CUSTOMER SEGMENTATION USING A COMBINATION OF RFM-BASED K-MEANS AND DECISION TREE METHODS (Case Study: Deepublish Publishing)

Muhamad Khusnur Rofiq
Data Science Study Program, Faculty of Science and Technology
University of Technology Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail: muhamadrofiq163@gmail.com

ABSTRACT

Deepublish Publishing faces challenges in understanding and managing the diverse transaction behaviors of its customers, resulting in suboptimal marketing strategies. Additionally, the company requires a model capable of mapping new customers into pre-defined segments. This study addresses these challenges by segmenting customers based on Recency, Frequency, and Monetary (RFM) values using the K-Means clustering algorithm, and subsequently classifying new customers into these segments using a Decision Tree model. The dataset consists of 6,855 transaction records with 16 attributes, collected between January 1, 2020, and December 29, 2023. The K-Means clustering process identified three distinct customer segments—loyal customers, new customers, and inactive customers—each characterized by different behavioral patterns. The Decision Tree model successfully classified customers into these segments with an accuracy of 99.67%, demonstrating its effectiveness in customer segmentation prediction. This research offers valuable insights for the company by enabling a deeper understanding of customer behavior, supporting the development of more targeted marketing strategies, and facilitating strategic engagement with new customers from the onset of their interactions.

Keywords: Segmentation, K-Means, RFM, Decision Tree, Classification.