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

**Febyola Anggraini Putri**\*

1182002036

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

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

**Abstrak**

Indonesia sempat mengalami pelambatan ekonomi berturut – turut akibat pandemik Covid-19. Pelambatan tersebut berdampak negatif ke berbagai sektor di Indonesia, tak terkecuali sektor properti. Sebagian besar harga properti seperti rumah, apartemen, hingga kendaraan bermotor mengalami penurunan yang cukup tajam. Hal ini dikarenakan adanya penurunan permintaan terhadap properti seiring dengan meningkatnya kehati–hatian masyarakat dalam melakukan pengeluaran di tengah pandemi. Untuk mengetahui apakah perusahaan mengalami penurunan atau kenaikan diperlukan analisa dari data laporan keuangan tahunan, dan untuk mengelompokan perushaan yang mengalami kenaikan atau penurunan atau stabil diperlukan clustering data. Clustering digunakan untuk menganalisis pengelompokkan berbeda terhadap data, menyerupai klasifiaksi, namum pengelompokkan belum di definisikan sebelum dijalankannya tool Data Mining. Clustering menggunakan metode K-Means. Dan pada kasus ini penulis menggunakan data dari sektor properti yang diambil dari laporan keuangan tahunan dari situs BEI. Setelah diolah data nya dapat terbentuk 2 cluster, d) Untuk cluster yang memiliki nilai yang bagus ada di dalam kategori 2.

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

1. **PENDAHULUAN**
   1. **Latar Belakang**

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.

* 1. **Tujuan Penelitian**

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.

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

Diagram

Description automatically generated

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].

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. **K-means**

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

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.

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

Industri properti dan real estate adalah industri yang bergerak di bidang pengembangan jasa dengan memfasilitasi pembangunan kawasan kawasan yang terpadu dan dinamis. Produk yang dihasilkan industri ini dapat berupa perumahan, apartemen, ruko, gedung perkantoran, dan pusat perbelanjaan.

Properti dinilai masih menjadi primadona investasi karena memiliki nilai aset yang terus meningkat tanpa banyak mengalami fluktuasi dan risiko yang relatif cukup rendah.

Berdasarkan Market Behavior Survey yang disusun oleh Indonesia Property Watch (IPW) pada September 2020, sebesar 68,09 persen responden tertarik membeli properti pada masa pandemi Covid-19 dengan beberapa alasan utama, yaitu harga yang lebih murah, penawaran danpromosi menarik daripengembang, serta rencana angsuran yang fleksibel.

Selain kebutuhan primer masyarakat, properti merupakan produk investasi yang cukup menjanjikan, khususnya jenis properti real estate seperti rumah, tanah dan bangunan lainnya. Nilai aset yang terus meningkat, fluktuasi harga yang jarang terjadi, ditambah risiko yang relatif rendah menjadikan properti salah satu produk investasi yang didam- idamkan masyarakat.

* 1. **Data Understanding**

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

|  |  |
| --- | --- |
| Emitten | Company |
| AMAN | Makmur Berkah Amanda Tbk |
| APLN | Agung Podomoro Land Tbk |
| ASRI | Alam Sutera Realty Tbk |
| ATAP | Trimitra Prawara Goldland Tbk |
| BAPA | Bekasi Asri Pemula Tbk |
| BAPI | Bhakti Agung Propertindo Tbk |
| BCIP | Bumi Citra Permai Tbk |
| BEST | Bekasi Fajar Industrial Estate Tbk |
| BIKA | Binakarya Jaya Abadi Tbk |
| BKDP | Bukit Darmo Property Tbk |
| BKSL | Sentul City Tbk |
| BSDE | Bumi Serpong Damai Tbk |
| CITY | Natura City Developments Tbk |
| COWL | Cowell Development Tbk |
| CPRI | Capri Nusa Satu Properti Tbk |
| CTRA | Ciputra Development Tbk |
| DADA | Diamond Citra Propertindo Tbk |
| DART | Duta Anggada Realty Tbk |
| DMAS | Puradelta Lestari Tbk |
| DUTI | Duta Pertiwi Tbk |

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

|  |
| --- |
| Emitten,DER,Operating Profit Margin,ROA,Sales Growth,Category |
| AMAN,0.4,0.047536,0.01,-3.797435, |
| APLN,1.79,11.536165,-0.02,8.184296, |
| ASRI,1.35,7.757196,-0.01,6.69231, |
| ATAP,0.24,0.233544,0.05,5.698101, |
| BAPA,0.07,0.237532,-0.03,-3.317679, |
| BAPI,0.47,0.048601,-0.01,-0.096221, |
| BCIP,1.14,-0.025778,-0.01,-6.929469, |
| BEST,0.42,8.523578,-0.02,-8.44809916, |
| BIKA,-13.78,10.738797,0.01,6.9733069, |
| BKDP,0.73,1.860947,-0.05,-9.498563, |
| BKSL,0.61,2.104998,0.02,3.730487, |
| BSDE,0.69,12.893797,0.02,-7.213595, |
| CITY,0.12,0.256031,0.06,-7.213594691, |
| COWL,3.1,0.818589,-0.01,0.12, |
| CPRI,0.13,0.012406,-0.01,0.820829, |
| CTRA,1.16,15.0190759,0.05,4.08626, |
| DADA,1,0.052904,0.03,-18.2316939, |
| DART,1.57,14.0140375,-0.06,-16.572531, |
| DMAS,0.12,10.33586,0.13,10.221301, |
| DUTI,0.4,13.622598,0.04,-14.6164778, |

* 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, histogram

Description automatically generated

Table

Description automatically generated

**Hasil Clustering Operating Profit Margin Perusahaan**

Chart, scatter chart

Description automatically generated

Pada hasil cluster Oerating Profit Margin di sektor perusahaan properti ada 11 perusahaan dengan posisi OPM nya rendah dan sisa nya ada di posisi tengah dan juga atas, untuk nilai OPM tertinggi ada di perusahaan CTRA.

**Hasil Clustering Sales Growth**

Untuk hasil clustering dengan scatter plot dengan melihat sales growth nya, terdapat perusahaan yang nilai sales growth nya kecil yaitu seperti pada DADA, dan yang tertinggi ada pada DMAS

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 3.

Cluster berdasarkan Operating Profit Margin dan DER. Dapat dilihat bahwa terdapat clustering dengan DER tinggi namun Operating Profit Margin nya rendah, dan ada yang DER tinggi dan Operating Profit Margin nya tinggi juga, lalu ada yang DER rendah dan Operating Profit Margin berada di posisi tengah.

Chart, scatter chart

Description automatically generated

**3D scatterplot**

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated**X dan Y View**

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

**Result**

Table

Description automatically generated

Untuk hasil dari result nya terdapat 2 kategori yaitu kategori 2 dan 4. Untuk kategori 2 lebih bagus dikarenakan untuk di kategori 4 ada angka minus di nilai DER sedangkan di kategori 2 tidak ada.

**Centroid**

Chart

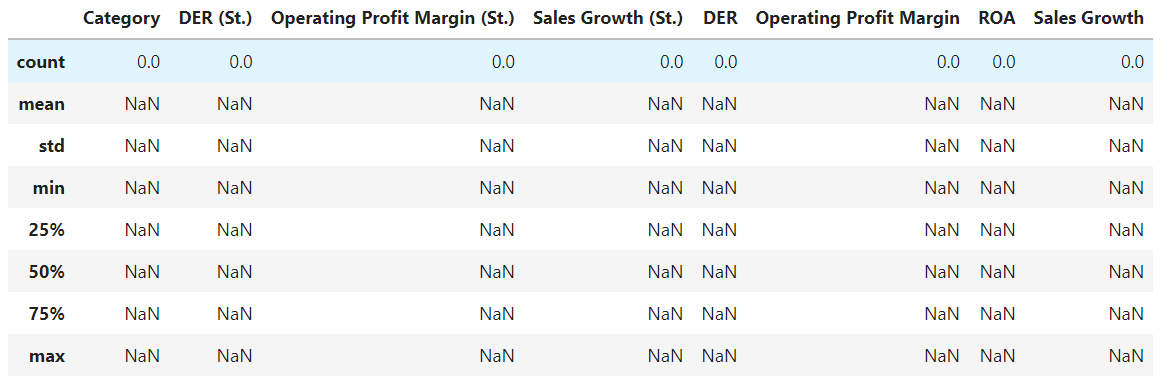
Description automatically generated

Untuk titik centroid nya berada di DER antara 0.0 sampai -0.5 dan sales growth nya sekitar dari -1 sampai 0.

**Cluster Description**

Graphical user interface

Description automatically generated



Graphical user interface

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

**KESIMPULAN**

Indonesia sempat mengalami pelambatan ekonomi berturut – turut pada kuartal II, III dan IV tahun 2020 akibat pandemik Covid-19. Pelambatan tersebut berdampak negatif ke berbagai sektor di Indonesia, tak terkecuali sektor properti. Sebagian besar harga properti seperti rumah, apartemen, hingga kendaraan bermotor mengalami penurunan yang cukup tajam. Hal ini dikarenakan adanya penurunan permintaan terhadap properti seiring dengan meningkatnya kehati–hatian masyarakat dalam melakukan pengeluaran di tengah pandemi. Namun adanya penurunan harga properti, tidak selalu berdampak buruk. Adanya penurunan harga properti dapat menguntungkan bagi sebagian masyarakat yang tetap membutuhkan properti sebagai kebutuhan primer. Terlebih banyaknya penawaran penurunan suku bunga dan kemudahan transaksi, cukup banyak menarik minat masyarakat untuk bertransaksi properti. Dan setelah dilakukan pengelompokan pada perusahaan industri properti untuk mencari tau perkembangan perusahaan nya apakah mengalami kenaikan atau penurunan melalui persebaran menggunakan K−Means clustering maka dapat di simpulkan sebagai berikut :

1. Setelah dilakukan pengelompokkan perusahaan properti melalui persebaran menggunakan K−Means clustering terbentuk menjadi 3 cluster
2. Cluster berdasarkan Operating Profit Margin dan DER. Dapat dilihat bahwa terdapat clustering dengan DER tinggi namun Operating Profit Margin nya rendah, dan ada yang DER tinggi dan Operating Profit Margin nya tinggi juga, lalu ada yang DER rendah dan Operating Profit Margin berada di posisi tengah
3. berdasarkan hasilnya terdapat 2 cluster yang optimal yaitu di dalam cluster category 4 dan 2.
4. Untuk cluster yang memiliki nilai yang bagus ada di dalam kategori 2
5. Untuk perusahaan yang masuk kategori 4 yaitu BIKA;Binakarya Jaya Abadi Tbk, yang mengalami penurunan

**DAFTAR PUSTAKA**

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

**Penulisan *Journal Article***