

## ABSTRAK

Di era digital ini, data menjadi inti dalam kelangsungan sebuah bisnis. Kebutuhan akan praktisi data semakin tinggi dalam industri saat ini. Data sendiri akan bisa menjadi informasi yang berguna apabila dilakukan pemrosesan data. Manusia dihadapkan pada melimpahnya data yang bisa didapatkan tanpa menguasai kemampuan untuk ekstraksi informasi di dalamnya. Dengan menggunakan data science dan algoritma Machine Learning, perusahaan dapat mengolah data dan memberikan nilai yang lebih untuk keperluan strategi bisnis. Salah satu contoh hasil pengolahan data dalam bisnis yaitu dengan membuat segmentasi pelanggan, segmentasi pelanggan bermanfaat untuk mengenali dan memfilter pelanggan dengan data yang ditentukan. Analisis dari segmentasi mampu memberikan alokasi target pasar yang lebih efektif, strategi marketing atau promosi yang lebih akurat, dan masih banyak lagi. Karena segmentasi bertujuan untuk memisahkan pelanggan ke beberapa grup atau *cluster* maka algoritma clustering dapat digunakan.

Pada Proyek Akhir ini akan dilakukan pengelompokan data pelanggan kartu kredit menggunakan algoritma K-Means, untuk menentukan jumlah *cluster* yang optimal, digunakan 2 metrik yaitu metode *elbow* dan *silhouette score* untuk memberika kurva yang memperlihatkan jumlah *cluster* yang optimal untuk digunakan pada data. *Principal Component Analysis* (PCA) adalah metrik yang akan digunakan pada Proyek Akhir ini untuk me-reduksi dimensi data menjadi 2 *principal component* yang memiliki nilai *varians* tertinggi, PCA 1 dengan 28% dan PCA 2 dengan 23%, jika dijumlahkan makan menjadi 51% dari 100% data.

Setelah dilakukan *data preprocessing* dan *features selection*, dari 18 *features* dan 8950 *records* yang terdapat pada data pelanggan kartu kredit, hanya akan digunakan 14 *features* saja untuk proses modeling, metode *elbow* dan *silhouette score* memberikan hasil bahwa dengan 14 *features* yang digunakan menghasilkan jumlah *cluster* yang optimal adalah sebanyak 2 *cluster*, hasil pengelompokan akan dibuat dalam visualisasi histogram dan *scatter plot* untuk mempermudah analisis pada setiap *cluster*.

**Kata Kunci:** *Data Science, Machine Learning, Clustering, K-Means,.*