

Deliverable 1: Final Year Dissertation

Sahil Pattni

BSc. Computer Science (Honours)

Supervisor: Neamat El Gayar

November 28, 2020

Declaration

I, Sahil Pattni, confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed: SMP

Date: November 28, 2020

Abstract

In this study, we will extract a minimum spanning tree (MST) from a data set using machine learning techniques. We will then use this minimum spanning tree to segment products together, and produce association rules and extract the most interesting rules. The resulting rule-set will be compared to rules generated by the established Apriori Algorithm [1], which is the foundation of association rule mining. Additionally, we will study how the structure of our MST would change before and after promotional events, leading to insights that may help the management in such firms make better informed decisions about the type of promotions they would like to run.

Contents

1	Introduction	3
1.1	What is a Minimum Spanning Tree?	3
1.2	Aims	3
1.3	Objectives	3
2	Background	3
3	Research Methodology	3
4	Evaluation Strategy	3
5	Project Management	3

1 Introduction

1.1 What is a Minimum Spanning Tree?

Given an undirected graph G with edges E and vertices V (i.e. $G(E, V)$), a *spanning tree* can be described as a subgraph that is a tree [2] which includes all the vertices V of G with the minimum number of edges required. A *minimum spanning tree* is the spanning tree with the smallest sum of edge weights.

1.2 Aims

The aim of this study is to study the effectiveness of a minimum spanning tree in product classification and association rule mining, in addition to determining how the architecture of the MST will change during and after large events such as promotions. The model will be tested on a relatively large dataset of sales data.

1.3 Objectives

The research objectives for this project have been laid out below, in the order that they will be carried out.

1. Acquire a suitable dataset upon which the MST can be constructed.

2 Background

3 Research Methodology

4 Evaluation Strategy

5 Project Management

References

- [1] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *Proceedings of the 20th International Conference on Very Large Databases*, 1994.
- [2] E. Williamson, *Lists, Decisions and Graphs*. S. Gill Williamson, p. 171.

Appendix