DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM PADA DAFTAR EMITEN SEKTOR KEUANGAN

**Hizvanno Ezra Herwindo**\*

1 2 Program Studi Sistem Informasi, Fakultas Teknik Ilmu Komputer Universitas Bakrie, Jakarta, Indonesia

E-mail: 1 vannochristian@gmail.com\*

**Abstrak**

***Kata kunci****: Cluster, Data Mining,K-Means, CRISP-DM, Keuangan*

**PENDAHULUAN**

**Latar Belakang**

Saat ini, konsep Data Mining semakin dikenal sebagai tools penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya. Data Mining sendiri sering disebut sebagai knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. Output dari Data Mining ini dapat digunakan untuk pengambilan keputusan di masa depan (Alfina, et al. 2012).

Salah satu teknik yang dikenal dalam Data Mining yaitu clustering. Pengertian clustering keilmuan dalam Data Mining adalah pengelompokan sejumlah data atau objek ke dalam cluster (group) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model cluster dan menghitung jumlah cluster yang optimal sehingga dapat dihasilkan cluster yang paling baik (Alfina, et al. 2012).

Clustering adalah suatu teknik dalam Data Mining yang digunakan untuk memasukkan data ke dalam grup yang bersesuaian tanpa pengetahuan yang mendalam tentang grup tersebut (Santosa, 2007).

Clustering digunakan untuk menganalisis pengelompokkan berbeda terhadap data, menyerupai klasifiaksi, namum pengelompokkan belum di definisikan sebelum dijalankannya toolData Mining.

Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota suatu kelas dan meminimumkan kesamaan antar cluster yang berbeda.

Adapun tujuan dari clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi antar cluster (Agusta, 2007).

**Tujuan Penelitian**

Tujuan dari penulisan penelitian ini adalah untuk mencari pusat cluster yang paling optimum dengan melakukan modifikasi terhadap algoritma K-Mean Clustering.

**METODOLOGI**

1. **Pengklusteran (Clustering)**

Pengklusteran merupakan pengelompokkan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Cluster adalah kumpulan recordyang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam cluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam

pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

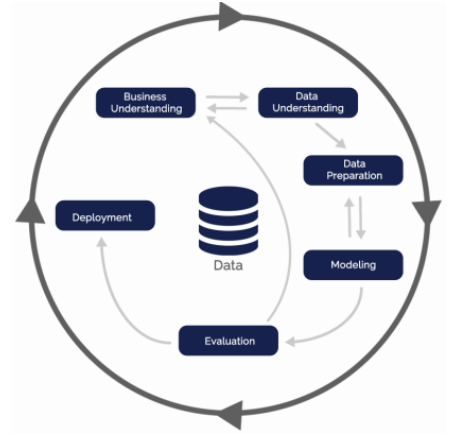
Contoh pengklusteran dalam bisnis dan penelitian adalah :

1. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.
2. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam keadaan baik ata mencurigakan.
3. Asosiasi (Assosiation)

Tugas asosiasi dalam Data Mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah :

1. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.
2. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
3. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons posistif terhadap penawaran upgrade layanan yang diberikan.
4. **Model Crisp-DM**

Pelaksanaan penelitian dilakukan dengan menggunakan metode Cross Industry Standard Process Model for Data Mining (CRISP-DM). Berikut gambaran alur proses yang terjadi pada CRISP-DM:



Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase. Yaitu:

*Business Understanding* adalah pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

*Data Understanding* adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasikan masalah yang berkaitan dengan kwalitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

*Data preparation* sering disebut sebagai fase yang padat karya. Aktivitas yang dilakukan antara lain memilih table dan field yang akan ditransformasikan ke dalam database baru untuk bahan data mining (set data mentah).

*Modeling* adalah fase menentukan tehnik data mining yang digunakan, menentukan tools data mining, teknik data mining, algoritma data mining, menentukan parameter dengan nilai yang optimal.

*Evaluation* adalah fase interpretasi terhadap hasil data mining yang ditunjukan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

*Deployment* atau penyebaran adalah fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining (Shearer, 2000).

1. **K-means**

Konsep dasar dari algoritma K-means adalah pencarian pusat cluster (centroid points) secara iteratif. Pusat cluster ditetapkan berdasarkan jarak setiap data ke pusat cluster. Proses clustering dimulai dengan mengidentifikasi data yang akan dicluster, xij (i=1,...,n; j=1,...,m) dengan n adalah jumlah data yang akan dicluster dan m adalah jumlah variabel.

**HASIL DAN PEMBAHASAN**

**Daftar Emitten**

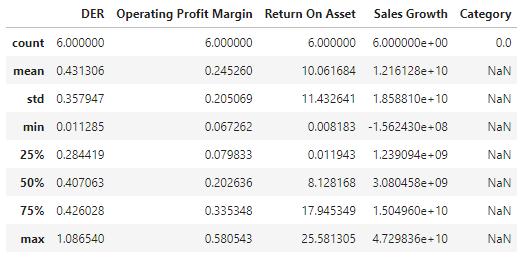
Daftar perusahaan yang dianalisa dari industri Keuangan terdapat 6 perusahaan seperti didalam tabel dibawah ini:



**Data financial Perusahaan dari daftar Emitten**

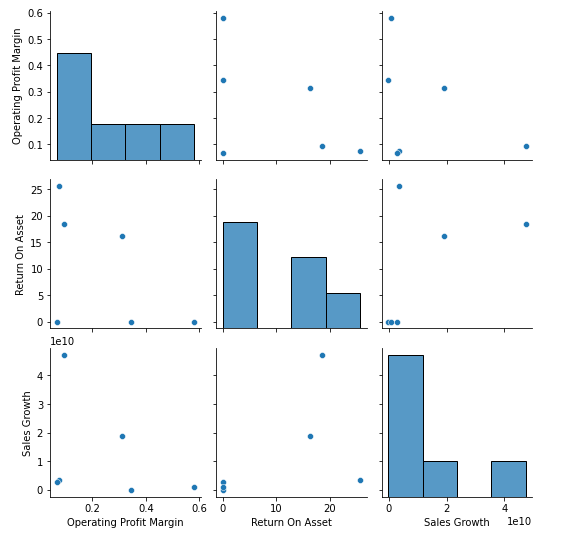
Data financial yang diambil dari yahoo finance terhadap keuangan di beberapa perusahaan beserta describe load data nya adalah seperti pada tabel berikut ini:



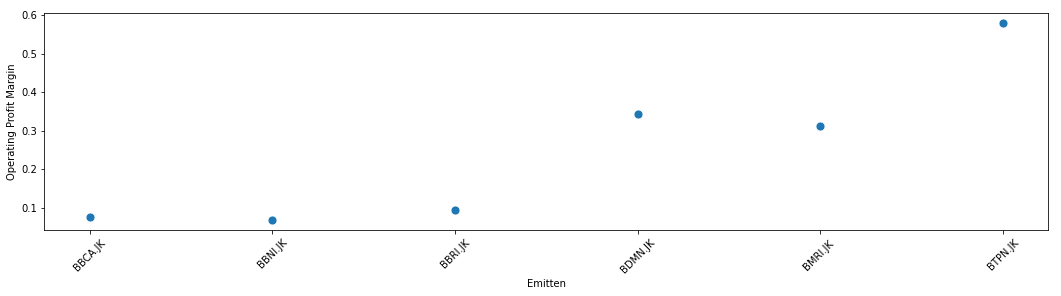


**Hasil scatter**

Dari hasil dataset yang sudah ada dapat dibuat penyebaran dalam bentuk scatter.



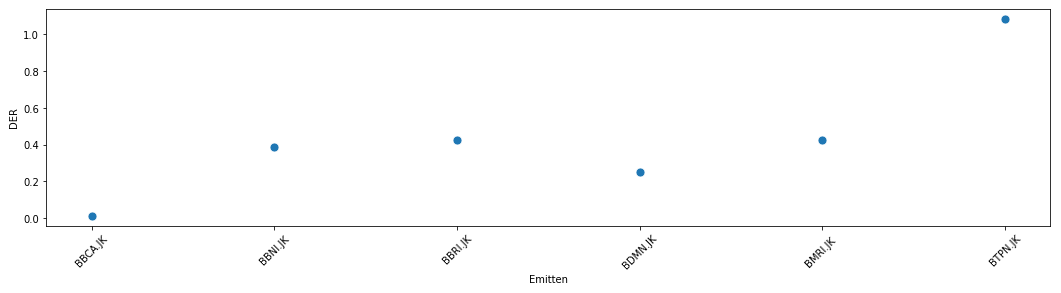
**Hasil Clustering Profit Margin Perusahaan**



Pada hasil cluster Profit Margin di sektor perusahaan keuangan dapat dilihat pada Bank BTPN.JK memiliki angka profit margin lebih tinggi dari perusahaan yang lainnya.

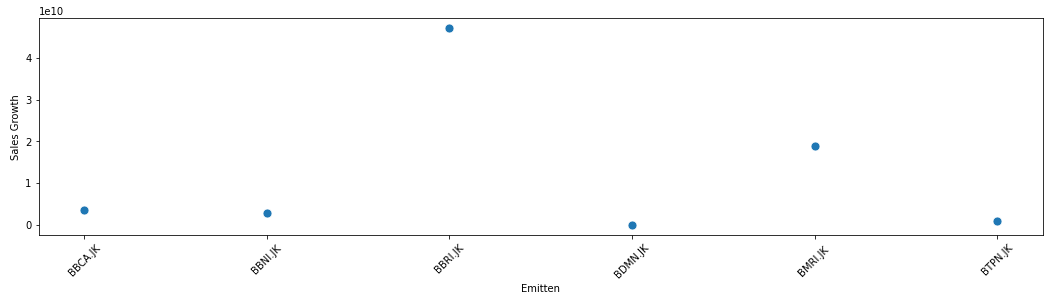
**Hasil Clustering DER Perusahaan**

Berikut adalah hasil clustering pada DER perusahaan, dan dapat dilihat pada perusahaan BTPN.JK lebih tinggi dan paing rendah pada perusahaan BBCA

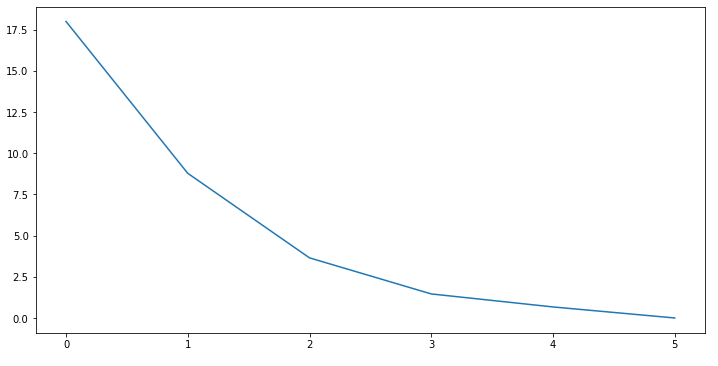


**Hasil Clustering Sales Growth**

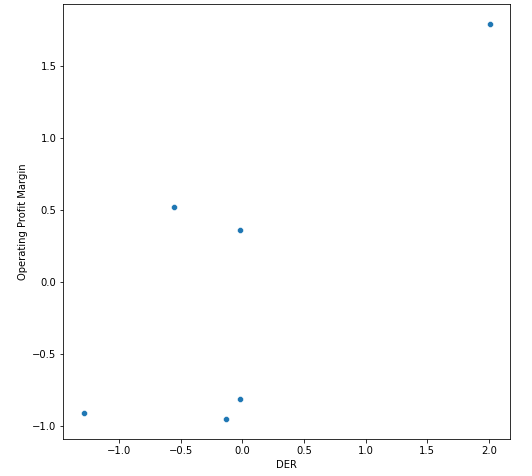
Dari hasil clustering untuk sales Growth yang tertinggi yaitu pada BBRI.JK



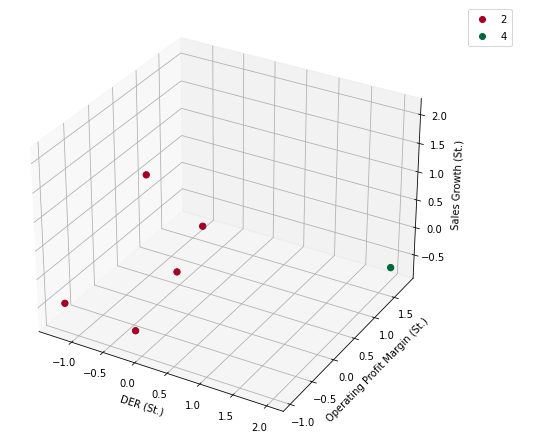
**K-Means**



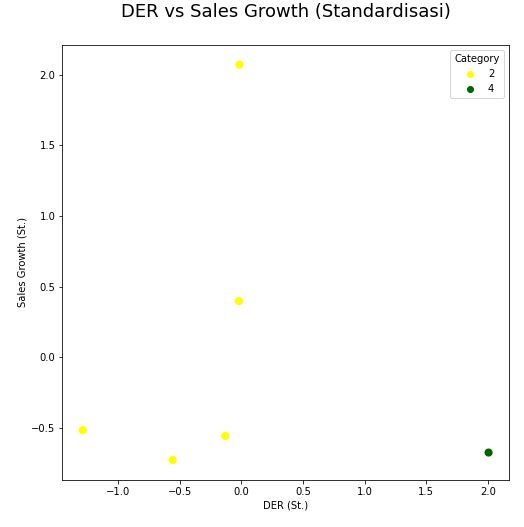
**Scatterplot**

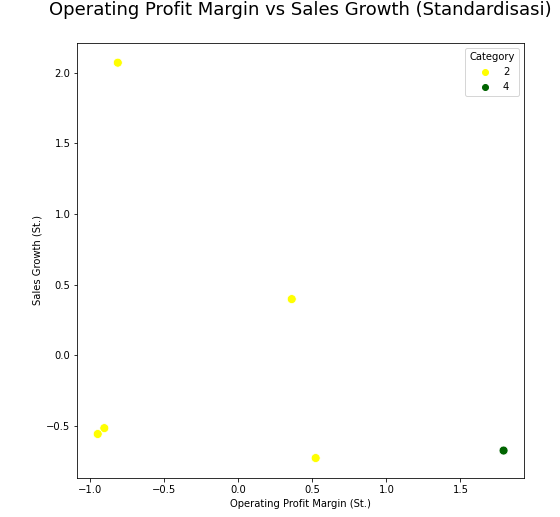


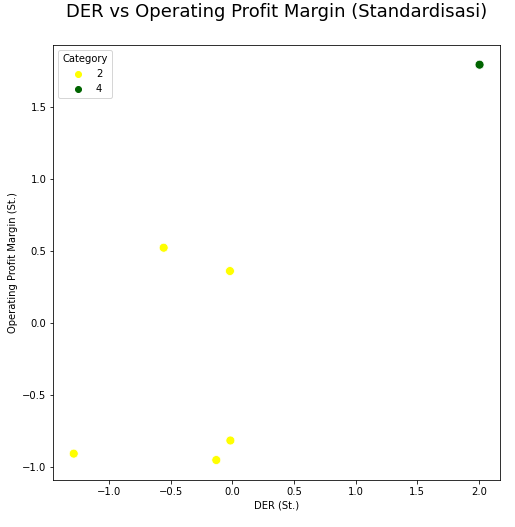
**3D scatterplot**



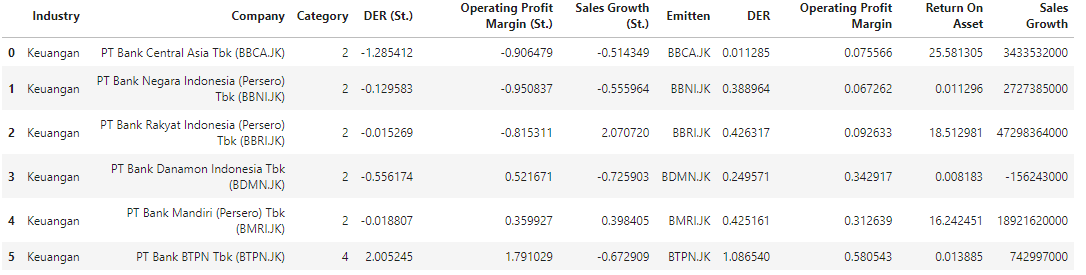
**X dan Y View**



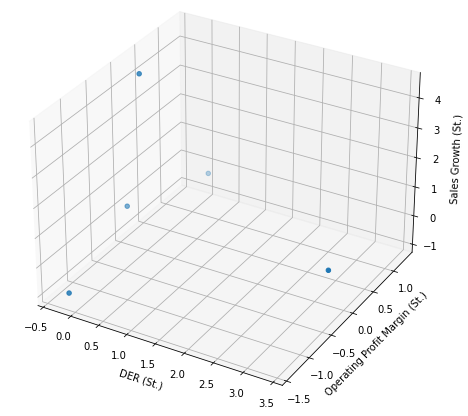




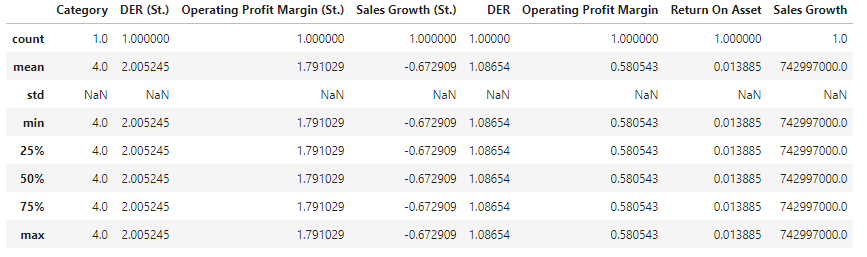
**Result**

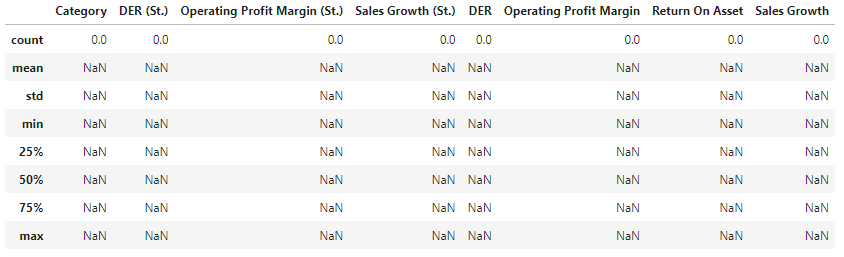


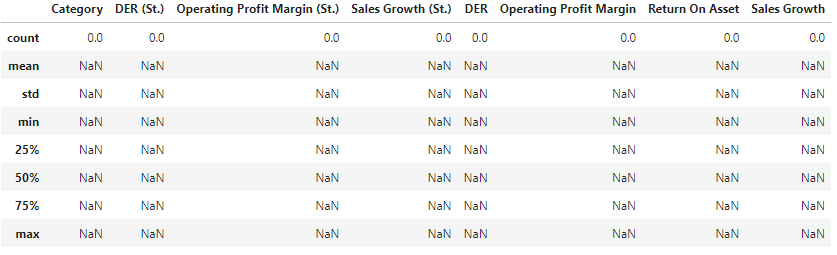
**Centroid**

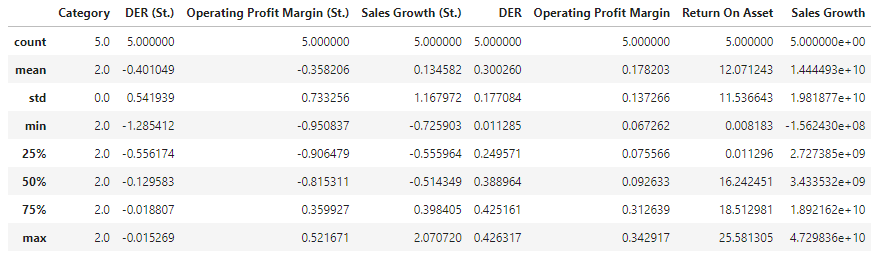


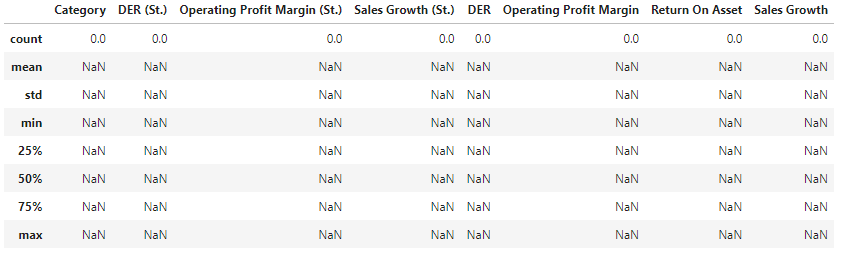
**Cluster Description**

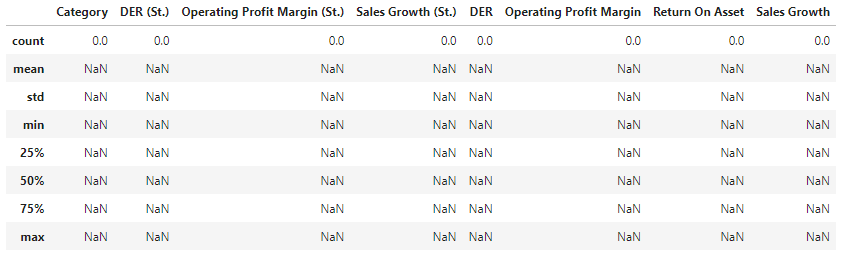












**KESIMPULAN**

Setelah dilakukan pengelompokan perusahaan industry keuangan melalui persebaran menggunakan K−Means clustering maka dapat di simpulkan sebagai berikut :

a. Setelah dilakukan pengelompokkan perusahaan keuangan melalui persebaran menggunakan K−Means clustering terbentuk menjadi 5 cluster

b. berdasarkan hasilnya terdapat 2 clluster yang optimal yaitu di dalam cluster 4 dan 2

**UCAPAN TERIMA KASIH**

**DAFTAR PUSTAKA**

**Penulisan *Journal Article***