PROJECT REPORT

K-Means Clustering of Imported Products Based on HS Code and Product Attributes

Maura Dinda Apprilliani N.M
18.K1.0047

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT

As Indonesia continues to engage in international trade agreements like AFTA, the number of imported products has grown significantly—making it harder to manage import data efficiently. Relying on manual classification using the HS (Harmonized System) Code is no longer practical, especially when dealing with large datasets that include various brands, product categories, and countries of origin. This study offers a more practical approach by using the K-Means clustering algorithm to group imported products based on their attributes and HS Codes. The dataset includes product names, categories, subcategories, brands, units, countries of origin, and HS Codes, and was processed using Orange Data Mining. The data went through several steps including cleaning, selecting relevant columns, normalization, and assigning cluster labels as a new target. To determine the best number of clusters, silhouette scores were used as an evaluation metric. The results showed that splitting the data into three clusters gave the clearest pattern, with silhouette scores ranging between 0.48 and 0.58. Each cluster grouped products with similar traits, such as originating from the same country or belonging to the same product group. In the end, this research demonstrates that K-Means is a useful method for uncovering hidden patterns and improving the classification of imported products.

Keywords: clustering, imported products, HS Code, K-Means, data mining