

**AN OPTIMIZED EARLY CHURN PREDICTION FOR
FOOD INDUSTRY BY DEEP NEURAL DECISION
FORESTS**

Kiruthiga Balendran

(219353R)

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

July 2023

**AN OPTIMIZED EARLY CHURN PREDICTION FOR
FOOD INDUSTRY BY DEEP NEURAL DECISION
FORESTS**

Kiruthiga Balendran

(219353R)

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree
Master of Science in Computer Science specialization in Data Science.

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

July 2023

DECLARATION

I declare that this thesis is my own work and it doesn't contain any sort of external material without an acknowledgement. Also, according to my knowledge, I haven't used any previously published materials to prepare the thesis. All the work carried out to prepare the thesis is individual work and do not copy any existing work without an acknowledgement.

I like to grant permission to the University of Moratuwa to reuse or share my work in any sort of format with external parties. Also, I like to keep my right to use my work in future.

Signature: B.Kiruthiga

Date:19-07-2023

The above candidate has researched the Master's thesis Dissertation under my supervision.

Name of the supervisor: Dr.Uthayasanker Thayasivam

Signature of the supervisor:

Date:

ABSTRACT

Customer retention is one of the main goals in large-scale food industries. This is achieved by predicting churn customers in advance and satisfying their needs. This research focuses on providing a higher accurate model than the other previous models. The deep neural decision forest model is a combination of Random Forest and Convolutional Neural Network models that help in fulfilling the objective. The dataset is taken from a US food industry to train and test the models. This model has achieved 92% accuracy in predicting the churn customers.

Keywords: Churn prediction, Deep neural decision forest, food service industry

ACKNOWLEDGEMENT

I wish to sincerely thank my supervisor, Dr Uthayasanker Thayasivam for providing me with the necessary guidance throughout the project. He provided me with ideas from the beginning of the project and helped me with finding the research area. Also, special gratitude should be given to our MSc coordinator for encouraging me to full fill my research under dedicated timelines.

Also, I would like to extend my gratitude to all the academic staff of the University of Moratuwa for the great contribution they made to me during my study. Also, I am grateful to my family members and all the friends who helped me throughout the project.

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT	ii
ACKNOWLEDGEMENT	iii
LIST OF FIGURES	vi
LIST OF TABLES	vi
LIST OF ABBREVIATIONS	vii
1 INTRODUCTION	1
1.1 Research Problem	2
1.2 Research Objectives	3
2 LITERATURE REVIEW	4
2.1 Introduction to Customer Relationship Management	4
2.2 Introduction to Customer Churn Prediction	4
2.3 Machine Learning Prediction Techniques	5
2.4 Deep Learning Prediction Techniques	7
2.5 Data Mining Techniques	8
2.6 Graph Representation Learning	8
2.7 Existing Research Summary	11
3 METHODOLOGY	14
3.1 Convolutional Neural Networks	14
3.2 Random Forests	14
3.3 Deep Neural Decision Forests	14
3.3.1 Learning the Forest	14
3.3.2 Learning the trees from Back-Propagation	16
3.4 System Architecture	17
3.5 Data Analysis	18
3.6 Data Preprocessing	25
3.6.1 Data Cleaning	25
3.6.2 Label Encoding	25
3.6.3 Data Transformation	26
3.6.4 Data Preparation	26

3.7	Implementation.....	26
3.8	Testing	26
4	EVALUATION.....	28
4.1	Test Results of CNN.....	28
4.2	Test Results of Random Forests	28
4.3	Test Results of Deep Neural Decision Forest	28
4.4	Evaluation of Model Parameters	29
4.5	Test Results for Bank Domain	29
4.6	Discussion	30
5	CONCLUSION.....	31
6	REFERENCES.....	32
7	APPENDIX A: Training Summary of Deep Neural Decision Forest.....	36

LIST OF FIGURES

Figure 1 - Integrated Telco Customer Analytics Framework	5
Figure 2 - A Framework with Ensemble Techniques	6
Figure 3 - Deep Learning Model in the Retail Domain	8
Figure 4 - Graphical representation of the undirected and directed graph.....	9
Figure 5 - Schematic diagram for proposed method in [19]	11
Figure 6 - Illustration of Deep Neural Decision Forest [26]	15
Figure 7 - Sample Routing Decision Function [26]	16
Figure 8 – High level Methodology	18
Figure 9 - Customer Churn Status.....	20
Figure 10 - Non-churned customer medal status	21
Figure 11 - Churned customer medal status.....	22
Figure 12 - Customer Region Level.....	23
Figure 13 - Market Name Vs Churn Status.....	24
Figure 14 - Data Model	25

LIST OF TABLES

Table 1 - Existing Research Summary	11
Table 2 - Data Description	19
Table 3 - Features in Dataset.....	19
Table 4 - Label Encoding for Customer Medal Status.....	26
Table 5 - Experiments of CNN	28
Table 6 - Experiments of Random Forest	28
Table 7 - Experiments of Deep Neural Decision Forest	29
Table 8 - Evaluation of Model Parameters	29
Table 9 - Test Results for Bank Domain.....	30

LIST OF ABBREVIATIONS

Abbreviation	Description
B2C	Business to Customer
B2B	Business to Business
CRM	Customer Relationship Management
DNN	Deep Neural Networks
FMCG	Fast-Moving Consumer Goods
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
RF	Random Forest
EGB	Extreme Gradient Boosting
CART	Classification and Regression Trees
LR	Logistic Regression
MARS	Multivariate Adaptive Regression Splines
SB	Stochastic Boosting
kNN	k-Nearest Neighbours
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
MLP	MultiLayer Perceptron
DT	Decision Trees
ANN	Artificial Neural Network
LDA	Linear Discriminant Analysis
SVMauc	SVM based on the AUC parameter-selection technique
FIM	Frequent Itemset Mining
H-MK-SVM	Hierarchical Multi-Kernal Support Vector Machine
GWO	Grey Wolf Optimization
MSISDN	Unique number to identify subscriptions in mobile network
CSV	Comma-Separated Values
PII	Personal Identifiable Information
GNN	Graph Neural Networks
SGD	Stochastic Gradient Descent