

# A Comparative Analysis Using RStudio for Churn Prediction

Vani Kapoor Nijhawan, Mamta Madan, Meenu Dave

**Abstract:** With the availability of numerous data, in each and every sphere, it has become significant to analyze the voluminous data, and utilize the generated patterns for the future predictions. This is what we refer to as data mining. This paper exploits, decision tree technique, to predict churning trends of telecom users. For this study, authors are making use of R and its GUI Rattle. In this paper, the focus is, to compare the variations in churning patterns of a number of users, based on the reflections made by different variables or factors and then make the predictions thereafter.

**Keywords:** Data mining, Decision Tree, Customer churn, RStudio, Rattle

## I. INTRODUCTION

Churning of customers is an area of concern in almost all spheres. Telecommunications, is one such field. Here, the users of mobile phones, keep on switching their providers every now and then. There are enormous reasons for the same. Out of the many factors, in the dataset, some have a strong impact on the churning and others have least or no effect. In this paper, the authors would like to explore the same and would like to highlight, a few factors, responsible, for the churning of telecom customers and identifying the least affecting one. This analysis would be done in RStudio using Rattle, as the tool, for generating the decision trees.

### Decision Tree

Decision tree is that the flow diagram representing a tree in top-down fashion, ranging from root node and moving towards the terminal nodes. Here, the inner nodes represent a take a look at or call and also the branches represent the outcomes. It's a well-liked technique attributable to its simplicity. [1].

### Software Tools Being Used R

R is a very popular open source statistics and graphics including hundreds of additional packages available for free, which are very useful for providing help in data mining, machine learning and statistics. [w1]

### R: Rattle package

Rattle stands for, R Analytic Tool To Learn Easily (Rattle). Rattle serve as an interface to R in helping the user to load

data directly from a CSV file (or using ODBC), transform it (if required), explore the data, apply various techniques, build models and analyse the results. [w2]

## II. RELATED WORK

S. Hussain et al.[2], have used data processing within the field of education to predict the performance of undergraduates and used R for analysis. The explanation for exploitation R is that, R provides heaps several applied mathematics techniques within the style of varied packages for modeling, analysing, clustering, classification etc. Also, the programmers having data of C, C++, Java, .NET or Python will write their own code to govern the R objects. apart from this, R contains graphical packages conjointly to supply sensible quality graphs. [3]  
Wonhee Cho et al.[3] in their paper on huge knowledge have compared and analysed interactive internet packages with R visual image packages. They additional that R has been improved for a giant knowledge analysis and mining tool. It's supported with multiple packages for various targets with visual image.

J. Rao, R. Kelappan, P. Pallath [4] in their paper tried to use, data processing ideas like regression, classification and hybrid approaches to effectively set up the work allocations within the designing section of package development exploitation R- scripts. These results can be accustomed predict the additional designing of the task and facilitate the management to achieve higher visibility.  
Authors in [5] conjointly found that R is one among the foremost widespread languages within the knowledge science, applied mathematics and machine learning community. They additional that several knowledge scientists are hindered by its limitations of obtainable functions to handle giant datasets with efficiency. So, the authors mentioned here regarding the solutions like providing a public code repository that attendees are ready to access and adapt to their own observe.  
Manpreetkaur and Dr Prema [6] in their paper, aforementioned that if the info already obtainable with the medium firms if analysed rigorously, will throw some light-weight on churning patterns of the purchasers. This info are often used for the present and approaching customers to style the retain policies.

## III. METHODOLOGY

The process of mining, the gathered dataset in rattle needs to undergo some basic steps of data cleansing and pre-processing.

Revised Manuscript Received on May 28, 2019.

Vani Kapoor Nijhawan, Assistant Professor, VIPS, GGSIPU Delhi  
Mamta Madan, Professor, VIPS, GGSIPU Delhi  
Meenu, VIPS, GGSIPU Delhi