Project Dissertation Report on

Churn Rate Prediction using Machine Learning

Submitted By:

SUDEEP KAPADIA

2K17/MBA/95

Under the Guidance of:

Mr. Yashdeep Singh

Assistant Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

CERTIFICATE FROM THE INSTITUTE

This is to certify that the Project Report titled **CHURN RATE PREDICTION USING MACHINE LEARNING**, is a bonafide work carried out by **Mr. SUDEEP KAPADIA** of MBA 2018-19 and submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 in partial fulfilment of the requirement for the award of the Degree of Masters of Business Administration.

Signature of Guide (DSM)	Signature of Head (DSM)
Mr. Yashdeep Singh	Dr. Rajan Yadav

Seal of Head

Place: New Delhi

Date: 29th May 2019

DECLARATION

I **Sudeep Kapadia**, student of MBA 2018-19 of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi-42 declare that Major Research Project Report on **Churn Rate Prediction using Machine Learning** submitted in partial fulfilment of Degree of Masters of Business Administration is the original work conducted by me.

The information and data given in the report is authentic to the best of my knowledge.

This Report is not being submitted to any other University for award of any other Degree, Diploma and Fellowship.

Sudeep Kapadia

Place: New Delhi

Date: 29th May 2019

ACKNOWLEGEMENT

I gratefully acknowledge my profound indebtedness towards my esteemed guide, Mr. Yashdeep Singh, Assistant Professor at Delhi School of Management, DTU, for his invaluable guidance, excellent supervision and constant encouragement during the entire duration of the project work. This project would never have been possible without his guidance and supervision.

I also express my sincere thanks to Dr. Rajan Yadav, Head of Department and all the faculty members of Delhi School of Management, DTU, for being a constant source of encouragement.

ABSTRACT

Customer Churn refers to a situation in which a customer has terminated a relationship with a company, and this is an expensive issue. Customers are the fuel that stimulates a company. The loss of customers affects sales. Moreover, it is much more difficult and expensive to win new customers than to keep existing customers. As a result, organizations need to focus on reducing customer churn.

Machine Learning can be advantageous in solving this problem. For many businesses that offer subscription based services, it's critical to both predict customer churn and explain what features relate to customer churn. Older techniques such as logistic regression can be less accurate than newer techniques such as deep learning, which this project demonstrates i.e an ANN in R with the keras package.

For many business problems, it's equally important to explain what features drive the model, which is why LIME package is used for explainability. LIME results are cross-checked with a Correlation Analysis using the corrr package for increased accuracy.

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. Artificial Neural Networks (ANN) are now a staple within the sub-field of Machine Learning called Deep Learning. Deep learning algorithms can be vastly superior to traditional regression and classification methods (e.g. linear and logistic regression) because of the ability to model interactions between features that would otherwise go undetected. The challenge becomes explainability, which is often needed to support the business case. Keras and LIME provides best of both techniques.

TABLE OF CONTENTS

CERTIFICATE FROM THE INSTITUTE ii
DECLARATION iii
ACKNOWLEGEMENT iv
ABSTRACT v
LIST OF FIGURES viii
1. INTRODUCTION 1
1.1 Telecommunication Industry 1
1.2 Customer Churn 2
2. LITERATURE REVIEW 6
2.1 Machine Learning 6
2.2 Deep Learning 7
2.3 Neural Networks 8
2.4 Artificial Neural Networks (ANN) 10
2.5 R Programming 11
2.5.1 R Packages 12
3. RESEARCH METHODOLOGY 14
4. CHURN MODELING WITH ARTIFICIAL NEURAL NETWORKS (KERAS) 15
4.1 Churn Modeling Overview 15
4.2 Scope of the Study 15
4.3 Objective of the Study 15
4.4 Data Collection Source 16
4.5 Loading Libraries 16
4.6 Importing Data 16
4.7 Preprocessing Data 17
4.7.1 Pruning the Data 17
4.7.2 Splitting into Train/Test Sets 18
4.8 Exploratory Data Analysis 19
4.9 Discretizing Data 21
4.10 One Hot Encoding 23
4.11 Transforming Data with recipe library 24
4.11.1 Creating the recipe 24
4.11.2 Applying the recipe with bake() function 25
4.11.3 Storing the actual values as y_train_vec and y_test_vec 26

4.12 Deep Learning Model....26
4.13 Making Predictions....28
4.14 Inspecting Performance With Yardstick....30
4.15 Explaining Model with LIME Package....31
4.16 Checking Explanations with Correlation Analysis....32
4.17 Feature Investigation....34
5. FINDINGS OF CHURN MODEL....36
5.1 Deep Learning Training Result....36
5.2 LIME Feature Importance Visualization....36
5.3 LIME Feature Importance Visualization....38
5.4 Churn Correlation Analysis....39
5.5 Feature Investigation Results....40
REFERENCES....43
ANNEXURE....44
List of Abbreviations....44

LIST OF FIGURES

Figure 1 Types of Churn……. ……..……..4

Figure 2 Layers of ANN…….…….....11

Figure 3 Steps of building Churn Model……14

Figure 4 Churn Model Overview……..……..15

Figure 5 Total Charges Histogram…….22

Figure 6 One Hot Encoding Chart…….....……23

Figure 7 Deep Learning Training Result……...…36

Figure 8 LIME Feature Visualization……37

Figure 9 Feature Importance Visualization……38

Figure 10 Churn Correlation Analysis……...….……39

Figure 11 Feature Correlation Charts…….40