PROPOSAL FOR DATA MINING – ONLINE RETAIL

Shannon School of Business, Cape Breton University

Data Mining – MGSC-5126 Section 22

Jamileh Yousefi

February 22nd, 2024

Team - 13

Name	Student Id	Student Email
Simarpreet Kaur	0285168	cbu22clbn@cbu.ca
Krina Chiragkumar Patel	0283599	cbu22cgxl@cbu.ca
Tanvi Atulbhai Ajmera	0283994	cbu22cgsc@cbu.ca
Prabhjot Singh	0278865	cbu22bxfn@cbu.ca

Project Description

Online retail has become a significant component of the global economy, with businesses increasingly relying on digital platforms to reach customers. The abundance of data generated from online transactions presents an opportunity to extract valuable insights through data mining techniques. This proposal outlines a data mining project focused on the Online Retail dataset available at the UCI Machine Learning Repository.

ONLINE RETAIL: https://archive.ics.uci.edu/dataset/352/online+retail

The Online Retail dataset contains transactional data from an online retail store based in the United Kingdom. It includes records of transactions between 01/12/2010 and 09/12/2011, involving customers from various countries. Each record contains information such as the invoice number, stock code, description of the product, quantity sold, transaction date, unit price, and customer ID.

- Dataset Characteristics - Multivariate, Sequential, Time-Series
- Subject Area – Business
- Associated Tasks - Classification, Clustering
- Instances – 541909
- Feature Type - Integer, Real
- Features – 6
- Has Missing Values - No

Objective Statement

The primary objectives of this data mining project are:

- To identify patterns and trends in customer purchasing behavior.
- To segment customers based on their buying habits and preferences.
- To forecast future sales and product demand.

- To optimize marketing strategies and improve customer retention.
- To detect anomalies or irregularities in transaction data.

Methodology

For the purpose of this project, the methodology to be used in the Recency, Frequency, and Monetary model. Using this model the customers will be segmented into various meaningful groups using the k-means clustering algorithm and decision tree induction.

Project Timeline

The proposed timeline for the data mining project is as follows:

- Data Preprocessing and EDA
- Customer Segmentation
- Association Rule Mining
- Predictive Modeling
- Anomaly Detection
- Report Writing and Presentation

The project will be completed by the march 18th 2024 after completing all the necessary modules and presentation.

Final Product/Expected

By the end of this project, we aim to understand the customer buying patterns in regards to what combination of products are frequently purchased together by which customers and customer groups. We are also expecting to enhance the merchant's website to track customer's shopping activities instantaneously and accurately. We will also be predicting each customer's lifecycle value to quantify the level of diversity of each customer.