e-Learning Product Sales Prediction using SMS Activation Data

V. S. W. Wickramasekara
189474B

Faculty of Information Technology
University of Moratuwa

July 2021

e-Learning Product Sales Prediction using SMS Activation Data

V. S. W. Wickramasekara
189474B

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of Degree of Master of Science in Information Technology.

July 2021

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student
V. S. W. Wickramasekara

Signature of Student

Date

Supervised by

Name of Supervisor
Mr. S. C. Premaratne

Signature of Supervisor

Date

Acknowledgements

I am so grateful to my supervisor Mr. Saminda Premaratne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for academic guidance, advice, time, encouragement and patience that has seen me through, making this work success.

Also, I would like to give a special thanks to my beloved family members for their support, love, advice and encouragement throughout the period of my study. I lastly wish to extend my sincere gratitude toward my colleagues who have tirelessly guided and encouraged me at all times through my course.

.

Abstract

Most companies nowadays using data mining techniques and algorithms for the decision-making process and to make profitable adjustments in operation and production.

This helpful to identifying possibilities of a customer buying products after buying a particular product, identifying a group of customers who's in the difficulty in the activation process/ identifying customers who are going to drop from the company/ how it effects for the company, identifying connection between customer purchasing methods, find the group of customers to promote particular product/s and forecasting the product sales process (required quantity of each product for special annual events like exhibitions).

Understanding of consumers helps companies to sell more. Customers are a valuable source of information, collection of data that organizations identify different customers and observe how they behave. The more knowledge about consumers and their different needs, it is easier to identify market opportunities to sell and invent new products and target consumers with appropriate offers. Use of company past data, production and operations can predict using data mining techniques and algorithms.

Table of Contents

Declaration i
Acknowledgements ii
Abstract iii
Table of Contents iv
List of Figures vi
List of Tables viii
Chapter 1 1
Introduction 1
1.1 Background and Motivation 1
1.2 Problem Statement 2
1.3 Aim and Objectives 2
1.3.1 Aim 2
1.3.2 Objectives 2
1.4 Scope of Research 3
1.5 Propose solution 3
1.6 Structure of the theses 4
Chapter 2 5
Literature Review 5
2.1 Introduction 5
2.2 The observation of objectives 5
2.3 Summary 8
Chapter 3 9
Environment Setup 9
3.1 Introduction 9
3.2 Machine Learning Tools 9
3.2.1 Packages and Tools 9
3.3 Summary 10
Chapter 4 11
Research Methodology 11
4.1 Introduction 11
4.2 Main Phases 11
4.3 Business understanding 12
4.4 Data understanding 12
4.5 Data Preparation, Modeling and Evaluate Algorithms 13

4.5.1 Area 1 14
4.5.2 Area 2 23
4.5.3 Area 3 35
4.5.4 Area 4 44
4.5.5 Area 5 60
4.6 Summary 70
Chapter 5 71
Conclusion and Future Work 71
5.1 Introduction 71
5.2 Conclusion 71
5.3 Future Developments 72
5.4 Summary 72
References 73

List of Figures

Figure 1: Main Phases 11
Figure 2: Database Activated Table 13
Figure 3: Area 1-Program code part 1 15
Figure 4: Area 1-Program code part 2 15
Figure 5: Area 1-Program code part 3 16
Figure 6: Area 1-Program code part 4 16
Figure 7: Area 1 arff data file 17
Figure 8: Apriori output part 1 18
Figure 9: Apriori output part 2 19
Figure 10: FilteredAssociator output part 1 20
Figure 11: FilteredAssociator output part 2 21
Figure 12: FilteredAssociator output part 3 22
Figure 13: Selected Best Rules from Apriori algorithm 23
Figure 14: Area 2-Program code part 1 24
Figure 15: Area 2-Program code part 2 25
Figure 16: Area 2-Program code part 3 26
Figure 17: Area 2-Program code part 4 27
Figure 18: Area 2-Program code part 5 27
Figure 19: Area 2 arff data file 28
Figure 20: Bayes.NaiveBayes output 29
Figure 21: Trees J48 output 30
Figure 22: Trees J48 Visualizer output 31
Figure 23: Trees.RandomForest output 32
Figure 24: knowledge flow 33
Figure 25: Multiple ROC Curves 34
Figure 26: Area 3-Program code part 1 36
Figure 27: Area 3-Program code part 2 37
Figure 28: Area 3-Program code part 3 38
Figure 29: Area 3-Program code part 4 38
Figure 30: Area 3 arff data file 39
Figure 31: Apriori output part 1 40
Figure 32: Apriori output part 2 41
Figure 33: FilteredAssociator output part 1 42
Figure 34: FilteredAssociator output part 2 43
Figure 35: Selected Best Rules from Apriori algorithm 44
Figure 36: Area 4-Program code part 1 45
Figure 37: Area 4-Program code part 2 46
Figure 38: Area 4-Program code part 3 47
Figure 39: Area 4 arff data file 48
Figure 40: EM with training set cluster output part 1 49
Figure 41: EM with training set cluster output part 2 50
Figure 42: EM with training set cluster output part 3 51
Figure 43: EM with class attribute cluster output part 1 52
Figure 44: EM with class attribute cluster output part 2 53
Figure 45: EM with class attribute cluster output part 3 54

Figure 46: SimpleKMeans with training set cluster output part 1 55
Figure 47: SimpleKMeans with training set cluster output part 2 56
Figure 48: SimpleKMeans with class attribute cluster output part 1 57
Figure 49: SimpleKMeans with class attribute cluster output part 2 58
Figure 50: SimpleKMeans with class attribute cluster output part 3 59
Figure 51: Area 5 arff data file 60
Figure 52: Linear Regression Prediction output 61
Figure 53: Linear Regression Error Rates evaluation 62
Figure 54: SMOreg Prediction output 63
Figure 55: SMOreg Error Rates evaluation 64
Figure 56: Multilayer Perceptron Prediction output 65
Figure 57: Multilayer Perceptron Error Rates evaluation 66
Figure 58: Gaussian Processes Prediction output 67
Figure 59: Gaussian Processes Error Rates evaluation 68

List of Tables

Table 1: Error values for four algorithms 69
Table 2: Linear Regression and Multilayer Perceptron Prediction values 69