**UNIVERSITY OF HERTFORDSHIRE, HATFIELD**

**SCHOOL OF ENGINEERING AND COMPUTER SCIENCE**

**MSC COMPUTER SCIENCE (DATA SCIENCE AND ANALYTICS)**

**MODULE: MSC CS PROJECT**

**DETAILED PROJECT PROPOSAL**

**PROJECT TITLE:**

**A STUDY ON CUSTOMER BEHAVIOUR ANALYSIS FOR RETAIL OPTIMIZATION USING K-MODE AND ASSOCIATION RULE MINING**

**STUDENT: AMOBI OBIDIEBUBE**

**LEVEL: 7**

**ACADEMIC YEAR: 2024-25**

**Table of Contents**

1. **Aim of the project -----------------------------------------------------**
2. **Research question/hypothesis ---------------------------------------**
3. **Project Objectives -----------------------------------------------------**
4. **DESCRIPTION OF PROJECT IDEA-----------------------------------------**
5. **METHODOLOGY ----------------------------------------------------------------**
6. **REFERENCES --------------------------------------------------------------------**

**1.0 Project Goal/Aim**

This project aims to develop a data-driven framework for analyzing customer behaviour using K-Modes clustering and association rule mining, to support retail optimization strategies. This includes identifying distinct customer groups and uncovering frequently co-purchased products to inform targeted marketing and inventory decisions.

**2.0** **Research Question/Hypothesis**

This study investigates:

1. How effectively K-Modes clustering can segment retail customers based on categorical behavioral attributes.
2. Whether association rule mining can uncover meaningful product purchase relationships within these segments.
3. Whether integrating K-Modes clustering and association rule mining leads to better retail decision-making than using either method independently.

**Hypothesis:** The combined use of K-Modes clustering and association rule mining will produce more actionable and insightful outcomes for customer behaviour analysis and retail optimization than using either technique alone.

**3.0 Project Objectives**

**Literature Review**: Examine current methods in customer segmentation and market basket analysis, focusing on K-Mode and association rule mining.

**\*** To acquire and preprocess a real-world retail dataset containing categorical variables..

\*To implement K-Modes clustering to identify meaningful customer segments

\*To apply association rule mining (Apriori or FP-Growth) to identify frequent product combinations.

\*To analyze the patterns revealed and present data-driven recommendations for retail optimization.

**\*** To compare the results with those from existing research studies..

**\*** Document the process, findings, and suggest real-world retail strategies based on insights.

**4.0 DESCRIPTION OF PROJECT IDEA**

This project is centered around creating a machine learning–based analytical system that aids retailers in understanding customer behaviour by analyzing categorical retail data. Using **K-Modes clustering**, customers will be grouped into meaningful segments such as frequent shoppers, high spenders, or category-specific buyers (Dinh et al., 2025; Thapaliya & Zhuang, 2025). **Association rule mining** techniques will then be applied to perform market basket analysis, revealing which products are frequently bought together (Hashad et al., 2024). This combination provides a deeper understanding of customer needs and product affinities.

The concept is inspired by the real-world retail industry's increasing reliance on AI and advanced analytics to enhance business intelligence and consumer targeting.

**5.0 METHODOLOGY**

**Data Collection**: A real-world dataset will be selected from platforms such as Kaggle or the UCI Machine Learning Repository, preferably one that has been used in academic research to enable comparative analysis. The dataset will include categorical fields such as product types, transaction types, payment methods, and user segments.

**Preprocessing**: This includes cleaning missing or inconsistent data, encoding categorical variables, and formatting for compatibility with the k-modes and association mining libraries in Python.

**Clustering**: Using the K-Modes algorithm to categorize customers based on their purchasing behaviours and other categorical features. The number of clusters will be determined using cost-based evaluation.

**Market Basket Analysis**: Utilizing association rule mining algorithms such as Apriori or FP-Growth to identify product associations within clusters. Evaluation metrics will include support, confidence, and lift.

**Evaluation**: Effectiveness of clustering will be assessed based on intra-cluster similarity and interpretability, while association rules will be evaluated using standard rule interestingness metrics. Results will be compared with related studies.

**6.0 REFERENCES**

* Dinh, T., Wong, H., Fournier‑Viger, P., Lisik, D., Ha, M.Q., Dam, H.C. & Huynh, V.N. (2025) 'Categorical data clustering: 25 years beyond K‑modes', *Expert Systems with Applications*, 272, Article 126608.
* Thapaliya, B. & Zhuang, Y. (2025) 'Fast clustering of categorical big data', *arXiv*, 2502.07081.
* Hashad, A.A., Khaw, K.W., Alnoor, A. & Chew, X. (2024) 'Exploratory analysis with association rule mining algorithms in the retail industry', *Malaysian Journal of Computing*, 9(1), pp. 1746–1758.