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

**Febyola Anggraini Putri**1\*

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

E-mail: 1Febbyyollaa72@gmail.com\*

**Abstrak**

Data Mining merupakan sebuah proses ekstraksi informasi untuk menemukan pola (pattern recognition) yang penting pada tumpukan data dalam database sehingga menjadi pengetahuan (knowledge discovery). Pada jurnal ini akan membuat clustering pada data keuangan pada beberapa perusahaan di industry property dengan menggunakan metode CRISP-DM, dan dengan K-Means, scatter plot untuk penyebarannya, sehingga dapat menghasilkan beberapa cluster.

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

**PENDAHULUAN**

Data Mining sebagai salah satu cabang ilmu yang relatif baru mempunyai potensi pengembangan yang sangat besar dan diprediksi akan menjadi salah satu yang paling revolusioner pada dekade ini (Larose, 2005). Data Mining merupakan sebuah proses ekstraksi informasi untuk menemukan pola (pattern recognition) yang penting pada tumpukan data dalam database sehingga menjadi pengetahuan (knowledge discovery). Fungsi-fungsi dalam data mining antara lain: fungsi deskripsi, fungsi estimasi, fungsi Prediksi, fungsi Klasifikasi, fungsi Clustering dan fungsi asosiasi (Larose, 2006). Clustering digunakan untuk pengelompokan data secara alamiah berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap kluster lain.

Clustering adalah pengelompokan menggunakan teknik unsupervised learning dimana tidak diperlukan pelatihan pada metode tersebut atau dengan kata lain, tidak ada fase learning serta tidak menggunakan pelabelan pada setiap kelompok.

Proses data mining harus terukur, dapat dipercaya dan memenuhi suatu standar yang telah disepakati. CRISP-DM adalah standarisasi data mining yang disusun oleh tiga penggagas data mining market. Yaitu Daimler Chrysler (Daimler- Benz), SPSS (ISL), NCR. Kemudian dikembangkan pada berbagai workshops antara 1997-1999.

**Tujuan**

Tujuan dari penelitian ini adalah menggali pengetahuan (discovering knowledge) tentang pola (pattern) proporsi pelaksanaan tridharma oleh dosen sehingga diketahui tingkat produktivitas, komposisi serta keseimbangan dalam pelaksanaan tridharma.

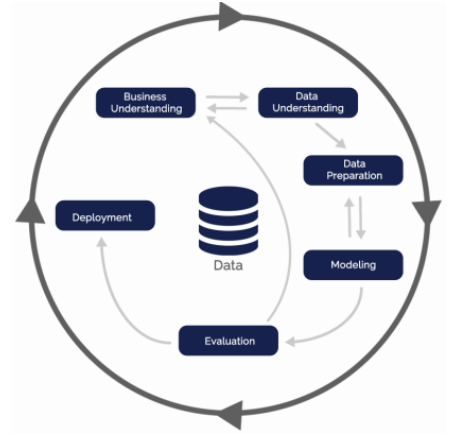
**METODOLOGI**

1. **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. **Clustering**

Clustering adalah studi formal untuk partisi atau pengelompokan data dengan tidak menggunakan pelabelan kategori. Clustering bersifat unsupervised learning atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi. Clustering digunakan untuk pengelompokan data secara alamiah berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap kluster lain [4].

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**

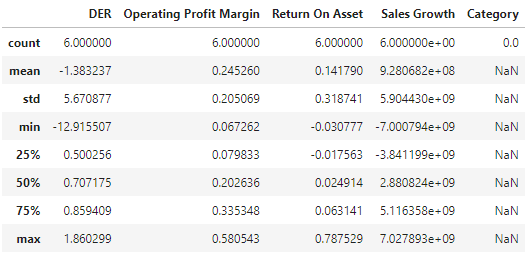
**Daftar Emitten**

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

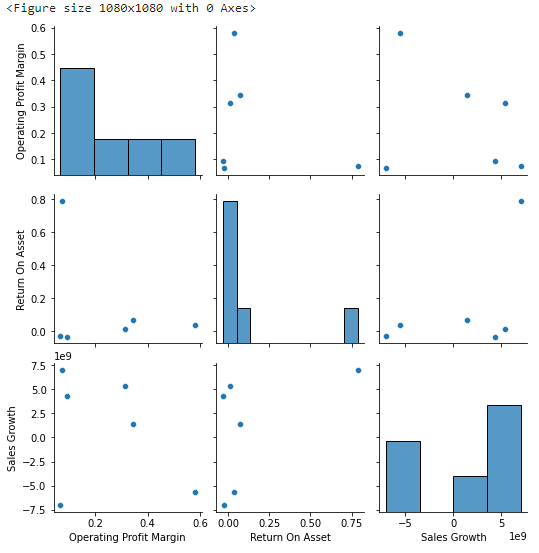
**Data financial Perusahaan dari daftar Emitten**

Data financial dari beberapa perusahaan diambil dari perhitungan data yang ada di yahoo finance. Dan dibawah ini adalah dataset yang dibuat dan disesuaikan dengan data yahoo finance beserta describe hasil load datanya.

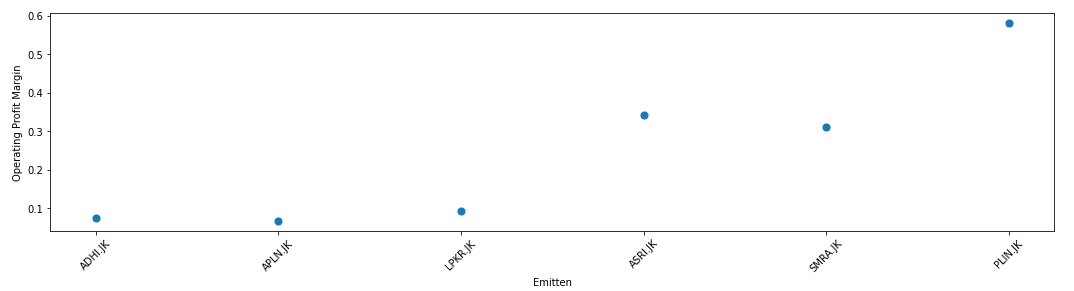




**Hasil scatter**

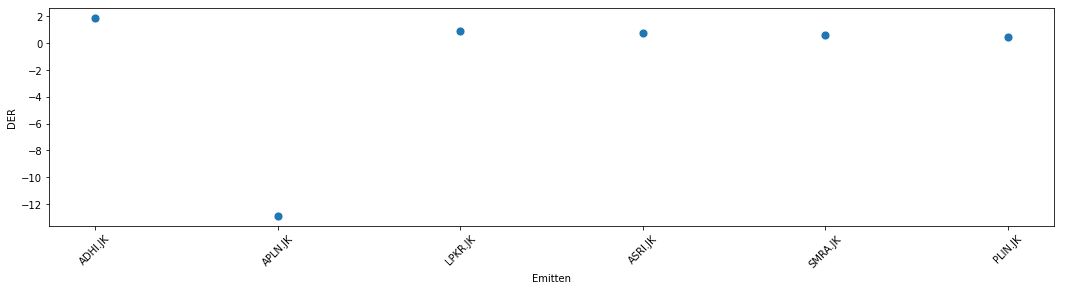


**Hasil Clustering Profit Margin Perusahaan**



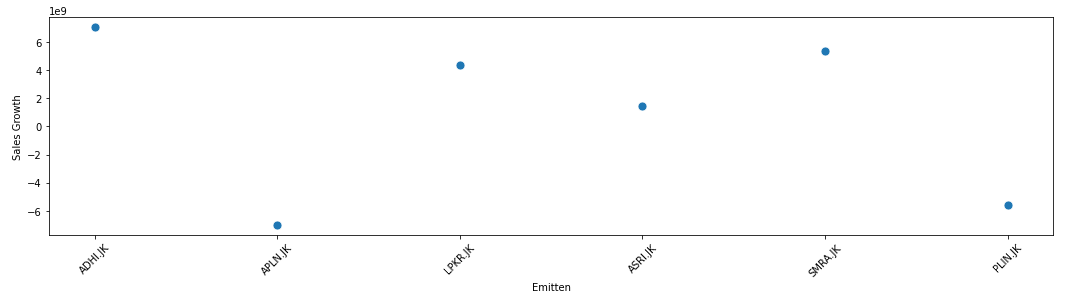
Pada hasil cluster Profit Margin di sektor perusahaan properti dapat dilihat pada ASRI.JK memiliki angka profit margin lebih tinggi dari perusahaan yang lainnya.

**Hasil Clustering DER Perusahaan**



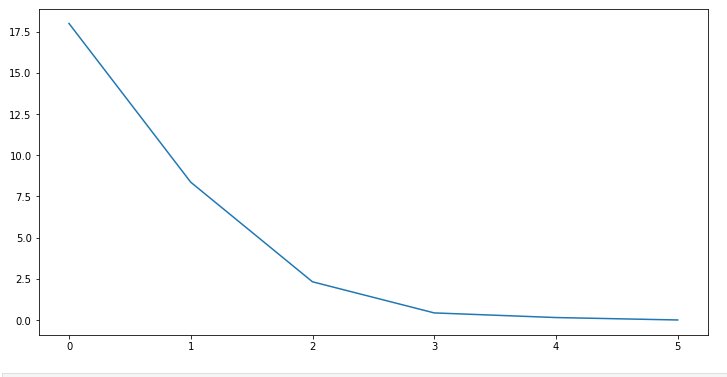
Berikut adalah hasil clustering pada DER perusahaan, dan dapat dilihat pada perusahaan ADHI.JK lebih tinggi dan paling rendah pada perusahaan APLN.JK

**Hasil Clustering Sales Growth**

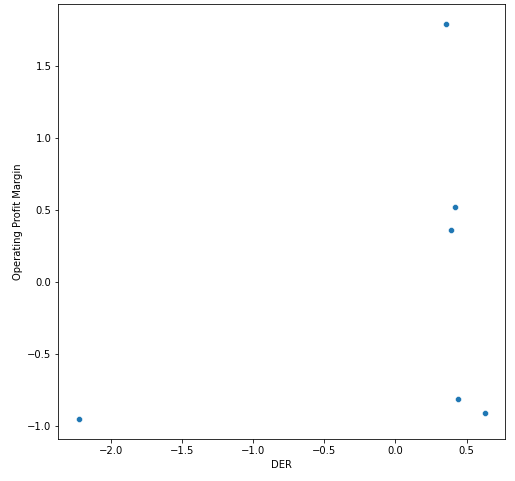


Dari hasil clustering untuk sales Growth yang tertinggi yaitu pada ADHI.JK

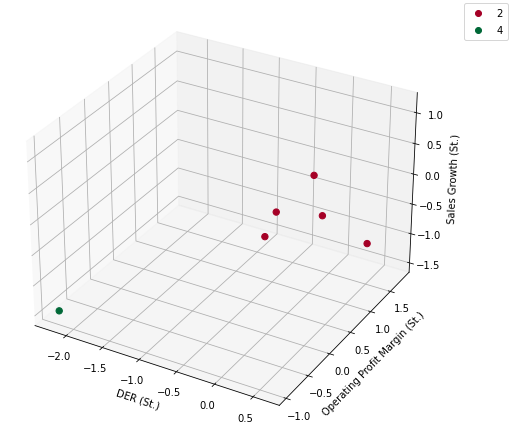
**K-Means**



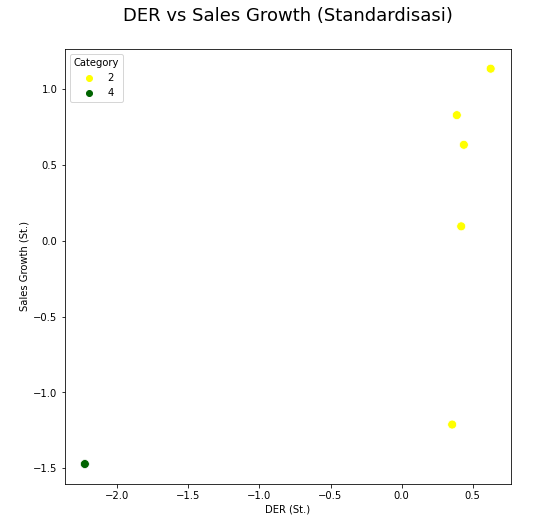
**Scatterplot**

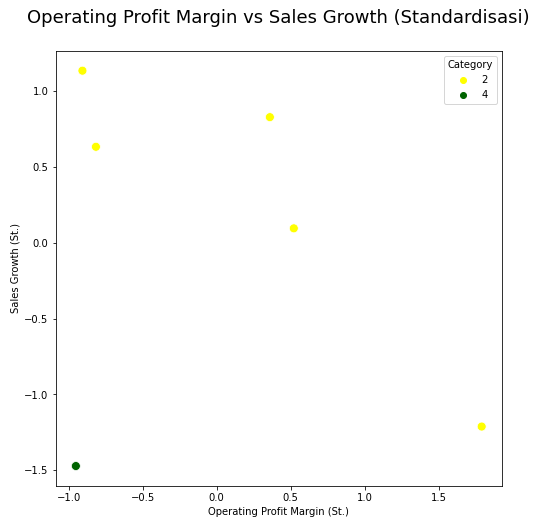


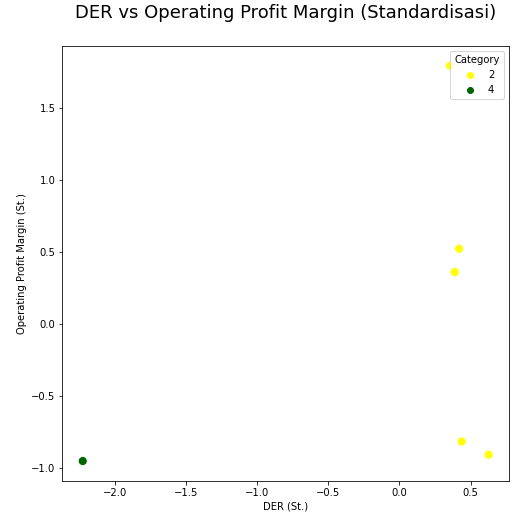
**3DER vs Operating Profit Margin scatterplot**



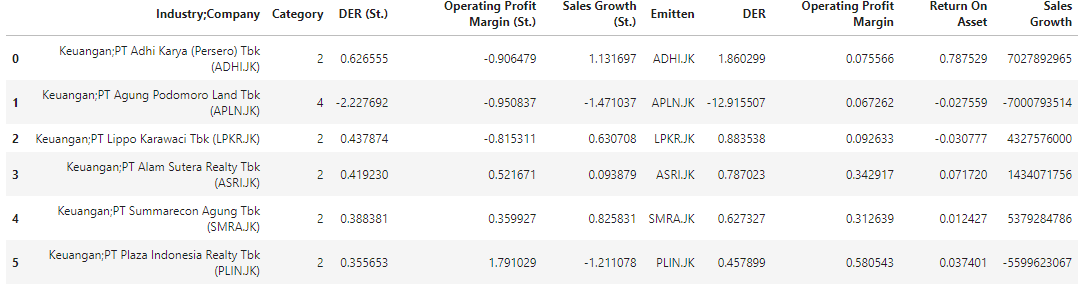
**X dan Y View**



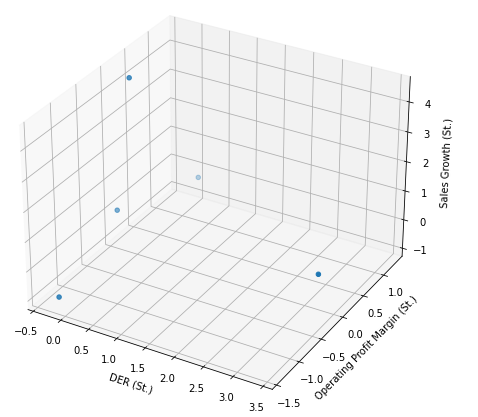




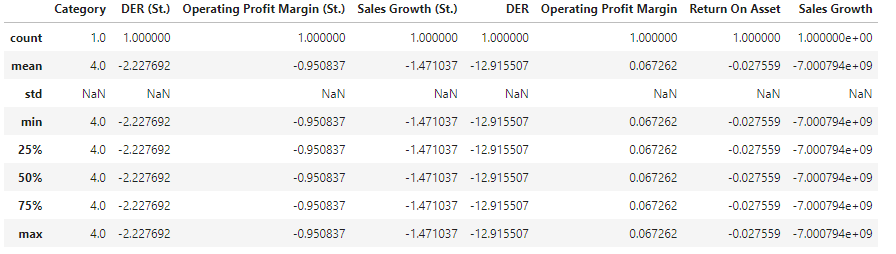
**Result**

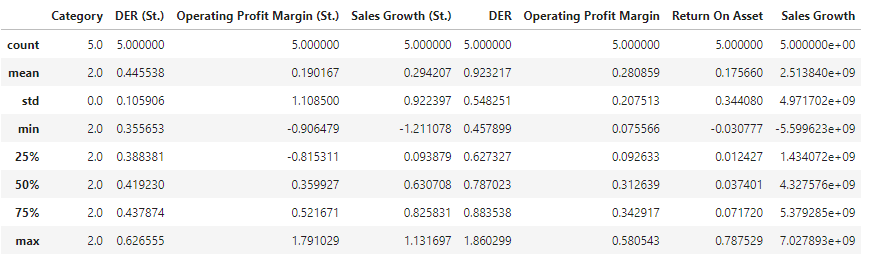


**Centroid**



**Cluster Description**





**KESIMPULAN**

Setelah dilakukan pengelompokan perusahaan industry Property 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**

Penulis mengucapkan terima kasih semoga dengan apa yang dibahas pada jurnal ini dapat membantu dalam menganalisa kenaikan dan juga penurunan yang terjadi pada perusahaan terutama pada bidang properti.