

ANALISIS SEGMENTASI PENGUNJUNG SUATU MALL DENGAN METODE KLASIFIKASI K-MEANS

Sagara Andi Ladzuardi Suwarsa

Program Studi Sains Data, Fakultas Sains dan Teknologi,
Universitas Koperasi Indonesia, Jatinangor, Indonesia
Email : asagara830@gmail.com

ABSTRAK

Diera global seperti sekarang ini, setiap pelaku bisnis disektor apapun harus mampu mengedepankan kebutuhan pasar dengan baik. Oleh sebab itu, mereka harus mampu melakukan analisis terhadap data yang dikumpulkan dari internal maupun external untuk dapat memberikan solusi terbaik pada setiap permasalahan yang ditemukan. Lebih daripada itu, mereka juga harus mampu melakukan peramalan berdasarkan data, dengan mensimulasikan semua kemungkinan yang ada tanpa harus menunggu ketidakpastian terjadi. Setidaknya mereka dapat meramalkan persentasi kemungkinan yang akan terjadi untuk dapat mengoptimalkan keuntungan dan sekaligus meminimalkan resiko.

Dalam penelitian ini, penulis akan melakukan penerapan salah satu aplikasi yang paling populer dalam konsep Mechine Learning yakni analisis segmentasi pengunjung suatu Mall dengan metode klasifikasi K-means. Kita semua tahu, untuk melakukan analisa dan menemukan segmentasi pelanggan terbaik diperlukan metodologi terbaik. Sehingga penelitian ini, melakukan analisis deskriptif terhadap data pengunjung suatu Mall dan kemudian mengimplementasikan beberapa model algoritma K-means. Setelah menemukan klasifikasi terbaik, maka direkomendasikan kepada para pelaku bisnis dalam Mall tersebut agar mereka meninjau ulang pruduk-pruduk yang seharusnya mereka sediakan untuk mengoptimalkan keuntungan dan meminimalkan risikonya. Tentu saja penelitian ini juga masih bisa dikembangkan lagi dengan menambahkan variabel-variabel lain yang saling mempengaruhi. Dapat juga dijadikan sebagai referensi untuk penelitian pada dataset yang berbeda dengan tujuan melakukan klasifikasi (klustering).

Kata kunci: Analisis Segmentasi, Statistik Deskriptif, Klasifikasi, Machine Learning, dan K-Mean Klastering.

1. PENDAHULUAN

1.1. Latar Belakang

Segmentasi pelanggan adalah aplikasi penting dalam dunia bisnis, terutama untuk analisis pemasaran. Teknik pengelompokan seperti K-means memungkinkan perusahaan mengidentifikasi segmen pelanggan yang berbeda untuk menargetkan basis konsumen potensial. Dalam penelitian ini, kita akan menggunakan K-means untuk mengelompokkan dataset yang tidak berlabel (unsupervised).

Segmentasi pelanggan adalah proses membagi basis pelanggan menjadi kelompok-kelompok individu yang memiliki kesamaan dalam berbagai aspek yang relevan dengan pemasaran, seperti jenis kelamin, usia, minat, dan kebiasaan pengeluaran. Perusahaan menganggap bahwa setiap pelanggan memiliki kebutuhan yang berbeda dan memerlukan pendekatan pemasaran yang khusus. Dengan segmentasi, perusahaan dapat memahami preferensi pelanggan lebih dalam, memenuhi kebutuhan spesifik mereka, dan memaksimalkan keuntungan.

Menurut Kotler (dalam Precylia Cyndi Kembuan, 2012), segmentasi penting karena memungkinkan perusahaan fokus dalam mengalokasikan sumber daya, menentukan posisi pasar, dan menyusun strategi pemasaran yang efisien. Segmentasi juga membantu perusahaan mengalahkan pesaing dengan melihat pasar dari sudut pandang yang unik.

Sanusi dan Anuar (2015) menunjukkan bahwa variabel yang paling efektif dalam membentuk positioning secara berurutan adalah proses, pelayanan, tempat, harga, dan promosi. Teknik segmentasi bergantung pada pembeda utama seperti demografi, geografi, status ekonomi, dan pola perilaku.

1.2 Rumusan Masalah

Adapun yang menjadi permasalahan dalam penelitian ini adalah bagaimana untuk menemukan menemukan konsumen yang lebih potensial untuk dilakukan pendampingan yang lebih tepat untuk meningkatkan

keuntungan yang akan diperoleh pelaku bisnis dalam Mall. Masalah lainnya adalah bagaimana untuk menemukan model yang tepat untuk pembagian segmentasi konsumen berdasarkan data yang Ada.

1.3 Tujuan Penelitian

tujuan penelitian ini adalah:

- Untuk melakukan analisis statistik deskriptif terhadap data pengunjung.
- Menemukan karakteristik pengunjung berdasarkan penghasilan, usia dan pengeluaran (biaya belanja yang biasa dilakukan) oleh pengunjung.
- Untuk menemukan berapa banyak klasifikasi segmentasi yang tepat berdasarkan data pengunjung Mall tersebut.

1.4. Manfaat Penelitian

Melalui tulisan ini dapat diambil beberapa manfaat antara lain:

- Bagi Peneliti: berguna untuk menjelaskan teori-teori statistika yang dapat diterapkan dalam penelitian dan dalam kehidupan sehari-hari. Sebagai contoh; Penerapan analisis Statistik Deskriptif dan Metode K-Means Klastering (Machine Learning) dalam menentukan Segmentasi Pelanggan.
- Bagi Pelaku Bisnis: rekomendasi untuk memahami segmentasi pelanggan mereka, sehingga dapat menentukan target pasar dan produk-produk yang seharusnya mereka sediakan

2. TINJAUAN PUSTAKA

2.1. Peneliti Terdahulu

Gunawan Adi Chandra Makens (2016) menyatakan bahwa segmentasi demografis membagi pasar berdasarkan usia, jenis kelamin, pendapatan, pekerjaan, pendidikan, agama, ras, dan kebangsaan. Variabel demografis paling sering digunakan karena kebutuhan, keinginan, dan tingkat penggunaan konsumen berhubungan erat dengan variabel ini. Kotler (2007) menambahkan bahwa segmentasi demografis dapat dibagi berdasarkan usia, ukuran keluarga, tahap kehidupan, jenis kelamin, pendapatan, generasi, dan kelas sosial. Segmentasi ini penting karena mempengaruhi preferensi konsumen terhadap produk.

Sanusi dan Anuar (2015) menunjukkan bahwa variabel proses, pelayanan, tempat, harga, dan promosi efektif dalam membentuk positioning. Segmentasi pelanggan dapat didasarkan pada demografi, geografi, status ekonomi, dan perilaku. Penelitian ini fokus pada variabel jenis kelamin, usia, penghasilan, dan pengeluaran.

2.2. Statistik Deskriptif

Statistik deskriptif adalah cara mengumpulkan, mengolah, menganalisis, dan menafsirkan data. Hasan (2001) menjelaskan bahwa statistik deskriptif mempelajari cara pengumpulan dan penyajian data sehingga mudah dipahami. Statistik ini mencakup distribusi frekuensi, ukuran nilai pusat, ukuran dispersi, kemencengan, keruncingan kurva, angka indeks, deret waktu, korelasi, dan regresi sederhana.

2.3. Machine Learning

Salah satu metode yang paling populer dalam Machine Learning adalah Algoritma K-means, berikut akan dilakukan review mengenai teori dan konsepnya.

K-means Algorithm

K-means clustering adalah metode clustering non-hierarkis untuk mengklasifikasikan data dalam satu bentuk atau lebih cluster. Dalam hal ini, data memiliki karakteristik yang sama diklasifikasikan dalam satu cluster, dan data memiliki karakteristik yang berbeda diklasifikasikan dengan cluster lain. Karena itu, jika data terdaftar dalam satu cluster berarti memiliki tingkat variansi yang kecil. Metode Klastering K-means mendeklarasikan nilai awal cluster, di mana pusat cluster adalah yang pertama dipilih secara acak, kemudian menghitung jarak setiap cluster ke pusat data menggunakan rumus Euclidean dalam. Menurut jurnal Nanda, Dkk (2010), rumus untuk Metode Klastering K-means adalah:

$$D_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad \dots(1)$$

Dimana,

- D_{ij} : Jarak data dari i ke pusat cluster j .
- X_{ki} : data i pada atribut k .
- X_{kj} : titik pusat ke j di atribut ke k .

3. METODE PENELITIAN

3.1. Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif dimana data yang dinyatakan dalam angka dan dianalisis dengan teknik statistik. Metode yang digunakan adalah metode asosiatif yaitu bentuk penelitian dengan menggunakan minimal dua variabel yang dihubungkan. Sejatinya, Metode asosiatif merupakan suatu penelitian yang mencari hubungan sebab akibat antara satu variabel independen (variabel bebas) yaitu Gender (X_1), Usia (X_2), Pendapatan_Tahunan_Ribuan_USD (X_3), dan Pengeluaran_USD (X_4) dengan variabel dependen (variabel terikat) yaitu keputusan pembelian (Y). Namun karena keterbatasan data, maka dibagian awal penelitian melakukan analisis statistik deskriptif untuk melihat sebaran data, frekuensi, densitas, dan korelasi antar-variabel. Kemudian menerapkan metode klasifikasi K-mean (Unsupervised Machine Learning) untuk melihat segmentasi pengunjung dalam Mall tersebut.

3.2. Metode Pengumpulan Data

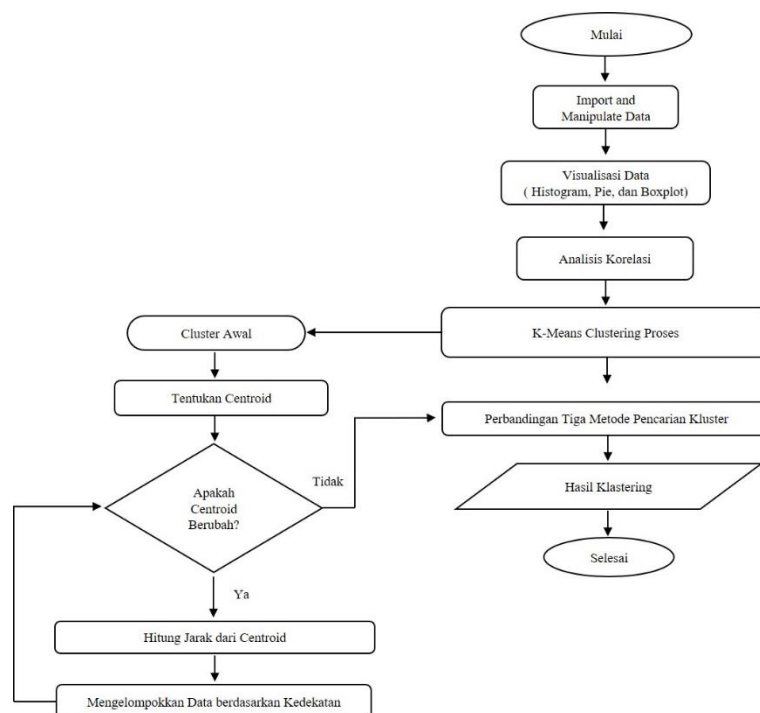
Penelitian ini menggunakan data sekunder yang merupakan tipikal penelitian kuantitatif, dataset yang digunakan sudah tersedia di Kaggle.com.

3.3. Objek Penelitian

Adapun yang menjadi objek dalam penelitian kuantitatif ini dibagi berdasarkan metode atau jenis penelitian, termasuk yang berikut ini:

- Objek Penelitian Deskriptif: melakukan analisis deskriptif terhadap data pengunjung Mall.
- Objek Penelitian Korelasi: dilakukan juga analisis korelasi untuk melihat keterkaitan antar variabel.
- Objek Penelitian Komparatif: melakukan perbandingan jenis kelamin, usia, Penghasilan dan pengeluaran. Serta melakukan perbandingan beberapa metode dalam menentukan kluster yang optimal untuk diterapkan dalam Klasifikasi K-means.

3.4. Flowchart



Gambar 1. Flowchart Penelitian

4. HASIL PENELITIAN DAN PEMBAHASAN

Pada langkah pertama dari proyek ini, akan dilakukan eksplorasi data. Mungkin saja anda akan mengimpor beberapa packages yang diperlukan dan kemudian membaca data. Berikut ini dilakukan import data, melihat struktur data, dan mengulas sepias mengenai ringkasan data statistiknya. Hal ini dilakukan untuk mendapatkan pengetahuan dasar mengenai data dan mungkin diperlukan dalam menemukan atau menentukan asumsi awal.

ID_Pelanggan	Gender	Usia	Pendapatan_Tahunan_Ribuan_USD
Min. : 1.00	Pria : 88	Min. :18.00	Min. : 15.00
1st Qu.: 50.75	Wanita:112	1st Qu.:28.75	1st Qu.: 41.50
Median :100.50		Median :36.00	Median : 61.50
Mean :100.50		Mean :38.85	Mean : 60.56
3rd Qu.:150.25		3rd Qu.:49.00	3rd Qu.: 78.00
Max. :200.00		Max. :70.00	Max. :137.00

Pengeluaran_USD

Min. : 1.00
1st Qu.:34.75
Median :50.00
Mean :50.20
3rd Qu.:73.00
Max. :99.00

13.9690073315589

26.2647211652712

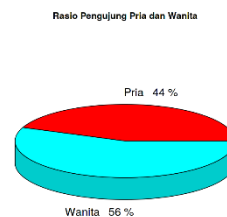
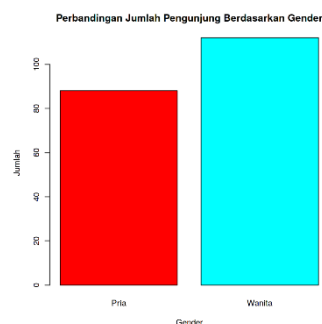
25.8235216683702

Dari hasil yang diperlihatkan diatas, disimpulkan bahwa struktur data gender (jenis kelamin) adalah faktor (kualitatif) sehingga untuk menganalisa variabel ini lebih tepat dengan melihat frekuensi atau persentasi antar faktor (dalam hal ini pria dan wanita) yang akan diperlihatkan degan histogaram atau pie chart setelah bagian ini. Berdasarkan summary statistiknya, diperoleh bahwa pengunjung wanita lebih banyak dibanding pria. Mean (Rata-rata) usia, pendapatan, dan pengeluaran secara berturut-turut adalah 38.85, 60.56, dan 50.20. Olehkarena itu, ditarik kesimpulan bahwa usia, pendapatan, dan pengeluaran pengunjung sangat beragam karena rentan antara nilai min dan max cukup besar. Hal ini, juga diperkuat besarnya nilai standar diviasi untuk setiap variabel. Standar diviasi atau simpangan baku biasa digunakan untuk mengukur sebaran variansi data. Semakin besar nilai standar diviasinya makan semakin beragam (semakin luas variansi sebaran datanya).

4.1.Visualisasi Data

Jenis Kelamin Pelanggan

Sejalan dengan kesimpulan sebelumnya, visualisasi plot histogram dan pie chart dibawah ini juga memperlihatkan perbandingan jumlah pengunjung wanita lebih tinggi dibanding pria (Itu sudah hal yang lumrah, tetapi itu penting untuk dipertimbangkan dalam analisis).



Usia Pelanggan

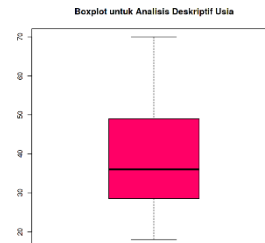
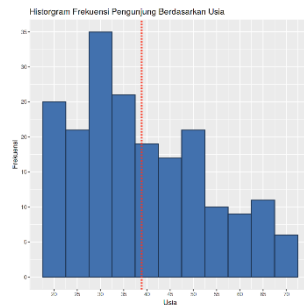
Pada bagian ini digunakan plot histogram untuk melihat distribusi sebaran frekuensi usia pelanggan. Pertama-tama dilakukan analisis, mean (rata-rata), Median (Nilai Tengan), Modus (Data yang sering Muncul) untuk melihat kecondongan data.

38.85

36

32

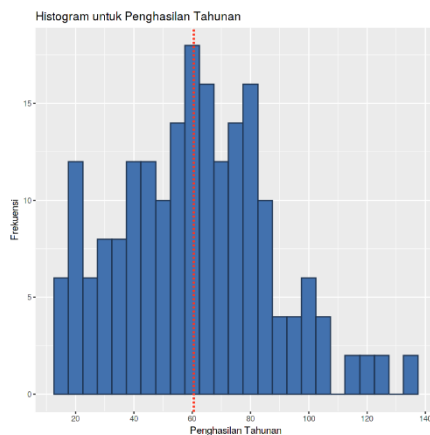
Menurut jurnal (Hippel,2005), jika mean berada di kanan median, dan median di kanan mode maka data berdistribusi Kemiringan condong ke kanan (Positively Skewed Distribution). Artinya, Usia pengunjung di Mall tersebut lebih didominasi usia muda sekitar umur 32 meskipun rata-rata pengunjungnya berusia 38.85 tahun. Jadi, boleh direkomendasikan juga produk-produk yang dijual di Mall tersebut harus lebih memperhatikan kebutuhan pengunjung umur 32 tahun (perlu penelitian lanjut). Kemudian, dilanjutkan dengan histogram dan Boxplot dibawah ini:
Dari dua visualisasi di atas, disimpulkan bahwa usia pelanggan maksimum adalah antara 30 dan 35. Usia



minimum pelanggan adalah 18, sedangkan, usia maksimum adalah 70.

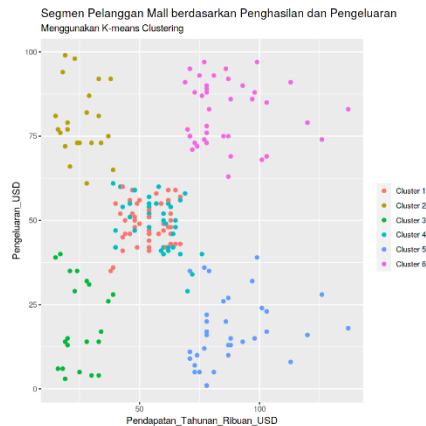
Pendapatan Tahunan Pelanggan

Pada bagian ini, dilakukan analisis pendapatan tahunan pelanggan dengan visualisasi distribusi densitasnya. Pertama-tama digunakan histogram dan kemudian melanjutkannya menggunakan plot densitas.



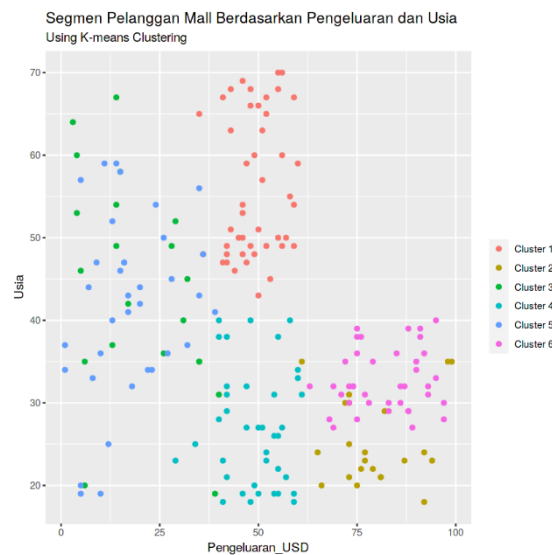
Dari analisis deskriptif berdasarkan kedua grafik diatas, disimpulkan bahwa pendapatan tahunan minimum pelanggan adalah 15 dan pendapatan maksimum adalah 137. Orang-orang yang mendapatkan penghasilan rata-rata 70 memiliki jumlah frekuensi tertinggi dalam distribusi histogram. Gaji rata-rata semua pelanggan adalah 60,56. Melalui Plot Kepadatan (Densitas) terlihat bahwa pendapatan tahunan memiliki distribusi normal.

Visualisasi Hasil Klustering



Dari visualisasi di atas, Segmen Pelanggan Mall berdasarkan Pendapatan dan Pengeluaran diperoleh kesimpulan berikut:

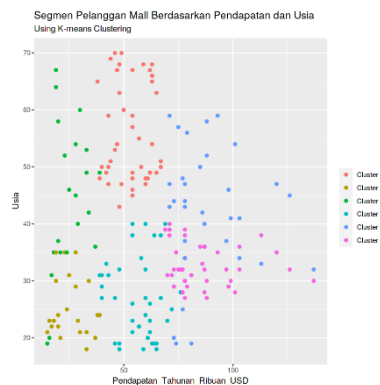
- Cluster 1 dan 4: Cluster ini mewakili pengunjung dengan gaji menengah dan juga gaji tahunan menengah.
- Cluster 2 : Cluster ini menunjukkan pengunjung dengan pendapatan tahunan yang tinggi dan pengeluaran tahunan yang rendah.
- Cluster 3 : Cluster ini menunjukkan pengunjung dengan pendapatan tahunan rendah serta pengeluaran tahunan rendah.
- Cluster 5 : Cluster ini mewakili pengunjung dengan pendapatan tahunan yang rendah tetapi pengeluarannya yang tinggi.
- Cluster 6 : Cluster ini mewakili pengunjung yang memiliki pendapatan tahunan tinggi dan pengeluaran tahunan tinggi.



Dari visualisasi di atas, Segmen Pelanggan Mall Pengeluaran dan Usia diperoleh kesimpulan berikut:

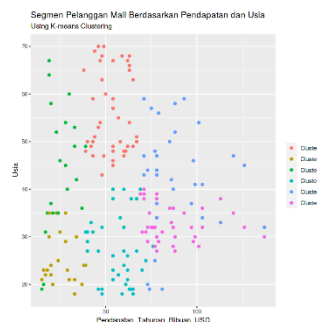
- Cluster 1 dan 6: mewakili pengunjung usia muda (25-32) dengan pengeluaran tinggi.
- Cluster 2 dan 5: menunjukkan pengunjung usia paling dominan (30-60) dengan pengeluaran yang rendah.
- Cluster 3 : menunjukkan pengunjung usia tua (45 keatas) dengan pengeluaran menengah (USD 50-an).

- Cluster 4 : menunjukkan pengunjung usia muda (35 kebawah) dengan pengeluaran menengah (USD 50-an).

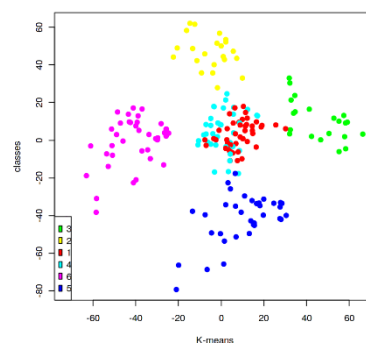


Dari visualisasi di atas, Segmen Pelanggan Mall Pendapatan dan Usia diperoleh kesimpulan berikut:

- Cluster 1 dan 2: mewakili pengunjung pendapatan rendah dibawah USD 50 pertahun Usia muda dan tua.
- Cluster 3 dan 4: menunjukkan pengunjung pendapatan menengah dan usia lebih besar dari 45 dan lebih kecil 40.
- Cluster 5 dan 6: menunjukkan pengunjung pendapatan tinggi dan usia lebih besar dari 45 dan lebih kecil 40.



Untuk menarik kesimpulan berdasarkan ketiga parameter diatas, dapat digunakan metode Principal Component Analysis (PCA) sebagai berikut:



- Cluster 3 dan 4: menyatakan pelanggan berpendapatan dan pengeluaran menengah dengan umur berpariatif
- Cluster 1: menyatakan pelanggan berpengeluaran tinggi, pendapatan rendah dan umur kisaran 25 kebawah (muda).

- Cluster 6: menyatakan pelanggan berpengeluaran dan pendapatan tinggi dengan umur kisaran 32 (sedang).
- Cluster 5: menyatakan pelanggan berpengeluaran rendah, umur tua dan penghasilan tinggi.
- Cluster 2: menyatakan pelanggan berpengeluaran dan pendapatan stabil tetapi umur tua.

5. KESIMPULAN

Dengan bantuan klasifikasi metode klustering K-means, peneliti dapat memahami variabel jauh lebih baik, mudah untuk mengambil keputusan dengan baik dan benar. Dengan identifikasi pelanggan, perusahaan dapat merilis produk dan layanan yang menargetkan pelanggan berdasarkan beberapa parameter seperti pendapatan, usia, pola pengeluaran, dll. Selain itu, penambahan variabel baru yang lebih kompleks seperti mempertimbangkan ulasan produk dapat memberikan gambaran segmentasi yang lebih baik.

Peneliti berikutnya dapat mengembangkan dan menggunakan konsep machine learning ini terhadap data unsupervised lainnya (yakni data yang tidak berlabel). Secara khusus, dapat menggunakan algoritma pengelompokan yang disebut K-means clustering. Dalam menganalisis dan memvisualisasikan data dan kemudian melanjutkan untuk mengimplementasikan algoritma ini.

6. DAFTAR PUSTAKA

- Hasan, M. I. (1999). *Pokok-Pokok Materi 1 Statistika Deskriptif*. Retrieved from <https://opac.perpusnas.go.id/DetailOpac.aspx?id=68018>
- Kotler, P., Armstrong, G., Nurmawan, I., & Sumiharti, Y. (1997). *Prinsip-Prinsip Pemasaran*. *Online Public Access Catalog, III*(1), 1-440. Retrieved from <https://www.scribd.com/document/436592785/Prinsip-Prinsip-Pemasaran-Philip-Kotler-Gary-Armstrong-Jilid-1-Edisi-12-1-pdf>
- Siregar, B. (2020). *Data-Pengunjung-Mall*. Retrieved from <https://www.kaggle.com/datasets/baktisiregar/datapengunjungmall>
- Suryoatmono, B. (n.d.). *Kursus Statistika Dasar*. Retrieved from <https://www.scribd.com/document/29477374/Kursus-Statistika-Dasar>