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**

Pada masa pandemi saat ini banyak sektor industri yang terkena dampaknya, salah satunya adalah industri keuangan. Untuk menentukan apakah perusahaan mengalami kenaikan atau penurunan perlu diperhatikan laporan keuangan tahunan nya. Untuk menentukan perusahaan mana saja yang mengalami kenaikan atau penurunan dilakukan clustering data. 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 tool Data Mining. Dan pada saat ini penulis akan memperlihatkan bagaimana clustering dapat digunakan sebagai analisa terhadap data. Dan pada kasus ini penulis menggunakan data dari sektor keuangan atau bank yang diambil dari laporan keuangan tahunan dari situs BEI. Setelah diolah data nya dapat terbentuk 2 cluster yang optimal. Untuk cluster yang optimal ada dalam cateogry 4.

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

1. **PENDAHULUAN**
   1. **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).

* 1. **Tujuan Penelitian**

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

1. **METODOLOGI**
2. **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.

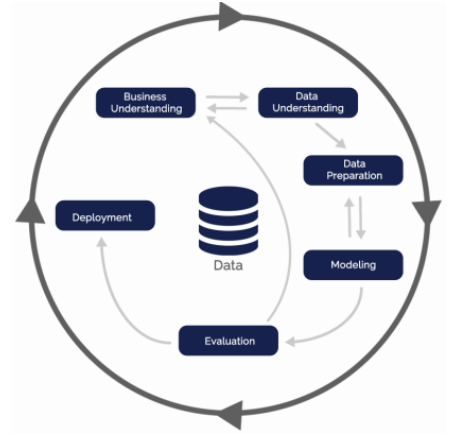
Contoh pengklusteran dalam bisnis dan penelitian adalah :

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**

K-Means merupakan algoritma yang umum digunakan untuk clustering dokumen. Prinsip utama K-Means adalah menyusun k prototype atau pusat massa (centroid) dari sekumpulan data berdimensi. Sebelum diterapkan proses algoritma K-means, dokumen akan di preprocessing terlebih dahulu. Kemudian dokumen itu akan dipresentasikan sebagai vektor yang memiliki term dengan nilai tertentu. Algoritma k-means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau dapat disebut sebagai centroid cluster.

• Konsep dasar dari K-Means adalah pencarian pusat cluster 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.

• Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang),

ckj (k=1,...,K; j=1,...,m).

• Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster.

• Untuk melakukan penghitungan jarak data ke-i (Xi) pada pusat cluster ke-k (Ck), diberi nama (dik), dapat digunakan formula Euclidean, yaitu:

1. **HASIL DAN PEMBAHA SAN**
   1. **Business Understanding**

Industri Jasa Keuangan (IJK) adalah kumpulan perusahaan/institusi dan lembaga pendukungnya yang berusaha di bidang jasa keuangan. Jenis Industri Jasa Keuangan yang diatur dan diawasi oleh Otoritas Jasa Keuangan meliputi; industri perbankan, industri pasar modal dan industri keuangan non bank yang terdiri dari Perasuransian, Dana Pensiun, Perusahaan Pembiayaan dan Pegadaian.

Investor non-residen masih mencatatkan inflow di pasar saham dan SBN. Spread yield US Treasury dibandingkan SBN sedikit mengalami penurunan. Hal yang sama juga terjadi penurunan yield pada korporasi di Indonesia

Dari sisi perbankan pada saat masa pandemi ini menurut Wimboh, serupa dengan kondisi pasar saham, bahkan sektor perbankan masih mencatatkan kinerja yang positif. Hal itu terlihat dari beberapa indikator.

* 1. **Data Understanding**

Daftar perusahaan yang dianalisa dari industri Keuangan terdapat 21 perusahaan seperti didalam tabel dibawah ini:

|  |  |
| --- | --- |
| Emitten | Company |
| AGRO | PT Bank Rakyat Indonesia Agroniaga Tbk |
| BBCA | PT Bank Central Asia Tbk |
| BBHI | PT Bank Harda Internasional Tbk |
| BBKP | Bank Bukopin Tbk |
| BBMD | PT Bank Mestika Dharma Tbk |
| BBNI | PT Bank Negara Indonesia (Persero) Tbk |
| BBRI | PT Bank Rakyat Indonesia (Persero) Tbk |
| BBSI | PT Bank Bisnis Internasional Tbk |
| BBTN | PT Bank Tabungan Negara (Persero) Tbk |
| BBYB | PT Bank Neo Commerce Tbk |
| BCIC | PT Bank JTrust Indonesia Tbk |
| BEKS | PT Bank Pembangunan Daerah Banten Tbk |
| BJBR | Bank Pembangunan Daerah Jawa Barat dan Banten Tbk |
| BJTM | Bank Pembangunan Daerah Jawa Timur Tbk |
| BKSW | PT Bank QNB Indonesia Tbk |
| BMRI | PT Bank Mandiri (Persero) Tbk |
| BNBA | Bank Bumi Arta Tbk |
| BNGA | PT Bank CIMB Niaga Tbk |
| BNII | PT Bank Maybank Indonesia Tbk |
| BNLI | Bank Permata Tbk |

Data financial yang diambil adalah data dari laporan keuangan di BEI terhadap keuangan di beberapa perusahaan

|  |
| --- |
| Emitten, DER,Operating Profit Margin,ROA,Sales Growth |
| AGRO,7.15,110.5129503,-0.09,-185.032137 |
| BBCA,4.88,984.5625988,0.03,-1485.583 |
| BBHI,4.36,0.122580674,0.01,121.8580899 |
| BBKP,10.03,56.90089514,-0.01,-1445.75 |
| BBMD,2.90,25.00227592,0.03,61.5454751 |
| BBNI,6.52,790,5351928,0.01,-4246.426 |
| BBRI,4.78,1043.240932,0.01,6531.747 |
| BBSI,0.49,0.269026929,0.04,22.55646717 |
| BBTN,15.82,218.4573561,0.01,129.291 |
| BBYB,6.31,3.922798839,-0.04,167.3732876 |
| BCIC,11.53,52.31315789,-0.02,-24.753 |
| BEKS,4.17,17.73977165,-0.03,-93,899 |
| BJBR,11.34,259.5426513,0.01,649.269 |
| BJTM,8.42,215.7843662,0.02,424.496 |
| BKSW,3.82,110.2022495,-0.02,-309.781 |
| BMRI,5.91,2143.12025,0.01,7034.205 |
| BNBA,3.49,0.361939939,0.01,-81.85642106 |
| BNGA,5.49,710.2705373,0.01,-1001.648 |
| BNII,4.50,64.46326531,0.01,-1690.584 |
| BNLI,5.07,22.41592814,0.01,263.701 |
| BTPN,3.94,390.3372123,0.01,-1667.699 |
| BTPS,0.36,74.46883943,0.08,567.248 |
| MASB,7.90,8.795872752,0.01,-2756.119855 |
| MEGA,5.54,179.2985033,0.03,335.8793442 |
| NISP,5.64,477.0781305,0.01,-1059.743 |
| PNBN,3.16,472.7519480,0.02,-1195.535 |
| SDRA,3.68,63.26626506,0.01,29.416 |

* 1. **Data Preparation**

Dibawah ini adalah contoh dari dataset yang sudah di read pada saat mau melakukan pengkodingan dan hasil nya seperti di tabel ini

Text

Description automatically generated

Table

Description automatically generated

Berdasarkan data nya didapat non-nuull count nya sebanyak 20 non-null dengan dtype float64 kecuali pada emitten dtype nya adalah object.

Chart, scatter chart

Description automatically generated

Table

Description automatically generated

* **Hasil Clustering Operating Profit Margin Perusahaan**

Chart, scatter chart

Description automatically generated

Pada hasil cluster Profit Margin di sektor perusahaan keuangan/Bank dapat dilihat bahwa BBNI memiliki Operating profit margins memiliki angka profit lebih tinggi dari perusahaan yang lainnya. Dan untuk perusahaan lainnya berada di nilaii Operating Margin perusahaannya rendah

**Hasil Clustering DER Perusahaan**

Berikut adalah hasil clustering pada DER perusahaan, BBNI lebih tinggi dan yang perusahaan lainnya rendah.

Chart, scatter chart

Description automatically generated

**Hasil Clustering Sales Growth**

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

Chart, scatter chart

Description automatically generated

* 1. **K-Means**

Chart, line chart

Description automatically generated

Dari hasil K- Means yang didapat jika diperhatikan titik siku nya ada di titik 2. Jadi untuk hasil data nya dihasilkan 2 cluster yang akan optimal

Chart, scatter chart

Description automatically generatedOperating Profit Margin dan DER, dapat dilihat hasil scatter plot nya bisa di perhatikan ada 3 cluster yang pertama nilai Operating Profit Margin nya tinggi, rendah dan di nilai tengah

**3D scatterplot**

Chart, scatter chart

Description automatically generated

**X dan Y View**

Chart, scatter chart

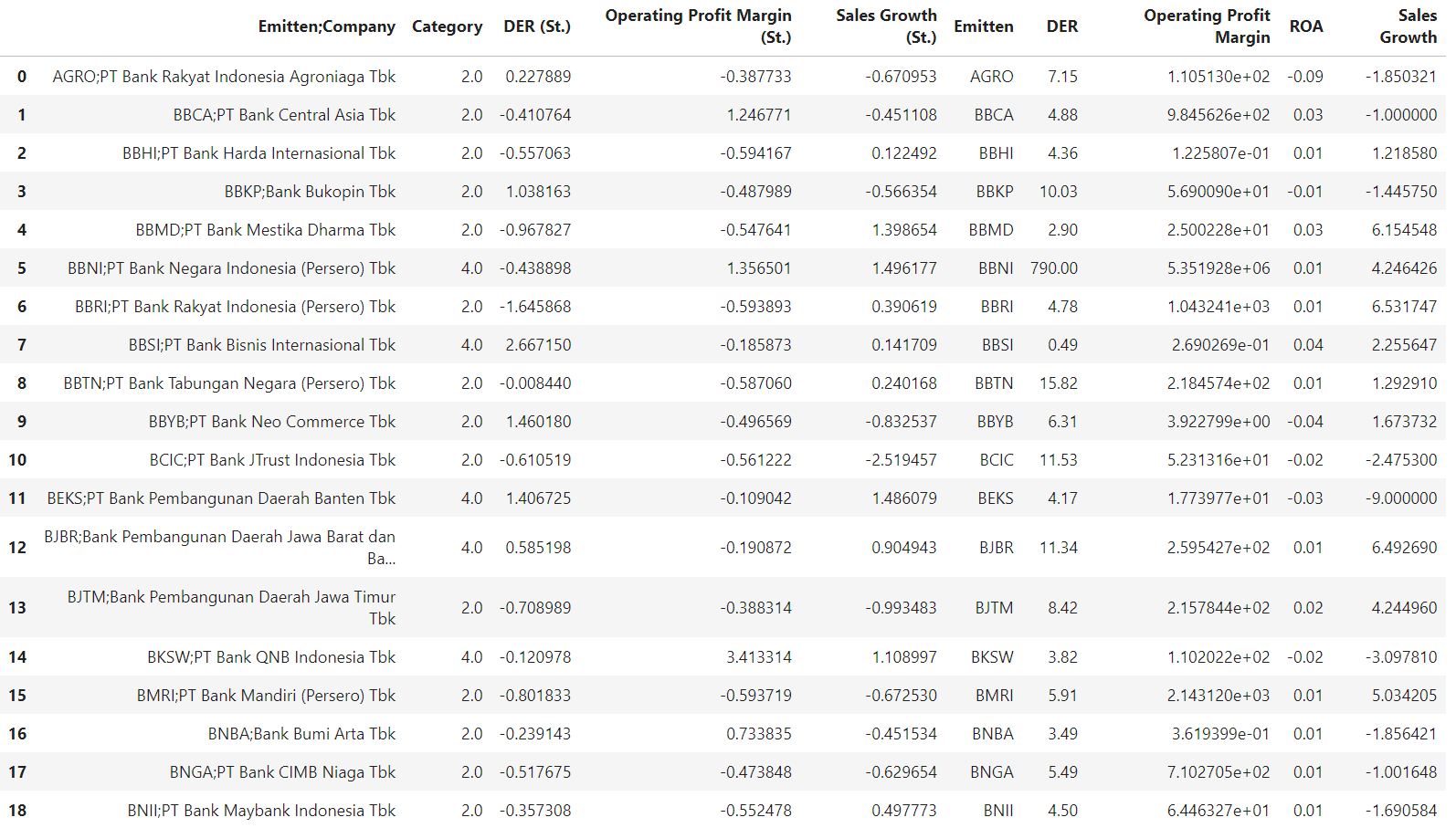
Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

**Result**

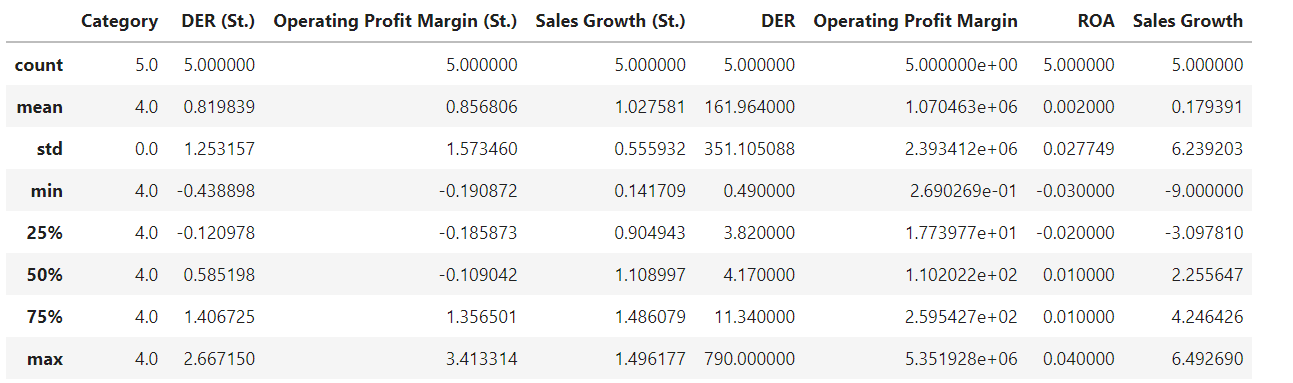


**Centroid**

Chart

Description automatically generated

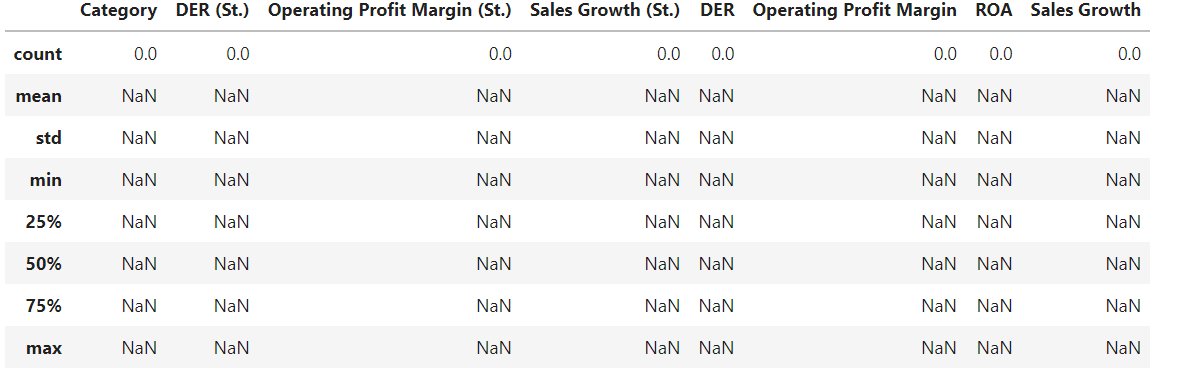
**Cluster Description**

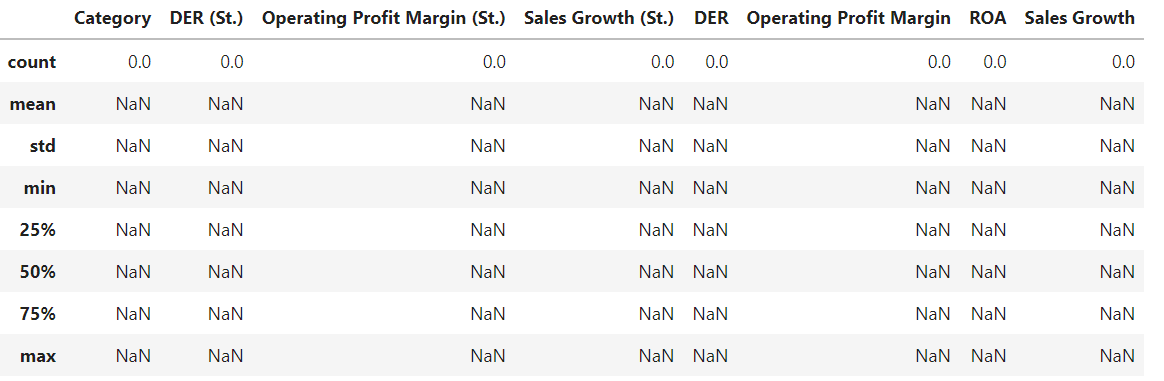




Graphical user interface, application

Description automatically generated





**KESIMPULAN**

Di masa pandemic ini industry dengan sector keuangan sangat terkena dampak nya, banyak beberapa Bank yang mengalami penuruan dalam keuangannya. Dan setelah dilakukan pengelompokan pada perusahaan industri keuangan atau bank melalui persebaran menggunakan K−Means clustering maka dapat di simpulkan sebagai berikut :

1. Setelah dilakukan pengelompokkan perusahaan keuangan melalui persebaran menggunakan K−Means clustering terbentuk menjadi 5 cluster
2. berdasarkan hasilnya terdapat 2 cluster yang optimal yaitu di dalam cluster category 4 dan 2.
3. Untuk cluster yang memiliki nilai yang bagus ada di dalam kategori 4
4. Untuk perusahaan yang memiliki DER dan Operating Profit Margin tinggi adalah BBNI.

**DAFTAR PUSTAKA**

[*https://www.republika.co.id/berita/qznqi9457/ojk-ungkap-kondisi-industri-keuangan-saat-pandemi*](https://www.republika.co.id/berita/qznqi9457/ojk-ungkap-kondisi-industri-keuangan-saat-pandemi)

**Penulisan *Journal Article***