 3.000
 Median :100.0   Median : 9.050   Median :10.30   Median : 4.000
 Mean   :100.1   Mean   : 9.039   Mean   :10.24   Mean   : 4.479
 3rd Qu.:113.0   3rd Qu.:10.590   3rd Qu.:12.10   3rd Qu.: 6.000
 Max.   :175.0   Max.   :17.770   Max.   :20.00   Max.   :20.000
```

```
  Intl.Charge    CustServ.Calls     Churn.
 Min.   :0.000   Min.   :0.000   False.:2850
 1st Qu.:2.300   1st Qu.:1.000   True. : 483
 Median :2.780   Median :1.000
 Mean   :2.765   Mean   :1.563
 3rd Qu.:3.270   3rd Qu.:2.000
 Max.   :5.400   Max.   :9.000
```

4.5. Decision Tree for Churn (using J48)

m2 <- J48(`Churn.` ~ ., data = churn)
m2

```
> m2 <- J48(`Churn.` ~ ., data = churn)
> m2
J48 pruned tree
------------------

Day.Mins <= 264.4
|   CustServ.Calls <= 3
|   |   Int.l.Plan = no
|   |   |   Day.Mins <= 223.2: False. (2221.0/60.0)
|   |   |   Day.Mins > 223.2
|   |   |   |   Eve.Mins <= 242.3: False. (296.0/22.0)
|   |   |   |   Eve.Mins > 242.3
|   |   |   |   |   VMail.Plan = no
|   |   |   |   |   |   Night.Mins <= 174.2
|   |   |   |   |   |   |   Day.Mins <= 246.8: False. (12.0)
|   |   |   |   |   |   |   Day.Mins > 246.8: True. (5.0/1.0)
|   |   |   |   |   |   Night.Mins > 174.2: True. (50.0/8.0)
|   |   |   |   |   VMail.Plan = yes: False. (20.0)
|   |   Int.l.Plan = yes
|   |   |   Intl.Calls <= 2: True. (51.0)
|   |   |   Intl.Calls > 2
|   |   |   |   Intl.Mins <= 13.1: False. (173.0/7.0)
|   |   |   |   Intl.Mins > 13.1: True. (43.0)
|   CustServ.Calls > 3
|   |   Day.Mins <= 160.2
|   |   |   Eve.Charge <= 19.83: True. (79.0/3.0)
|   |   |   Eve.Charge > 19.83
|   |   |   |   Day.Mins <= 120.5: True. (10.0)
|   |   |   |   Day.Mins > 120.5: False. (13.0/3.0)
|   |   Day.Mins > 160.2
|   |   |   Eve.Charge <= 12.05
|   |   |   |   Eve.Calls <= 125: True. (16.0/2.0)
|   |   |   |   Eve.Calls > 125: False. (3.0)
|   |   |   Eve.Charge > 12.05: False. (130.0/24.0)

Day.Mins > 264.4
|   VMail.Plan = no
|   |   Eve.Mins <= 187.7
|   |   |   Day.Mins <= 280.4: False. (30.0/7.0)
|   |   |   Day.Mins > 280.4: True. (27.0/9.0)
|   |   Eve.Mins > 187.7: True. (101.0/5.0)
|   VMail.Plan = yes: False. (53.0/6.0)

Number of Leaves  :     19

Size of the tree :      37

> m3<- table(churn$`Churn.`, predict(m2))
> m3

          False. True.
  False.    2822    28
  True.      129   354
> plot(m3)
```

Fig. 1 depicts the churn values from table formed by predicting the values of J48 decision tree on churn parameter.

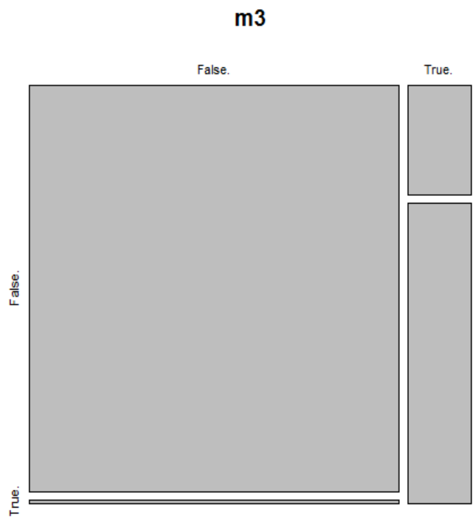

Figure 1: Churn value prediction

4.6. Classification Tree for all the Calls (using rpart)

library(rpart)
f<-rpart(Churn.~CustServ.Calls+Eve.Calls+Intl.Calls+Night.Calls +Day.Calls,method="class", data=churn)
plot(f, uniform=TRUE,main="Classification Tree for Churn")
text(f, use.n=TRUE, all=TRUE, cex=.7)

Fig. 2 represents the classification tree for all the Calls considered in churn Dataset. The decision is made on basis of call number and the churn factor having values true and false.

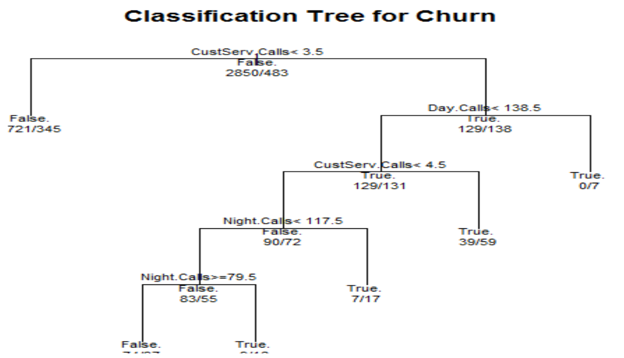

Figure 2: Classification Tree Based on Calls

4.7. Using rpart

library(rpart)
f<-rpart(Churn.~CustServ.Calls+Eve.Charge+Intl.Charge+Night. Charge+Day.Charge, method="class", data=churn)
plotcp(f,lty=4,col="red")

Fig. 3 represents Applied on the set of possible cost- complexity pruning of a tree from a nested set. A cross- validation is already performed by rpart on the geometric means of the Interval values of cp where pruning is optimal. The mean and standard deviation of errors in cross- validated prediction against each of the geometric means is stored in cptable in 'f' are plotted by this function.

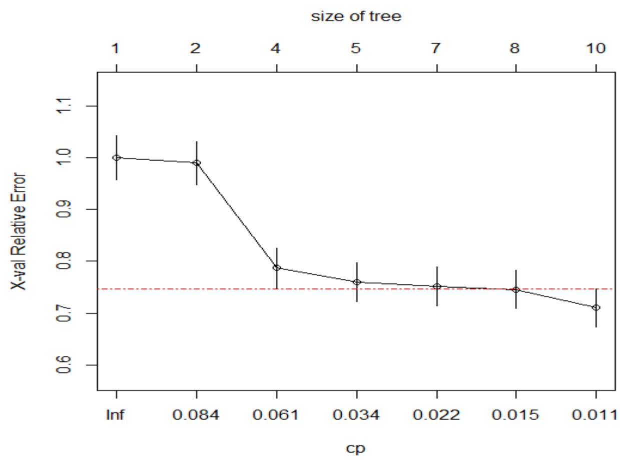

Figure 3: Plotcp function

4.8. Using Plot Function

plot(Churn. ~., data = churn, type = "c")
lines(Churn.~ Day.Charge,type="l")

Fig. 4 represents the graph of customer service calls with respect to churn factor that has just two values True and false

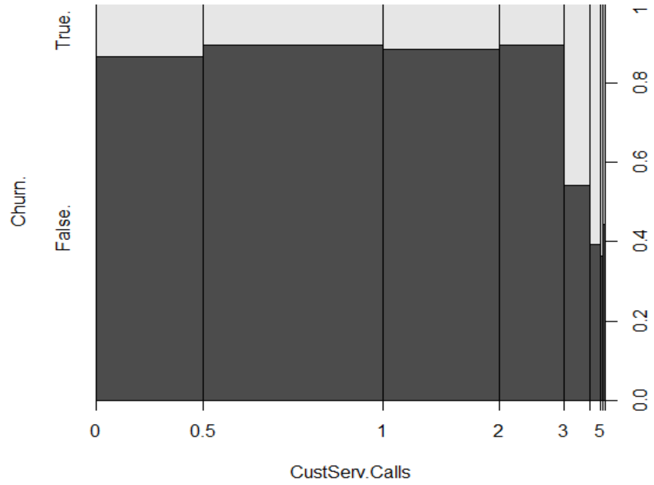

Figure 4: Churn and Customer Service Calls

Fig. 5 represents the graph of churn factor with respect to all the states. The number of churns can easily be observed state wise in the graph.

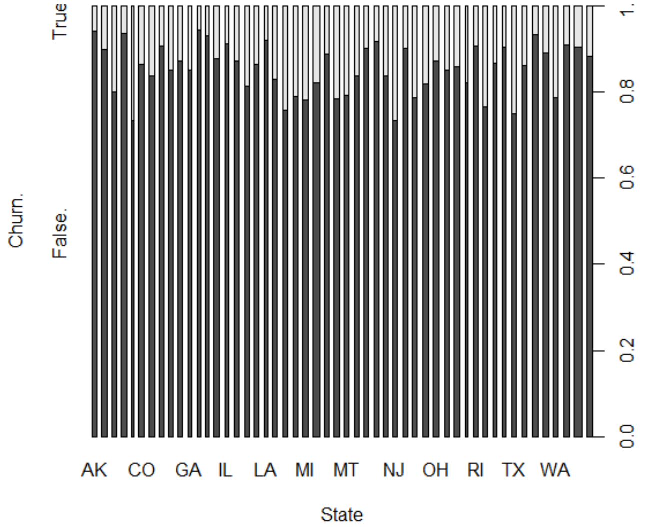

Figure 5: Churn with respect to different States

Fig. 6 represents the graph of churn factor with respect to International plan.

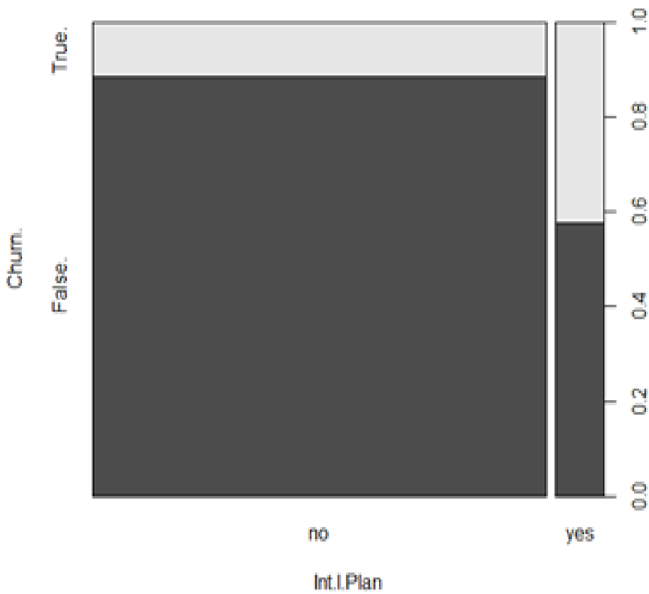

Figure 6: Churn and International Plan

4.11. Line Charts

4.11.1. Line Chart for Day Calls and Customer Service Calls

```
# convert factor to numeric for convenience
churn$Churn. <- as.numeric(churn$Churn.)
ntrees <- max(churn$Churn.)
# get the range for the x and y axis
xrange <- range(churn$Day.Calls)
yrange <- range(churn$CustServ.Calls)
# set up the plot
plot(xrange, yrange, type="n", xlab="Day.Calls (num)",
ylab="CustServ.Calls(num)" )
colors <- rainbow(ntrees)
linetype <- c(1:ntrees)
plotchar <- seq(15,15+ntrees,1)
# add lines
for (i in 1:ntrees) {
 tree <- subset(churn, Churn.==i)
 lines(tree$Day.Calls, tree$CustServ.Calls, type="b", lwd=1.5,
lty=linetype[i], col=colors[i], pch=plotchar[i]) }
# add a title and subtitle
title("Churn", "line plot")
# add a legend
legend(xrange[1], yrange[2], 1:ntrees, cex=0.8,
col=colors,pch=plotchar, lty=linetype, title="Tree")
```

Fig. 7 shows the line chart of Day calls and Customer Service calls using numbers as range and considering the Churn factor. The number of churns increase with the increase in customer service calls.

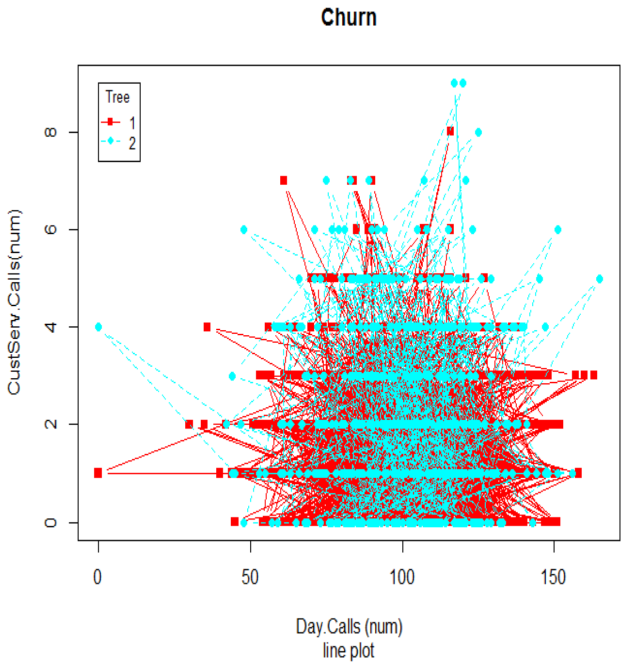

Figure 7: Line Chart Day Calls and Customer Service Calls

4.12. Using qplot function

4.12.1. qplot(Day.Calls, CustServ.Calls, data = churn,colour=Churn.)

Fig. 8 shows us relativity between customer service calls and Day calls with respect to the churn factor that is represented by two colors. By the color in the graph we see that churners are more in high number of customer service calls. Blue color represents the customers who churned.

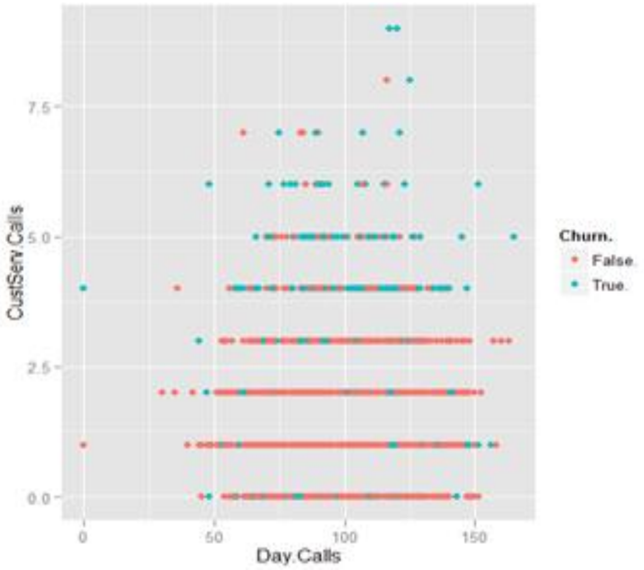

Figure 8: Qplot with three parameters

4.12.2. qplot(Day.Calls,Night.Calls, data = churn,geom = c("point", "smooth"),color=Churn.)

Fig. 9 shows the relativity in number of night calls and day calls. We observe that they are relatively dense in the same area. Whereas, the third parameter Churn factor represented by the color shows that True churns are less in number. The line in the graph is the smoother that depicts the trend followed by the data in graph. Here it depicts the relatively same number of night calls and Day calls.

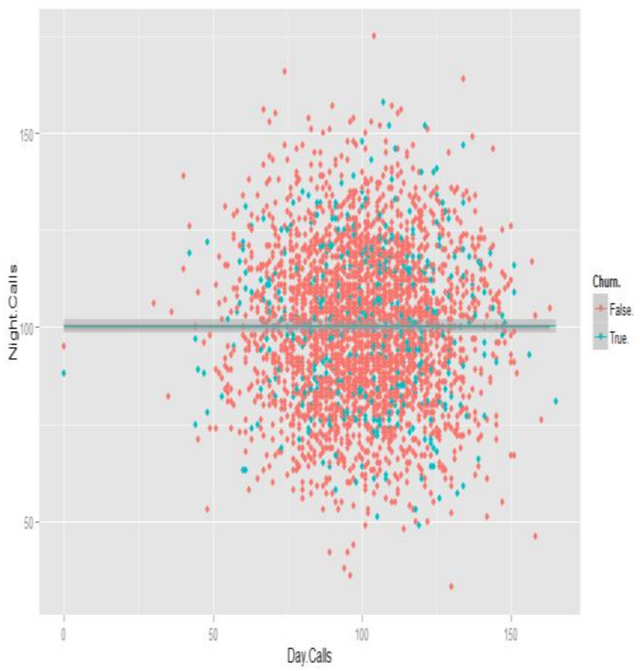

Figure 9: Relativity in Night Calls and Day Calls with churn factor using smooth curve

4.12.3. dsc<- churn[sample(nrow(churn),100),]

qplot(Day.Calls,CustServ.Calls, data = dsc, geom = c("point", "smooth"),color=Churn.)

Fig. 10 shows the relativity in number of customer service calls and day calls on the subset of Data. The third parameter Churn factor represented by the color. The smooth lines in the graph show clearly that Churns are more in case of high customer service calls. Whereas, they don't vary much with the day calls.

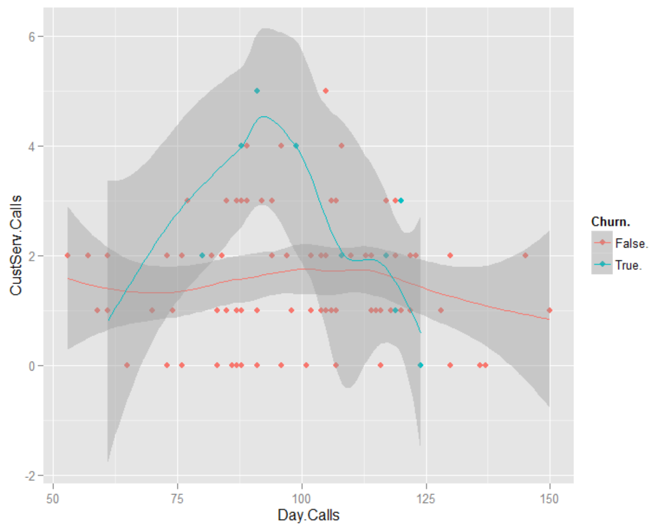

Figure 10: Relativity In Customer Service Calls, Day Calls and Churn factor

4.12.4. qplot(Day.Calls,CustServ.Calls, data=churn,facets=Churn.~Area.Code)

Fig. 11 shows the relativity in number of customer service calls and day calls on the subset of Data. The third parameter Churn factor and fourth is Area Code. The facets are representing third and fourth parameter. We can observe the churns in particular area code with respect to number of day calls and customer service calls.

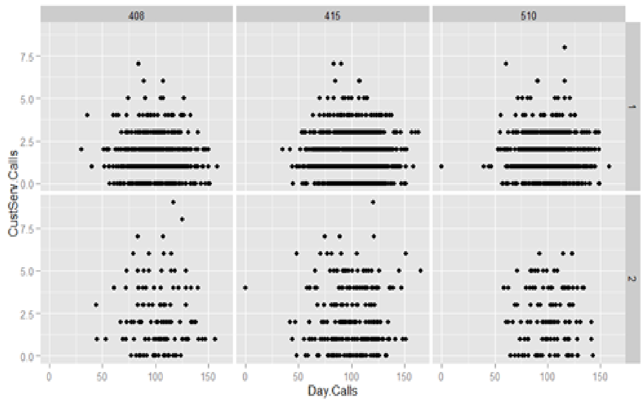

Figure 11: Relation of Churn in Area code w.r.t. Calls

4.12.5. dsc<- churn[sample(nrow(churn), 100),]

qplot(Day.Calls,Churn., data = dsc,geom = c("point", "smooth"),color=State)

Fig. 12 shows the graph of day calls and churn factor on the subset of Data. The third parameter state represented by the color shows the churns in various states.

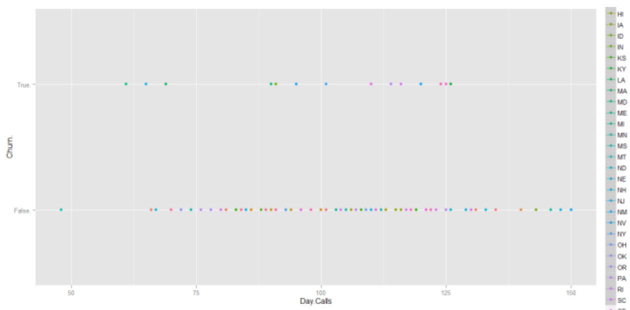

Figure 12: State wise Churn factor

4.12.6. qplot(Area.Code,Night.Mins, data=dsc) fig. 13 shows the relativity in Night minutes and Area code. We see that there is no calls made in Area code between 425 and 500.

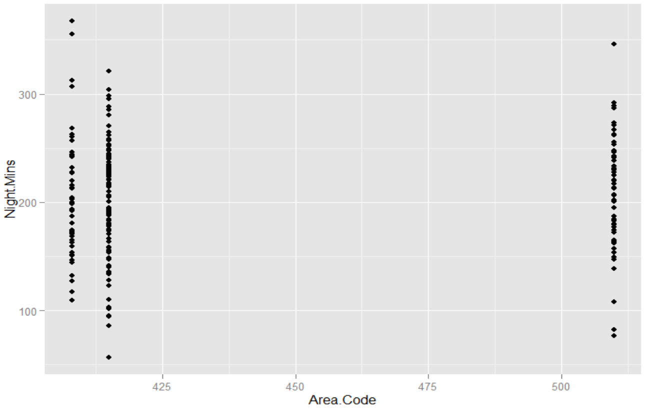

Figure 13: Relativity in Night min and Area code

4.12.7. qplot(Day.Calls,Night.Calls, data = churn, alpha=I(1/2))

Fig. 14 shows the use of alpha filter that shows transparency. It shows where the majority of points lie in the graph.

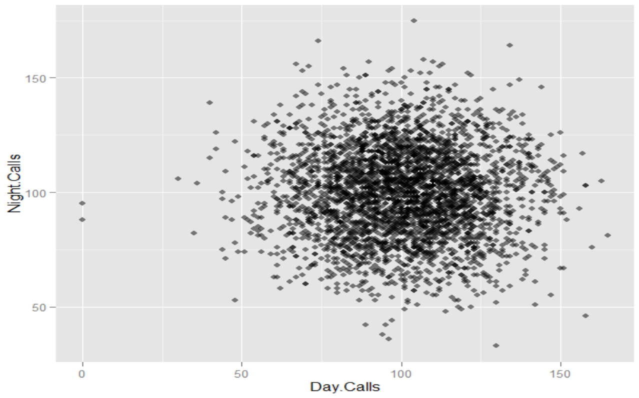

Figure 14: Alpha Filter

4.12.8. qplot(Day.Calls, data = dsc,geom = "histogram",fill=Churn.)

Fig. 15 shows the histogram, in which color is done using the Churn factor and represents Day Calls.

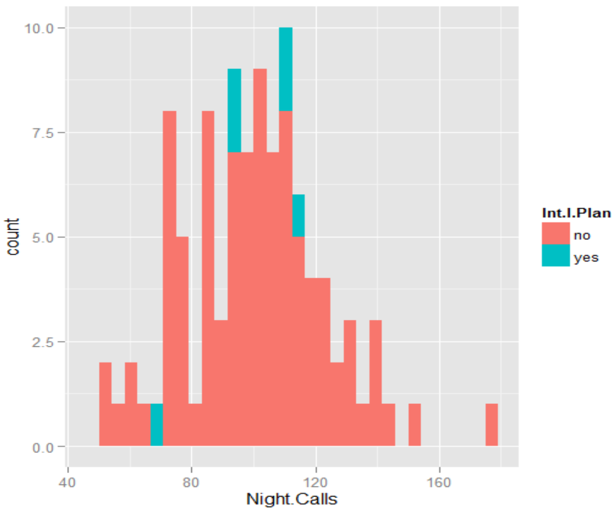

Figure 15: Histogram for Day Calls

4.12.9. qplot(Night.Calls, data = dsc,geom = "histogram",fill=Int.l.Plan)

Fig. 16 shows the histogram, color is done using the Churn factor and represents Night Calls.

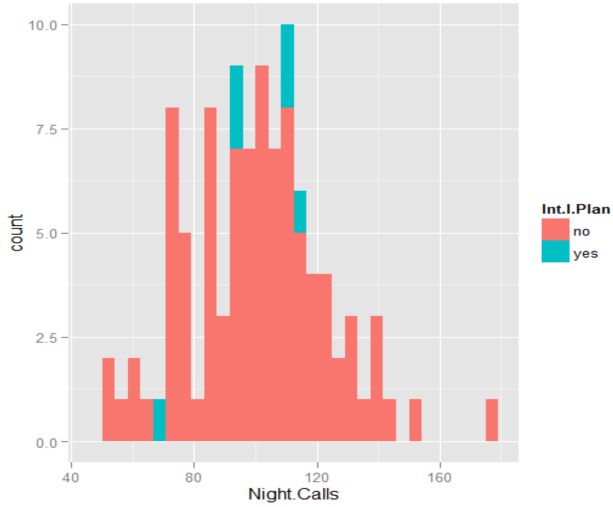

Figure 16: Histogram for Night Calls

V. CONCLUSION

The proposed research has used data mining technique and R package to predict the results of churn customers on the benchmark Churn dataset available at http://www.sgi.com/tech/mlc/db/ and http://www.dataminingconsultant.com/data/churn.txt. It has evaluated, the number of churns using the classification technique J48 tree. The R tool has represented the large dataset churn in form of graphs which depicts the outcomes vividly and in a unique pattern visualization manner. The Churn Factor is used in many functions to depict the various areas or scenarios when the churn rate is high. The study predicts that there is a huge deviation in graph of churners when customer service calls are measured. The graphs are made taking churn factors as the deciding parameters. Graphs represent the different ways of observing the number of churners from the dataset. Once the root area is recognized the steps can be taken by Telecom Company to improve their services and retain their old customers from churning

REFERENCES

[1] "Applying Data Mining Techniques in Telecom Churn Prediction" , International Journal of Advanced Research in Computer Science and Software Engineering, N.Kamalraj Dr.A.Malathi , Department of Computer Technology PG and Research Department of Computer Science, Dr. SNS Rajalakshmi College of Arts and Science Govt. Arts College.

[2] Anshul Goyal and Rajni Mehta, Assistant Professor,CSE. "Performance Comparison of Naïve Bayes and J48 Classification Algorithms."

[3] R Data: http://cran.r-project.org/

[4] Vladislav Lazarov, Technische Universität München and Marius Capota, Technische Universität München, "Churn Prediction",

[5] Prof S. Chandrasekhar, "Predicting The Churn In Telecom Industry"'.

[6] Dr. Rajni Jain, "Introduction to Data Mining Techniques".

[7] Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", Expert Systems with Applications, 37 (2010) 4710–4712.

[8] Gary M. Weiss, "Data Mining in the Telecommunications Industry", Data Mining and Knowledge Discovery Handbook, 2005 - Springer

[9] "Data Mining And Business Analytics with R", Johannes Ledolter, Dept. of Management Studies, University of IOWA.

[10] "Predicting Customer Churn in Mobile Telephony Industry Using robabilistic Classifiers in Data Mining" ,Clement Kirui1, Li Hong2, Wilson Cheruiyot3 and Hillary Kirui4.
[11] " Applying Data Mining to Insurance Customer Churn Management", Reza Allahyari Soeini and Keyvan Vahidy Rodpysh
[12] Han and Kamber. Data Mining: Concepts and Techniques. Second Morgan Kaufman Publisher, 2006
[13] "A Proposed Churn Prediction Model", Essam Shaaban, Yehia Helmy, Ayman Khedr, Mona Nasr,Department of Information Systems, MUST University, 6 October, Cairo, Egypt and Department of Information Systems, Helwan University, Cairo, Egypt.
[14] "Data Mining Applications in Customer Churn Management", Sahand KhakAbi, Industrial Engineering Department, Iran University of Science and Technology, Tehran, Iran, Mohammad R. Gholamian, Morteza Namvar.
[15] (IJACSA) International Journal of Advanced Computer Science and Applications, "Churn Prediction in Telecommunication Using Data Mining Technology", Rahul J. Jadhav of Bharati Vidyapeeth Deemed University,Pune, Yashwantrao Mohite of Institute of Management, Karad and Usharani T. Pawar2, Shivaji University Kolhapur.
[16] International Journal of Computer Applications (0975 – 8887) Volume 42– No.20, March 2012, " Applications of Data Mining Techniques in Telecom Churn Prediction", V.Umayaparvathi Department of Computer Science DDE, Madurai Kamaraj University, K. Iyakutti School of Physics Madurai Kamaraj University Madurai, Tamilnadu, India.
[17] Yu-Teng Chang, "Applying Data Mining To Telecom Churn Management", IJRIC , 2009 67 – 77.
[18] "Telecommunication Subscribers' Churn Prediction Model Using Machine Learning", Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal Department of Computing, (NUST), Islamabad, Ahsan Rehman, Business Analytic Consultant, IBM .
[19] Goran Kraljevi´c, Sven Gotovac, "Modeling Data Mining Applications for Prediction of Prepaid, Churn in Telecommunication Services".
[20] Bong-Horng Chu, Ming-Shian Tsai and Cheng-Seen Ho, "Toward a hybrid data mining model for customer retention," Knowledge-Based Systems, vol. 20, 2007, pp. 703–718.
[21] X. Hu, "A Data Mining Approach for Retailing Bank Customer Attrition Analysis," Applied Intelligence, vol. 22, 2005, pp. 47–60, Springer.
[22] "Data Mining in the Telecommunications Industry", Gary M. Weiss, Fordham University, USA.
[23] "Customer Churn Prediction for Telecom Services", Utku Yabas, Turker Ince, School of Engineering and Computer Science, Izmir University of Economics, Izmir, Turkey, Hakki Candan Cankaya, Computer Science and Engineering Department, Southern Methodist University Dallas, Texas, 2012 IEEE 36th International Conference on Computer Software and Applications.
[24] "The Study on Feature Selection in Customer Churn Prediction Modeling", Yin Wu, Jiyain Qi, School of Econ+omic and Management, Beijing University of Posts and Telecommunications, Beijing, China, Chen Wang, IBM China Research Laboratory, IBM, Beijing, P.R. China.
[25] "Predicting Customer Behavior in Telecommunications," L. Yan, R.H. Wolniewicz, R. Dodier, IEEE Intelligent Systems, IEEE Computer Society.
[26] Wikipedia "www.wikipedia.com".
[27] Elegant Graphics for Data Analysis, Wickham, H. "http://www.springer.com/978-0-387-981040-6.
[28] "Telecommunication Subscribers' Churn Prediction Model Using Machine Learning" , Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal

Ms.Manpreet Kaur, Student, Guru Gobind Singh Indraprastha University, Institute Of Information Technology & Management, Mca, Trainee At Niit Technologies, Delhi , India

Dr. Prerna Mahajan, ,Professor, Department Of Computer Science,Guru Gobind Singh Indraprastha University/ Institute Of Information Technology & Management, Delhi/India