ABSTRACT

In today's digital age, customer segmentation based on numerical data has become a critical need in strategic decision-making, particularly in the e-commerce sector. This study aims to develop a customer classification system using a machine learning approach, with a focus on the Support Vector Machine (SVM) algorithm and feature selection methods based on the Reduct algorithm. The objective is to enhance classification efficiency without compromising accuracy and the quality of prediction results. The dataset used is sourced from the USDA's Feed Grains Database, which contains numerical attributes such as price, harvest yield, stock, and supply. The data was processed through manual feature selection and Reduct to filter out relevant features. After the classification process, the results showed that although the model's accuracy value remained relatively constant (57.14%), there was an increase in precision and recall values, indicating improved segmentation quality between the Cheap and Expensive classes. This study concludes that the application of the Reduct algorithm can simplify data structure, enhance system efficiency, and maintain classification performance. This system has the potential to be applied to various other domains that rely on numerical data-based classification and can be developed into a customer analysis tool in e-commerce information systems.

Keywords — machine learning, support vector machine, feature selection, Reduct, customer classification