

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Seiring perkembangan zaman, teknologi semakin berkembang dan maju. Selain lebih praktis dan mudah, teknologi yang semakin maju ini lebih menguntungkan dan mudah dijangkau terutama dibidang penjualan. Platform yang dipakai untuk penjualan ada berbagai macam, bisa berbentuk aplikasi atau *website*. Pencatatan penjualanpun kini sudah menggunakan teknologi dan sudah sangat jarang menggunakan pencatatan manual.

Persaingan di dunia teknologi semakin meningkat, salah satunya adalah *smartphone*. Berbagai jenis, spesifikasi, dan harga yang bervariasi jika dilakukan pencatatan secara manual akan sangat menguras waktu mengingat penjualan *smartphone* di seluruh Indonesia ini sangat banyak. Dalam pemasaran juga perlu adanya strategi untuk memperbanyak stok barang *smartphone* agar tidak kehilangan pasar konsumen. Dengan demikian, maka perlu dilakukannya pengelompokan produk yang penjualan paling tinggi, penjualan sedang dan penjualan rendah agar mempermudah dalam penentuan stok barang.

Pengelompokan data penjualan *smartphone* dapat dilakukan dengan teknik *data mining*. Teknik *data mining* yaitu mengelompokkan data berdasarkan suatu karakteristik data tersebut. Data dalam hal ini yaitu penjualan yang akan dikelompokkan menjadi produk yang tergolong penjualan paling tinggi, penjualan sedang, dan penjualan rendah. Salah satu kajian *data mining* yaitu *clustering*. Algoritma yang digunakan untuk proses *clustering* yaitu *K-Means*. *K-Means* merupakan salah satu metode data *clustering non hirarki* yang membagi data ke dalam beberapa *cluster* / kelompok (Sukanto et al. 2018).

Pada penelitian yang dilakukan oleh Syam (2017) dengan hasil penelitian yaitu pengelompokan mahasiswa berdasarkan kelompok

mahasiswa berprestasi, kelompok mahasiswa berpotensi berprestasi, dan kelompok mahasiswa berpotensi bermasalah. Penelitian dilakukan uji menggunakan *software Rapid Miner*, yang hasilnya sama dengan perhitungan analisa menggunakan *K-Means* yang dilakukan.

Adapula penelitian yang dilakukan oleh Ulil et al. (2021) yang melakukan penelitian menggunakan metode *k-means clustering*, dengan menggunakan beberapa contoh dari berbagai merk *smartphone*. Sehingga akan terbentuk beberapa *cluster*/kelompok produk *smartphone* yang diharapkan bisa membantu untuk pemilihan produk *smartphone* yang akan dibeli.

Penelitian yang dilakukan oleh Harani et al. (2020) menganalisa karakteristik pelanggan sebagai dasar penetapan segmentasi pelanggan dan customer profiling pelanggan produk *digital service add on Indihome* menggunakan Algoritma *K-Means*. Penelitian ini menggunakan 3 evaluasi *cluster*, yaitu *Davies Bouldin Index (DBI)*, *Silhouette index*, dan *Calinski Harabasz Index*. Pengolahan data pelanggan dibagi menjadi 3 simulasi dengan persentase data train dan data test 80% - 20%, 70% - 30%, 50% - 50%. Hasil dari penelitian menunjukkan dari ketiga simulasi didapatkan simulasi terbaik yaitu persentase data train dan data test 50% - 50%.

Penelitian ini membahas penggunaan algoritma *k-means* untuk melakukan *clustering* data penjualan *smartphone* dengan menggunakan 3 simulasi dengan persentase data *train* dan data *test* 80% - 20%, 70% - 30%, 50% - 50%, kemudian dilakukan proses evaluasi dengan 3 evaluasi yaitu *Silhouette Index*, *Davies Bouldin Index (DBI)*, dan *Calinski Harabasz Index* yang mengikuti dari *paper* Harani et al. (2020). Untuk menentukan evaluasi optimal dapat dilihat dari hasil evaluasi tersebut, dimana score dari *Davies Bouldin Index* nya mendekati 0, score *Silhouette Index* nya semakin mendekati 1, dan score *Calinski Harabasz* nya semakin besar.

## **B. Perumusan Masalah**

Berdasarkan latar belakang diatas, perumusan masalah pada penelitian ini adalah bagaimana menerapkan *data mining clustering* untuk

mengelompokkan data penjualan *smartphone* berdasarkan simulasi dengan algoritma *K-Means* dan dievaluasi menggunakan *Davies Bouldin Index*, *Silhouette Index*, dan *Calinski Harabasz Index*.

### C. Batasan Masalah

Batasan masalah dari penelitian ini yaitu:

1. Data yang digunakan adalah data penjualan produk *smartphone* yang diambil dari *website kaggle*.
2. Penelitian hanya menggunakan metode evaluasi *Davies Bouldin Index (DBI)*, *Silhouette Index (SI)*, dan *Calinski Harabasz (CH)*.

### D. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini diharapkan menghasilkan cluster yang berdampak pada kebutuhan *stock* data yang akan menjadi acuan perusahaan untuk mengetahui data yang tergolong penjualan tinggi, penjualan sedang, penjualan rendah berdasarkan simulasi terbaik dari perbandingan presentase data dari hasil optimal perhitungan evaluasi yang sudah ditentukan.

### E. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mengetahui penggunaan teknik *split data* dengan membagi ke dalam 3 simulasi dengan persentase data *train* dan data *test* 80% - 20%, 70% - 30%, 50% - 50%.
2. Mengetahui penggunaan evaluasi *Davies Bouldin Index (DBI)*, *Silhouette Index*, dan *Calinski Harabasz Index* pada algoritma *k-means clustering*.
3. Hasil penelitian ini dapat digunakan sebagai bahan perbandingan dan pengkajian pada pemilihan metode perhitungan evaluasi bagi para peneliti.