# ABSTRAK

IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENENTUKAN POLA PEMBELIAN KONSUMEN

**Oleh:**
**Faiz Khoirurrizqi Noor**
**1217050053**

Pertumbuhan data transaksi dalam dunia bisnis menuntut adanya pengolahan data yang tepat agar dapat menghasilkan informasi bernilai bagi pengambilan keputusan. Salah satu pendekatan yang efektif adalah segmentasi konsumen untuk memahami perbedaan perilaku pelanggan. Penelitian ini bertujuan untuk menerapkan algoritma K-Means clustering dengan pendekatan RFM (Recency, Frequency, Monetary) sebagai dasar segmentasi konsumen. Proses penelitian dilakukan menggunakan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) yang mencakup tahapan *business understanding, data understanding, data preparation, modeling, evaluation,* hingga *deployment* secara konseptual. Untuk menentukan jumlah cluster yang optimal secara lebih objektif, digunakan kombinasi empat metrik evaluasi internal, yaitu Elbow Method, Silhouette Score, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Score (CHS). Selanjutnya, kualitas cluster yang terbentuk divalidasi menggunakan Analysis of Variance (ANOVA) guna memastikan perbedaan antar cluster signifikan secara statistik. Hasil penelitian menunjukkan bahwa jumlah cluster optimal yang diperoleh adalah lima cluster (k=5). Pendekatan multi-metrik dengan CRISP-DM ini terbukti mampu meningkatkan objektivitas dalam penentuan jumlah cluster sekaligus menghasilkan profil konsumen dengan karakteristik perilaku berbeda yang dapat dijadikan dasar dalam perumusan strategi pemasaran perusahaan.

Kata Kunci: *K-Means*, Segmentasi Konsumen, RFM, CRISP-DM, Evaluasi Multi Metrik, *Elbow Method, Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Score,* ANOVA

## ABSTRACT

*IMPLEMENTATION OF THE K-MEANS ALGORITHM TO IDENTIFY CONSUMER PURCHASING PATTERNS*

***By:***
**Faiz Khoirurrizqi Noor**
**1217050053**

*The rapid growth of transaction data in the business sector requires proper processing to generate valuable insights for decision-making. One effective approach is customer segmentation to better understand differences in consumer behavior. This study aims to implement the K-Means clustering algorithm using the RFM (Recency, Frequency, Monetary) model as the basis for customer segmentation. The research process was carried out using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which includes the stages of business understanding, data understanding, data preparation, modeling, evaluation, and a conceptual deployment. To determine the optimal number of clusters more objectively, four internal evaluation metrics were applied: Elbow Method, Silhouette Score, Davies-Bouldin Index (DBI), and Calinski-Harabasz Score (CHS). Furthermore, the quality of the resulting clusters was validated using Analysis of Variance (ANOVA) to ensure that the differences among clusters were statistically significant. The results indicate that the optimal number of clusters is five clusters (k=5). The combination of a multi-metric evaluation and the CRISP-DM methodology effectively enhances the objectivity in determining the optimal cluster number and produces customer profiles with distinct behavioral characteristics, which can serve as a foundation for developing data-driven marketing strategies.*



*Keywords:* *K-Means, Customer Segmentation, RFM (Recency, Frequency, Monetary), CRISP-DM, Multi-Metric Evaluation, Elbow Method, Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Score*, ANOVA