Project Stage I Report on

Developing a Unified Customer Segmentation Framework using Multi-Industry Behavioral Data

Submitted in partial fulfillment of the requirements for the Degree of

Bachelor of Engineering in Artificial Intelligence & Data Science

by

Vipashyana Jawale (2347003) **Anish Date (2347005)**

Shubham Keskar (2347031) **Yashraj Devrat (2347075)**

Under the guidance of

Dr. P. M. Paithne

Vidya Pratishthans

Kamalnayan Bajaj Institute of Engineering and Technology

Baramati-413133, Dist-Pune India

Department of Artificial Intelligence & Data Science

Academic Year 2023-24

VPKBIET, Baramati

Certificate

This is to certify that the Project Stage I Report on

Developing a Unified Customer Segmentation Framework using Multi-Industry Behavioral Data

SUBMITTED BY

Vipashyana Jawale (2347003) **Anish Date (2347005)**

Shubham Keskar (2347031) **Yashraj Devrat (2347075)**

in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering in Artificial Intelligence & Data Science at Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, under the Savitribai Phule Pune University, Pune. This work is done during year 2023-24 Semester-I, under our guidance.

Dr. P. M. Paithne	Dr. P. M. Paithane	Dr. R. S. Bichkar
Project Guide	Head of Dept.	Principal

Examiner 1:——- Examiner 2: - - - - - - -

Acknowledgements

We would like to express our deepest gratitude to Dr. P. M. Paithane, whose unwavering guidance and expertise played a pivotal role in the successful completion of this project. His invaluable insights and mentorship were instrumental in shaping the direction and execution of our work. We are also thankfull to Dr. P. M. Paithane, Head of the Department, for his constant support and encouragement throughout this endeavour. His belief in the project and continuous motivation provided the impetus needed to overcome challenges and achieve our goals. We extend my heartfelt thanks to all the teachers and faculty members of the department for their invaluable contributions, both academically and personally. Their collective efforts created an environment conducive to learning and innovation. Finally, we would like to acknowledge the principal Dr.R.S.Bichkar , their support have been a source of inspiration and a guiding light throughout this journey.

Mr.Anish Date

Ms.Vipashyana Jawale

Mr.Yashraj Devrat

Mr.Shubham Keskar

Abstract

This study introduces a novel approach for customer segmentation in an application promotion system known as multi-behaviour RFM (MB-RFM). MB-RFM considers different user-item interactions such as clicking, favouriting, and adding to cart, as opposed to earlier RFM models that simply include purchase behaviour. To deconstruct numerous consumer actions and examine the weight relationship between users and things, the self-organizing map (SOM) method is used. Using the superiority chart and entropy value approaches, we examined the weight relationship between numerous user behaviours and products. Furthermore, several advertising methods are established based on client groups in order to increase application adoption and perform targeted promotions. The suggested method's usefulness is validated using real-world datasets, which show significantly enhanced classification performance compared to traditional methods, especially in sparse datasets.

Technical Keywords:

RFM modelling , Customer value, Cluster analysis, Self-Organizing Maps method (SOM), Multi-behaviour

List Of Figure

- PROPOSED SYSTEMS ARCHITECTURE
- DATA FLOW DIAGRAM LEVEL-0
- DATA FLOW DIAGRAM LEVEL-1
- USECASE DIAGRAM
- SEQUENCE DIAGRAM
- ACTIVITY DIAGRAM

Notation and Abbreviations

- RFM: Recency, Frequency and Monetary value
- MB-RFM: Multi-Behavior Recency, Frequency and Monetary value
- SOM: Self-Organizing Maps
- GRE: Graduate Record Examination
- ML: Machine Learning
- NN: Neural Network
- DBSCAN: Density-based spatial clustering of applications with noise

Chapter 1

Introduction

1.1 Introduction

In the limitations of conventional RFM models that primarily focused on purchase behavior, neglecting crucial user-item interactions, this study introduces a pioneering approach known as Multi-Behavior RFM (MB-RFM). Leveraging the self-organizing map (SOM) algorithm, MB-RFM systematically analyzes recency, frequency, and monetary values alongside various user behaviors like clicking, favoriting, and adding to cart. Through methods like the superiority chart and entropy value analysis, the study establishes weight relationships between these behaviors. The resulting MB-RFM model values are employed in an improved SOM neural network for customer segmentation. The study then tailors promotion strategies based on customer categories, aiming to enhance application utilization and implement targeted promotional efforts. Experimental validation using real-world datasets, particularly in sparse conditions, affirms the significant accuracy improvement in customer classification offered by the proposed MB-RFM method.

1.2 Motivation

The traditional rfm model only focused on purchase behavior, neglecting other important user behaviors such as clicking, favoriting, and adding items to the cart. However, these additional behaviors offer valuable insights into user preferences and interactions with the application. The novel solution, the multi-behavior RFM (MB-RFM) model, addresses this limitation by deconstructing multiple user behaviors and using the self-organizing map (SOM) algorithm for customer segmentation. By incorporating a broader range of user behaviors and analyzing their weight relationship with items, the MB-RFM model enables a more comprehensive understanding of customer preferences and facilitates the development of targeted promotion strategies. The ultimate goal is to enhance client segmentation efficacy, improve consumer categorization performance and accuracy, and support application providers in implementing more effective advertising tactics to increase client engagement and satisfaction.

Chapter 2

Literature Survey

2.1 Literature Survey

Sr.No.	Paper	Technique Used	Remarks
1	Juan Liao et.al	MB-RFM model to analyze multiple behaviors	Considers various behaviors instead of one.
2	A. Joy Christy et.al.	Repetative K-Means Algorithm	RM K-Means problem with clusters
3	A. Syaifudin et.al	Fuzzy C-Means clustering, Genetic Programming to optimize FCM	Comparison with other clustering algorithms not shown.
4	Hanaa Hachimi et.al	statistical clustering method	Insufficient explanation of CLV factor calculation in RFM-D model

Chapter 3

Proposed Work

3.1 Problem Definition

To enhance the previous RFM modeling technique to integrate different customer behaivours and implement effective marketing strategy.

3.2 Project Objectives

- To divide customers into homogeneous clusters based on their RFM values, identify distinct customer segments with specific characteristics and properties.
- Gain insights into customer preferences and needs and understand product development and innovation.
- Develop targeted acquisition strategies to attract new customers who align with existing segments.

3.3 Scope of Project

- Data Analysis: Analyze user behaviors, including purchase behavior and other interactions, using real-world datasets. Explore the relationships and weightage between these behaviors using advanced analytical techniques.
- Model Development: Develop a multi-behavior RFM (MB-RFM) model that incorporates the analyzed behaviors. Utilize the self-organizing map (SOM) algorithm for customer segmentation based on behavior patterns.
- Validation and Performance Evaluation: Evaluate the effectiveness and accuracy of the MB-RFM model using sparse datasets. Compare the classification performance of the proposed method with existing RFM models.
- Promotion Strategy Development: Design and develop various promotion strategies tailored to each customer category. Aim to improve application utilization and enhance the effectiveness of targeted promotions.
- Implementation and Results Analysis: Implement the developed MB-RFM model and promotion strategies in an application promotion system. Evaluate the system's performance and analyze the results to assess the impact on customer segmentation and promotional outcomes.

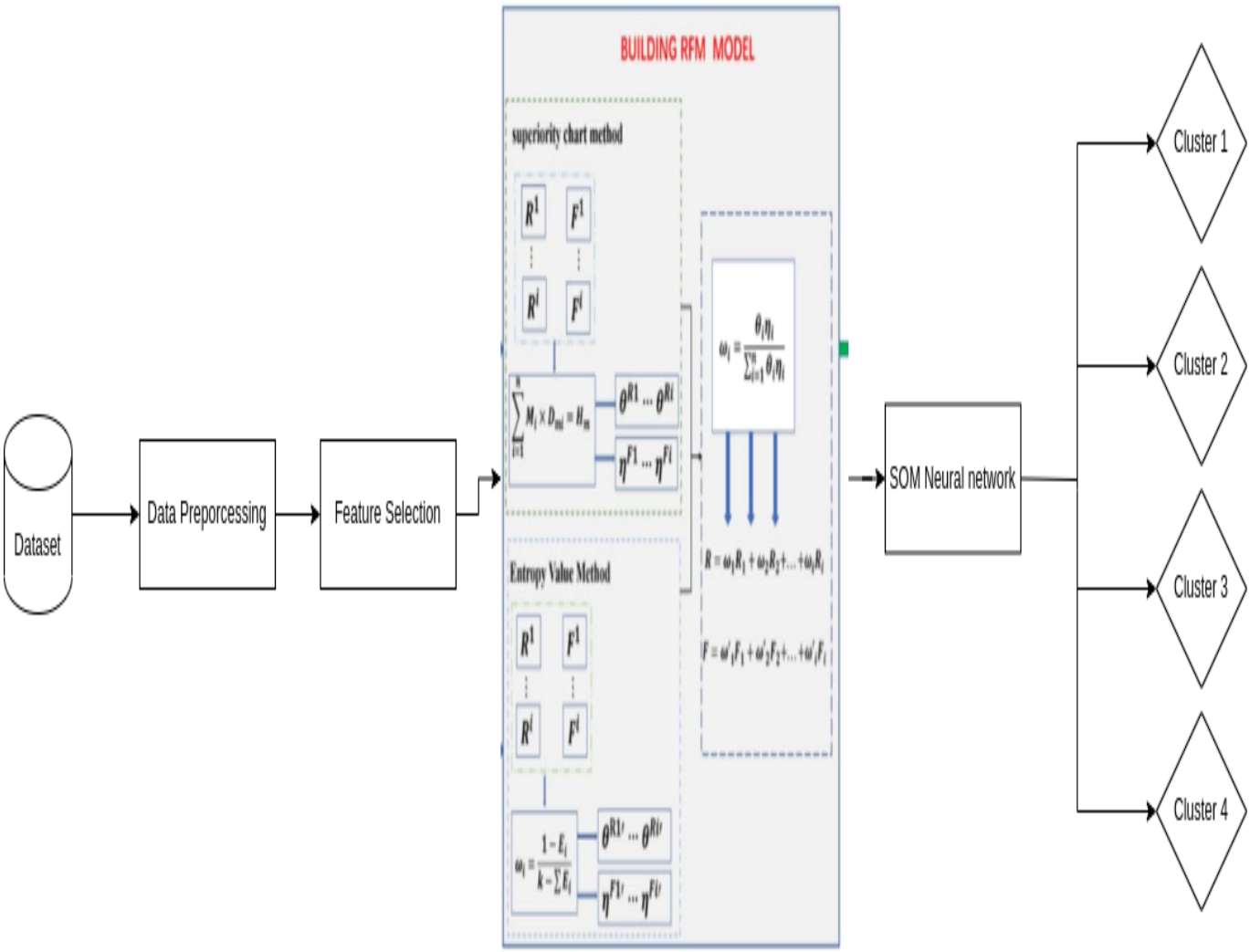
3.4 Project Constraints

- RFM models focused on analyzing only one type of user behavior data, such as purchase behavior, without considering other interactions between users and items.
- The proposed solution aims to deconstruct multiple behaviors of consumers, including clicking, favoriting, and adding to cart, within a specific period for customer segmentation.
- The self-organizing map (SOM) algorithm is used in the proposed multi-behavior RFM (MB-RFM) model for customer segmentation.
- The weight relationship between multiple behaviors of users and items is analyzed using the superiority chart and entropy value methods.
- The R, F, and M values are used to classify customers within the MB-RFM model, and an improved SOM neural network is employed for customer classification.
- The proposed method is evaluated using real-world datasets and demonstrates significantly more accurate classification performance in sparse datasets.
- The proposed model considers the particularity of different industrial sectors and incorporates additional factors, such as insurance compensation or customer lifetime value, to improve the RFM model.

Chapter 4

Proposed Systems Architecture

4.0.1 Proposed Systems Architecture



Chapter 5

Project Requirement Specification

5.1 Hardware Requirements

- Disk Space: 200 GB
- Processor: 11th Gen Intel Core i5
- GPU: NVDIA Corporation.
- RAM: 8 GB

5.2 Software Requirements

- Operating System: Linux
- OS Type: 64-bit
- Python Version: 3.11.4
- Tools: Google Colaboratory / Jupyter Notebook.

5.3 Performance Requirements

In designing a Self-Organizing Map (SOM) neural network for customer segmentation, it is imperative to establish specific performance requirements that align with business objectives. Accuracy and precision metrics should be defined to ensure the model accurately classifies customers into segments, minimizing false positives and negatives. Computational efficiency parameters, including training and inference times, must be set within acceptable limits to meet real-time processing needs. Scalability requirements should be considered to ensure the model can handle growing datasets or expanding customer bases. Robustness against noise and variations in input data, interpretability of segmentation results, and the ability to generalize to new data are crucial aspects to be addressed. Additionally, the model's adaptability to changing customer behavior patterns, user-friendly output formats, and compliance with data privacy and security regulations need to be specified. Establishing criteria for model validation and testing will further ensure ongoing performance evaluation. Collectively, these performance requirements create a comprehensive framework to guide the development and deployment of a SOM neural network for effective customer segmentation.

5.4 Software Quality Attributes/Requirements

Specify any additional quality characteristics for the product that will be important to either the customers or developers. some to consider are: adaptability, availability, correctness, flexibility, inter-operability, maintainability, portability, reliability, reusability, robustness, testability and usability. write these to be specific, quantitative and verifiable when possible. at the least, clarify the relative preferences for various attributes such as ease of use over ease of learning.

5.5 Security Requirements

Ensuring the security of a Self-Organizing Map (SOM) neural network for customer segmentation demands a comprehensive set of measures. Robust data security begins with the implementation of strong encryption protocols for both data at rest and in transit, safeguarding customer information from unauthorized access. Access controls, including role-based permissions, must be strictly enforced to limit system access to authorized personnel, ensuring only essential data is accessible. Rigorous authentication mechanisms and clear authorization protocols contribute to secure user interactions with the system. The establishment of detailed audit trails captures all system activities, providing a crucial tool for both monitoring and forensic analysis in the event of a security incident. Secure APIs, hosted in a protected environment, and regular security audits with penetration testing further fortify the system against potential vulnerabilities. Data minimization principles are essential, advocating for the collection and retention of only necessary customer data. Compliance with data protection regulations, such as GDPR or HIPAA, is non-negotiable, as is the development and maintenance of an incident response plan for swift and effective resolution of security breaches. Additionally, employee training programs ensure that personnel involved in system development and maintenance are well-versed in security best practices, mitigating both external and insider threats. By addressing these security requirements, the SOM neural network can operate within a robust and resilient framework, preserving the confidentiality and integrity of customer data throughout the segmentation process.

5.6 Safety

Ensuring the safety of a Self-Organizing Map (SOM) neural network for customer segmentation encompasses a multifaceted approach focused on ethical, transparent, and responsible use of the technology. Ethical considerations are paramount, guiding the collection and utilization of customer data in a manner that respects privacy and aligns with established ethical principles. Mitigating biases within the SOM is essential to guarantee fair and unbiased segmentation outcomes, preventing disproportionate impacts on specific demographic groups. Transparency and explainability are integral components, as clear insights into the model's decision-making processes foster user trust and understanding. Incorporating a human-in-the-loop approach allows human experts to intervene when necessary, ensuring oversight and addressing ethical concerns that may arise. Accountability mechanisms, adherence to regulations governing AI ethics, and robustness testing against adversarial attacks further fortify the model's ethical foundations. Continuous monitoring and regular reassessment of the model's impact on diverse customer groups contribute to ongoing safety. Additionally, user education initiatives clarify the SOM's capabilities and limitations, promoting responsible usage and managing expectations. By embracing these safety considerations, organizations can deploy the SOM for customer segmentation in a manner that prioritizes ethical standards, transparency, and the well-being of individuals impacted by the technology.

Chapter 6

Project Planning

6.1 Team Structure

Sr No.	Team Member	Role
1	Anish Date	Research, Documentation, Model Building
2	Yashraj Devrat	Research, Data Collection, Data Preprocessing
3	Vipashyana Jawale	Research, Documentation, Data Analysis
4	Shubham Keskar	Research, Model Validation, Deployment

Chapter 7

Project Schedule

7.1 Project Breakdown Structure

Month	Week	Action item
July-August	1-2	Project topic selection and synopsis submission
	3-4	Topic finalization
	5-6	Database collection
	7-8	Designing of algorithms and flow of project
Sept-October	1-2	Conduct market research to understand user needs and competition
	3-4	Define multi-criteria evaluation parameters for university selection.
	5-6	Design the user interface and wireframes for the application.

Project Plan 1.0

November - January	1-2	Set up the development environment and database.
	3-4	Develop the core functionality of the application.
	5-6	Implement the SOM Neural Network algorithm.
	7-8	Test the application for functionality and user experience.
February-March	1-2	Conduct thorough testing, including functional, usability, and performance testing.
	3-4	Collect user feedback through beta testing and make necessary adjustments.
	5-6	Address any bugs or issues identified during testing.

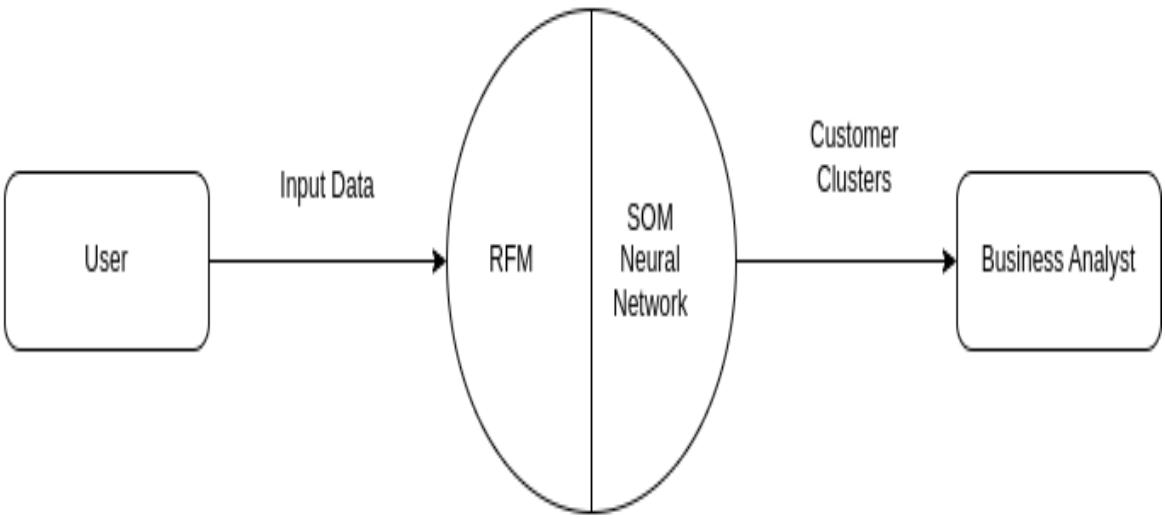
Chapter 8

Project Design

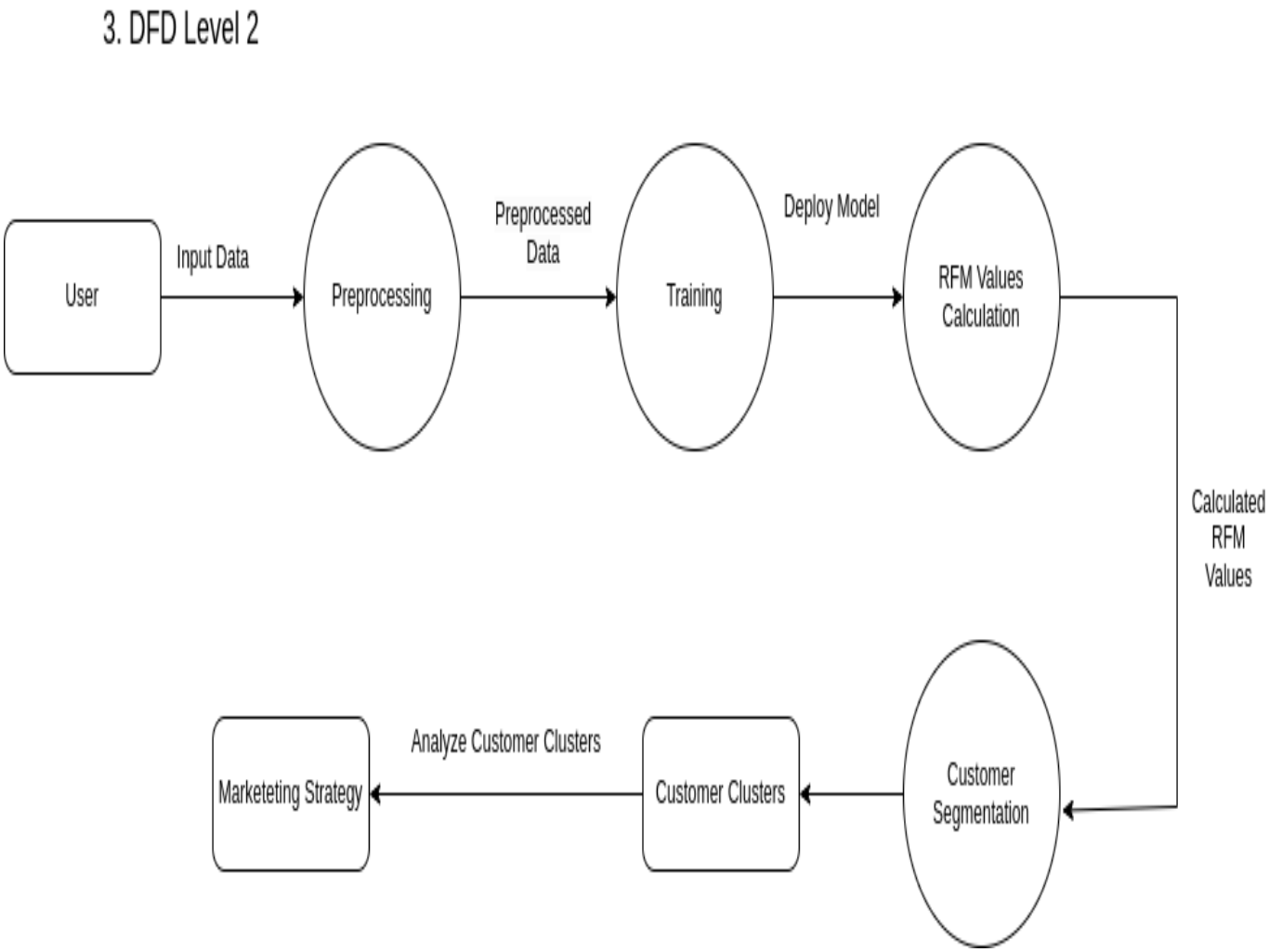
8.1 UML Diagrams / Data Flow Diagrams , Flowchart

8.1.1 DATA FLOW DIAGRAM LEVEL-0

1. DFD Level 0

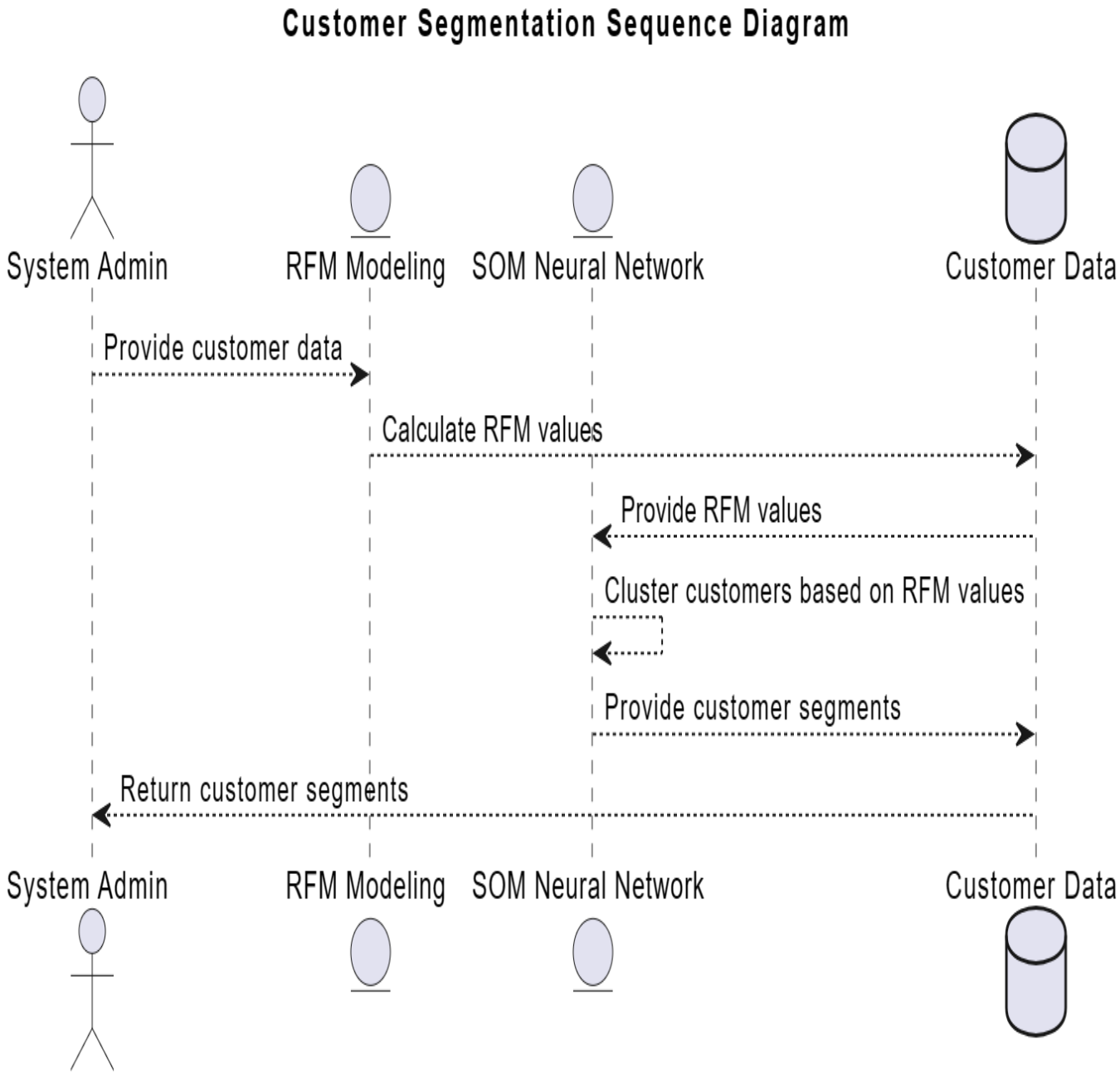


8.1.2 DATA FLOW DIAGRAM LEVEL-1

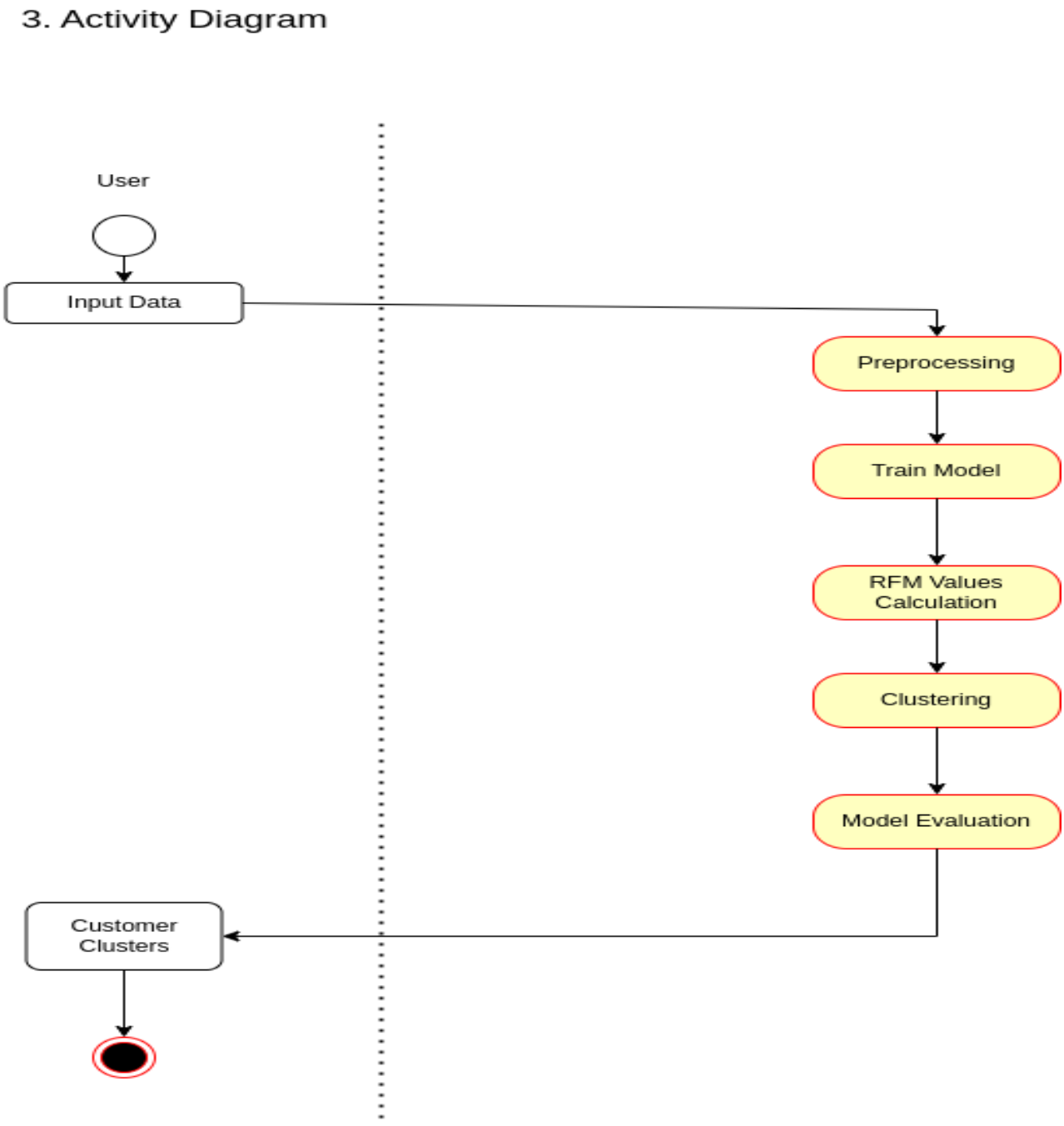


8.1.3 USE CASE DIAGRAM

8.1.4 SEQUENCE DIAGRAM



8.1.5 ACTIVITY DIAGRAM



Chapter 9

System implementation-code documentation

9.1 Customer Data Collection

This module is dedicated to the collection and processing of data related to customers behaviour. It involves several essential tasks, including web scraping, data cleaning, and formatting, all of which are aimed at preparing a high-quality dataset suitable for clustering purposes.

9.1.1 Code Documentation:

—

Project Overview: This code implements a Self-Organizing Map (SOM) neural network for customer segmentation. The SOM is designed to analyze customer data and identify distinct segments based on their features.

Requirements: - Python [version] - Libraries: numpy, sompy (if applicable), pandas, matplotlib (for visualization), etc.

Code Structure: The code begins by importing necessary libraries, including numpy, sompy (if applicable), pandas, and matplotlib for visualization. It then proceeds to load and preprocess customer data.

Main Class: ‘CustomerSegmentationSOM‘ The main class is ‘CustomerSegmentation-SOM‘, responsible for initializing and training the SOM model. Key parameters include the input data, SOM grid shape, number of epochs, and learning rate.

- Training (‘train‘): The ‘train‘ method trains the Self-Organizing Map on the input data over a specified number of epochs.

- Visualization (visualize segments): The visualize segments method generates visualizations of the segmented customers on the SOM grid.

Main Script:

The main script checks if the code is being run directly (‘if name == ”main”:‘ It loads and preprocesses customer data, initializes and trains the SOM model, segments new customers, and visualizes the results.

Usage Instructions: 1. Load and preprocess your customer data. 2. Initialize the ‘CustomerSegmentationSOM‘ class with your data. 3. Train the SOM model. 4. Segment new customers using the trained model. 5. Visualize the results.

Chapter 10

Conclusion

10.1 Conclusion

The proposed Multi-Behavior RFM (MB-RFM) model, integrating multiple user-item interaction behaviors through an improved SOM neural network, offers a more nuanced approach to customer classification compared to traditional RFM models. By leveraging transaction records from local applications in China, the MB-RFM model enables the extraction of valuable insights and the calculation of weights using advanced methods. The resulting SOM classification into seven customer categories provides a basis for tailored marketing strategies. From identifying key customers (Type 1) to understanding underappreciated clusters (Types 6 and 7), the MB-RFM model allows applications to strategically target specific customer segments, optimizing pricing policies, promotions, and personalized services for improved customer utilization and targeted item promotion.

References

[1] RFM ranking – An effective approach to customer segmentation, A. Joy Christy a, A. Umamakeswari a, L. Priyatharsini b, A. Neyaa 2018

[2] New RFM-D classification model for improving customer analysis and response prediction, Moulay Youssef SMAIL, Hanaa HACHIMI,2023

[3] Customer Segmentation Using Fuzzy-AHP and RFM Mode, Anu Gupta Aggarwal, Shweta Yadav, 2020.

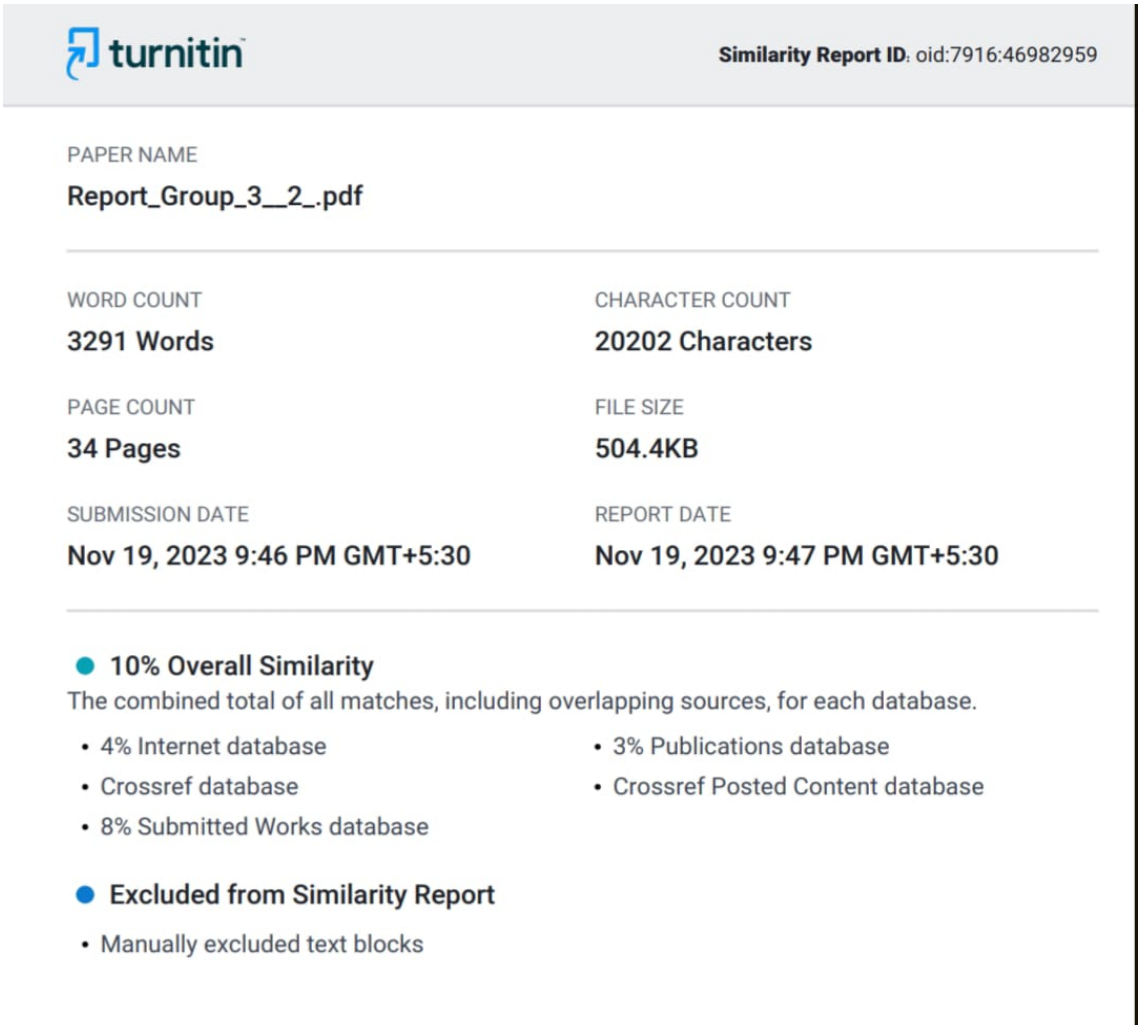
[4] Customer Segmentation with RFM Model using Fuzzy C-Means and Genetic Programming, Anas Syaifudin , Purwanto , Heribertus Himawan , M. Arief Soelema, 2023.

[5] Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods, Rahma Wati Br Sembiring Berahmana, Fahd Agodzo Mohammed, Kankamol Chairuang, 2020

Appendix A

Plagiarism Report

Plagiarism Report of Paper and Project Report from Turnitin software



turnitin

Similarity Report ID: oid:7916:46982959

PAPER NAME

Report_Group_3__2_.pdf

WORD COUNT

3291 Words

CHARACTER COUNT

20202 Characters

PAGE COUNT

34 Pages

FILE SIZE

504.4KB

SUBMISSION DATE

Nov 19, 2023 9:46 PM GMT+5:30

REPORT DATE

Nov 19, 2023 9:47 PM GMT+5:30

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 4% Internet database
- Crossref database
- 8% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Manually excluded text blocks

Appendix B

Base Paper

Juan Liao, Aman Jantan, Yunfei Ruan

"Multi Behavioral RFM Model Based on Improved SOM Neural Network"