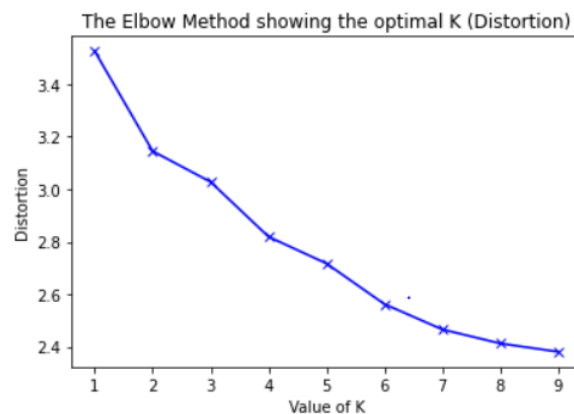


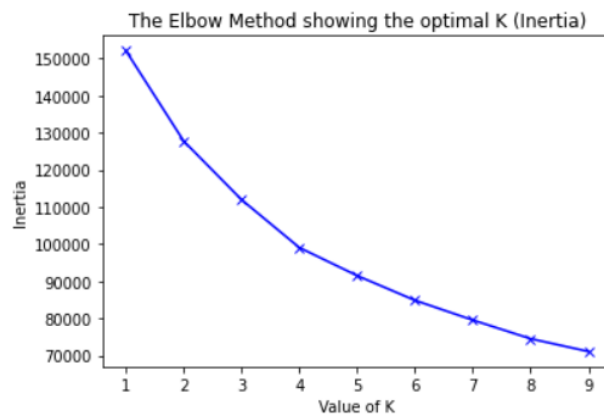
1. Buat pertimbangan dan analisis penentuan jumlah cluster. Hint: Untuk menentukan jumlah cluster optimal, gunakan elbow method.
2. Buat analisis dari tiap-tiap cluster yang terbentuk, contoh: Anggota cluster 1 memiliki rata-rata nominal CASHADVANCE (pengambilan uang) lebih besar daripada PURCHASE (pembelian), sehingga dapat disimpulkan bahwa anggota cluster 1 lebih membutuhkan kartu kredit untuk melakukan pengambilan uang tunai. Anggota cluster 2 memiliki rata-rata nilai ONEOFFPURCHASEFREQUENCY yang paling tinggi dibandingkan cluster lainnya sehingga anggota cluster ini kurang tepat apabila diberi promosi untuk cicilan. Dst.

Jawaban

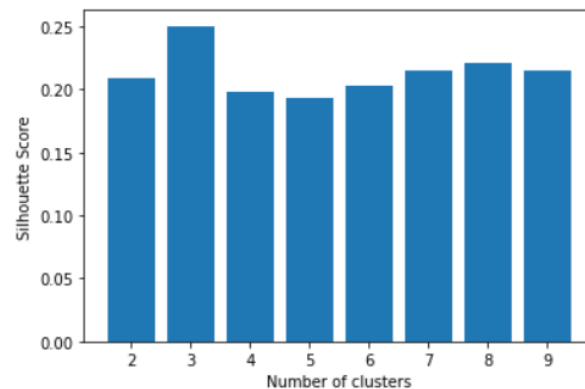
Pada proses clustering untuk dataset kartu kredit ini, metode yang digunakan yaitu K-means Clustering. Untuk penentuan jumlah cluster yang akan digunakan pada K-means Clustering ini yaitu dengan menggunakan elbow method. Pertama, kita akan melakukan elbow method dengan menggunakan distorsi yaitu rata-rata jarak kuadrat dari pusat cluster dari masing-masing cluster (dalam perhitungan ini metrik jarak Euclidean yang digunakan).



Tetapi ternyata elbow point tidak dapat ditentukan, maka selanjutnya kita akan melakukan elbow method dengan menggunakan SSE (Inertia).



Dengan melakukan elbow method dengan menggunakan SSE (Inertia), diketahui bahwa elbow point yang didapatkan yaitu 3. Setelah mendapat jumlah cluster yang optimal yaitu 3 pada elbow method, kita juga perlu melakukan perhitungan dengan menggunakan silhouette score untuk mengetahui apakah silhouette score yang didapatkan sesuai dengan hasil dari elbow method yang kita dapatkan tadi.



Dari silhouette score ini ternyata hasil yang didapatkan sesuai dengan hasil elbow method tadi, maka jumlah cluster yang kita gunakan pada K-means clustering ini yaitu 3 cluster.

Setelah melakukan K-means Clustering dengan jumlah cluster yaitu 3, berikut adalah analisa untuk masing-masing cluster yang dibentuk (cluster 0, cluster 1, cluster 2).

Pada variable yang pertama yaitu BALANCE, terlihat bahwa cluster 0 memiliki jumlah uang tersisa lebih sedikit dari cluster lainnya. Cluster 2 memiliki jumlah uang tersisa paling banyak dan cluster 1 berada diantara cluster 0 dan cluster 2. Selanjutnya, variabel yang kedua yaitu variable BALANCEFREQUENCY. Pada variable ini terlihat cluster 1 adalah yang paling sering memasukkan uang ke dalam kartu kredit. Diikuti dengan cluster 2 dan terakhir cluster 0. Untuk variable yang selanjutnya yaitu variable PURCHASES, ONEOFFPURCHASES, dan INSTALLMENTSPURCHASES. Dari ketiga variable ini dapat dilihat bahwa cluster 1 merupakan pengguna yang memiliki nominal transaksi terbesar baik itu merupakan transaksi sekali pembayaran maupun cicilan. Sedangkan kedua cluster lainnya memiliki nominal transaksi dibawah cluster 1. Selanjutnya, ada variable CASHADVANCE. Pada variable ini, cluster 2 terlihat sebagai cluster yang memiliki nominal paling terbesar untuk melakukan penarikan uang. Selanjutnya, ada tiga variable lainnya yaitu PURCHASESFREQUENCY, ONEOFFPURCHASESFREQUENCY, dan PURCHASESINSTALLMENTSFREQUENCY. Dari ketiga variable ini terlihat bahwa cluster 1 adalah pengguna yang paling sering melakukan transaksi baik itu transaksi sekali bayar atau cicilan. Selain itu, terlihat juga bahwa cluster 0 adalah pengguna yang lebih suka melakukan transaksi cicilan dan cluster 2 merupakan pengguna yang paling sedikit atau jarang melakukan transaksi. Selanjutnya, ada variable CASHADVANCEFREQUENCY. Dari variable ini terlihat bahwa cluster 2 adalah pengguna yang lebih sering melakukan transaksi penarikan uang. Selain itu, kedua cluster lainnya terlihat lebih sedikit melakukan transaksi penarikan uang. Selanjutnya ada variable CASHADVANCETRX dan PURCHASETRX. Dari kedua variable ini terlihat bahwa cluster 2 yang paling banyak melakukan transaksi penarikan uang dan cluster 1 yang paling banyak melakukan transaksi pembelian. Selanjutnya, ada variable CREDITLIMIT yang dimana cluster

0 memiliki batas nominal maksimal kartu kredit yang paling lebih rendah diantara kedua cluster lainnya. Selanjutnya, ada variable PAYMENTS. Pada variable ini, cluster 0 memiliki jumlah tagihan kartu kredit paling kecil dibanding cluster lainnya. Selanjutnya, ada variable MINIMUMPAYMENTS. Pada variable ini, cluster 0 memiliki batas pembayaran minimal paling terendah diantara cluster lainnya. Selanjutnya, ada variable PRCFULLPAYMENT. Pada variable ini terlihat bahwa cluster 1 adalah pengguna yang paling banyak melakukan pelunasan kartu kredit dengan pembayaran penuh. Untuk variable yang terakhir yaitu variable TENURE. Variable ini menunjukkan bahwa cluster 0 berisi pengguna yang memiliki jangka waktu kartu kredit paling lama.

Dari hasil analisa dari masing-masing cluster yang dibentuk, maka dapat disimpulkan bahwa cluster 0 berisi pengguna kartu kredit yang banyak melakukan transaksi pembelian dengan cara mencicil, cluster 1 berisi pengguna kartu kredit yang banyak melakukan transaksi pembelian dan sanggup membayar dengan sekali bayar, dan cluster 2 berisi pengguna kartu kredit yang jarang melakukan transaksi pembelian dan lebih sering menggunakan kartu kredit untuk menabung.