

BAB II

TINJAUAN PUSTAKA

Penelitian *clustering* dengan metode *K-Means* pernah diteliti oleh Zulrahmadi (2020) yang berjudul “Pemetaan Wilayah Potensial Terhadap Penjualan Sepeda Motor Honda Menggunakan *K-Means Clustering*”. Wilayah yang dijadikan penelitian terdiri dari 197 desa, 39 kelurahan, dan 20 kecamatan. Penelitian bertujuan untuk mengetahui wilayah potensial untuk penjualan motor sehingga bisa digunakan untuk pengambilan keputusan bagi PT. Capella Dinamik Nusantara. Penelitian menghasilkan tiga cluster yaitu *Cluster 0* (C0) yang dideskripsikan sebagai kurang berpotensi sebanyak lima wilayah, *Cluster 1* (C1) yang dideskripsikan cukup berpotensi sebanyak tiga wilayah, *Cluster 2* (C2) yang dideskripsikan sangat berpotensi sebanyak dua wilayah untuk penjualan motor. Iterasi terjadi sebanyak dua kali. Hasil *cluster* tersebut diharapkan bisa digunakan untuk mengambil keputusan terbaik sehingga keuntungan perusahaan bisa meningkat dan promosi bisa tepat sasaran. Data yang digunakan adalah data penjualan motor selama tiga tahun (2017, 2018, 2019) [6].

Terdapat penelitian lain menggunakan *clustering* dengan metode *K-Means* dengan studi kasus pada sekolah yang pernah diteliti oleh Tandy dan Assegaff (2019) dalam jurnal yang berjudul: “Analisis dan Perancangan *Clustering* Siswa Baru Menggunakan Metode *K-Means* pada SMK Negeri 1 Kota Jambi”. Kegiatan penerimaan siswa baru sekolah ini dilaksanakan setiap 1 tahun sekali dan dalam prosesnya memerlukan aplikasi yang efisien yaitu menggunakan aplikasi yang merubah sistem manual menjadi terkomputerisasi dan melakukan pembagian kelompok siswa baru berdasarkan jurusan di SMK Negeri 1 Kota Jambi yaitu Akuntansi, TKJ, Administrasi Perkantoran, dan Pemasaran. Penelitian ini bertujuan untuk memberi gambaran untuk segmentasi siswa baru berdasarkan jurusan dengan menggunakan data SMK Negeri 1 Kota Jambi. Penelitian ini menggunakan data siswa, data nilai hasil ujian nasional, jurusan yang tersedia, dan kuota penerimaan siswa. Hasil yang diperoleh yaitu laporan data pendaftar, laporan hasil

penyeleksian siswa baru berdasarkan kuota, dan hasil pengelompokan siswa yang telah di seleksi. Hasil tersebut berfungsi sebagai penimbang penerimaan siswa baru tiap tahun ajaran dan menjadikan sistem kerja yang efisien bagi pihak sekolah [7].

Selain itu Sembiring, Rohmah, dan Erfina pernah meneliti algoritme *clustering* dengan metode *K-Means* dengan studi kasus pada SMK YASPIM Gegerbitung dalam jurnal berjudul “Implementasi Algoritme K-Means Clustering Analysis untuk Menentukan Hambatan Pembelajaran Daring”. Menurut Kepala Sekolah SMK YASPIM Gegerbitung pembelajaran daring dari rumah menyebabkan kemunculan masalah baru seperti, membuat guru SMK YASPIM Gegerbitung keberatan karena harus ikut aturan Kemdikbud. Selain itu, siswa juga merasa keberatan karena proses belajar daring dari rumah tidak efektif dan tugas terlalu banyak, dan orang tua murid harus memberikan fasilitas untuk belajar daring di rumah. Tujuan dari penelitian ini adalah menentukan tingkat hambatan pembelajaran daring dengan data seperti sarana penunjang, stabilitas jaringan, dan nilai rata rata. Dalam penelitian ini penulis menganalisis tingkat hambatan pembelajaran daring di SMK YASPIM dengan mengaplikasikan algoritme *k-means clustering*. Hasil penelitian ini didapatkan 6 kelas *cluster* tinggi, dan 9 kelas *cluster* rendah pada hambatan pembelajaran daring di SMK YASPIM Gegerbitung [8].

Penelitian algoritme *clustering* dengan metode *K-Means* lainnya dilakukan oleh Muzakir dalam jurnal berjudul “Analisa dan Pemanfaatan Algoritme *K-Means* pada Data Nilai Siswa sebagai Penentuan Penerima Beasiswa”. Penelitian ini bertujuan untuk menentukan siswa mana yang pantas menerima beasiswa berdasarkan nilai siswa. Penelitian ini menggunakan 3 variabel dalam menentukan kelayakan dalam calon penerima beasiswa yaitu nilai matematika, nilai bahasa inggris, dan nilai komputer. Apabila ketiga nilai dikategorikan nilai baik maka siswa akan diberikan rekomendasi untuk memiliki beasiswa tersebut. Hasil yang diharapkan adalah penentuan data mahasiswa yang sesuai untuk memperoleh rekomendasi beasiswa ke perguruan tinggi dengan menggunakan variabel tertentu, seperti data tarif mahasiswa dari kelas 1 sampai kelas 2 dan data pendapatan orang

tua. Hasilnya berupa kelompok nilai baik (calon penerima beasiswa) dan kelompok nilai rendah (gugur). Alur penelitian ini dimulai dengan input data yang diperlukan, lalu di hitung dengan *K-Means*, dan *output* dalam bentuk *table* kelompok nilai baik (calon penerima beasiswa) dan kelompok nilai rendah (gugur) [9].

Ada pun penelitian yang dilakukan Muningsih dan Kiswati yang berjudul “Penerapan Metode *K-Means* Untuk *Clustering* Produk *Online Shop* Dalam Penentuan Stok Barang”. Manajemen stok yang berantakan dan asal menyebabkan biaya penyimpanan menjadi mahal, karena menyebabkan stok berlebih atau kurang pada beberapa produk. Hal ini sangat merugikan secara ekonomi bagi pengusaha seperti contohnya pada *online shop*. Maka dari itu penelitian ini berfungsi untuk menentukan produk mana yang harus memiliki stok banyak, sedang, atau sedikit. Metode *K-Means* yang diaplikasikan pada penelitian dan metode ini yang merupakan salah satu metode populer dan relatif baik untuk *clustering*. Variabel yang dijadikan acuan penelitian adalah rata-rata penjualan, kode produk, *volume* penjualan dan jumlah transaksi. Penelitian ini menggunakan *software RapidMiner* untuk mengolah data dengan metode *K-Means*. Hasil akhir dari penelitian adalah berupa suatu program aplikasi yang dapat mengelompokkan produk menjadi kategori jumlah stok sedikit, sedang dan banyak berdasarkan transaksi penjualan. Setelah tingkatan produk ditentukan, data ini menjadi informasi yang berguna untuk meningkatkan efisiensi penyetokan barang di produk *online shop* tersebut [10].

Sebuah penelitian lain dilakukan oleh Muningsih, Maryani, Handayani dengan jurnal berjudul “Penerapan Metode *K-Means* dan Optimasi Jumlah *Cluster* dengan *Index Davies Bouldin* untuk *Clustering* Propinsi Berdasarkan Potensi Desa”. Metode *K-Means* adalah metode yang seringkali dipraktikan dalam penelitian pengelompokan dan salah satu penelitiannya adalah penelitian ini. *K-Means* memiliki kelemahan dalam menentukan jumlah *cluster*. Penelitian ini akan mengaplikasikan *DBI* yang memiliki kepanjangan Indeks *Davies Bouldin* sebagai cara untuk mengoptimalkan jumlah *cluster* untuk segmentasi provinsi yang berdasar pada potensi desa dengan ragam aneka jenis industri milik wilayahnya masing-masing. Data yang digunakan adalah banyaknya kelurahan atau desa

menurut keberadaan dan jenis industri mikro dan kecil. Untuk mengolah data digunakan aplikasi *RapidMiner*. Pengujian dilakukan dengan mengambil nilai paling kecil dari DBI. Penelitian ini menghasilkan segmentasi desa berdasarkan potensi desa yang mana acuan potensi desa berdasarkan banyaknya jenis industri. Hasil *clustering* ini bisa dijadikan acuan pemerintah dalam membuat program untuk masing-masing provinsi kedepannya [11].

Metisen dan Sari pun pernah melakukan penelitian dengan judul “Analisis *Clustering* Menggunakan Metode *K-Means* dalam Pengelompokan Penjualan Produk pada Swalayan”. Penelitian ini menjelaskan tentang *software Tanagra* pada *data mining*. *Tanagra* merupakan *software* untuk *data mining* yang dapat memiliki akses terhadap metode metode *data mining* yang ada. Aplikasi ini menggunakan *dataset input* dan menggunakan tiga variabel yang berasal dari data barang di Swalayan Fadhilla Bengkulu. Variabel tersebut antara lain nama barang, jumlah(stok), dan jumlah barang terjual. Dalam penelitian ini, digunakan metode *clustering* dengan menggunakan algoritme *K-means*. Penelitian dengan sampel data yang diambil di Swalayan Fadhilla Bengkulu ini menghasilkan dua kelompok data yaitu data penjualan rendah dan data penjualan tinggi. Tujuan penelitian ini adalah pihak Swalayan Fadhilla dapat mengetahui jenis barang yang laris terjual dan tidak, sehingga gudang tidak menumpuk [12].

Penelitian yang dilakukan oleh Agustin, Fitria, dan Hanifah dengan studi kasus SMP Negeri 101 Jakarta dan berjudul “Implementasi *Algoritme K-Means* untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional”. Pengayaan materi adalah tahap persiapan murid dalam menghadapi UN. Di SMP Negeri 101 Jakarta ada dua pengayaan materi, yaitu pengayaan materi khusus dan pengayaan materi wajib. Pengayaan materi khusus berorientasi pada nilai akhir semester 5. Segmentasi kemampuan siswa diukur dengan nilai rapor semester 1 sampai semester 5 ditambah nilai asesmen akhir untuk data yang lebih akurat. Untuk itu diperlukan solusi yang dapat mengatasi kesulitan tersebut. Maka dari itu metode *clustering* algoritme *K-Means* diterapkan pada kasus ini. *Cluster* berjumlah empat sesuai jumlah mata pelajaran UN, sedangkan jumlah sampel data adalah 12 siswa dengan nilai terendah. Aplikasi ini menampilkan

segmentasi keahlian siswa pada mata pelajaran UN dan berfungsi untuk memantau perkembangan kemampuan setelah mengikuti pengayaan materi. Penelitian ini menghasilkan segmentasi kemampuan siswa pada UN dan dapat digunakan untuk memantau perkembangan kemampuan setelah mengikuti pengayaan materi [13].

Penelitian dari Wardhani dengan judul “Implementasi *Algoritme K-Means* untuk Pengelompokan Penyakit Pasien pada PUSKESMAS KAJEN PEKALONGAN”. Metode yang digunakan adalah metode *k-means*. Penelitian ini menggunakan alat *Rapid Miner 5.3*. Berdasarkan data dari Puskesmas KAJEN PEKALONGAN, hasil pengelompokan adalah 376 *item* dari akut dan 624 penyakit tidak akut dari 1000 total data. Penelitian ini menghasilkan informasi segmentasi dari penyakit akut dan tidak akut pasien puskesmas berdasarkan usia, jenis kelamin, lama penyakit dan diagnosis penyakit. Hasil tersebut dapat dijadikan bahan atau dasar penyuluhan kesehatan oleh dinas kesehatan setempat [14].

Salsabila pernah membuat penelitian dengan judul “Klasifikasi Barang Menggunakan Metode *Clustering K-Means* dalam Penentuan Prediksi Stok Barang”. Penelitian ini menggunakan metode *K-Means Clustering*. Penelitian ini menggunakan data berupa jumlah stok, harga pembelian bahan baku, total transaksi, keuntungan, tanggal transaksi. Penelitian ini menggunakan ROC (*Receiver Operating Characteristic*) sebagai perhitungan akurasi. Perhitungan *clustering* dengan metode *K-Means* menggunakan RFM (*Recency Frequency Monetary*) mendapatkan hasil 70% dan 76,67% tanpa menggunakan RFM. Dari akurasi tersebut dapat disimpulkan bahwa *clustering* dengan metode *K-Means* mendukung sistem dengan baik. Penelitian ini menghasilkan segmentasi produk dengan jumlah stok sedikit karena produk kurang diminati, stok sedang karena diminati, dan stok banyak karena paling diminati. Hasil tersebut berguna untuk meningkatkan efisiensi dalam penyetokan barang [15].

Tabel 2. 1 Perbandingan Penelitian yang telah Dilakukan

Pembanding	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	Penulis
Jumlah Sampel	10	600	15	60	31	34	10	12	1000	20	45
Jumlah Cluster	3	3	2	2	3	3	2	4	2	3	3
Objek	Desa	Siswa	Kelas	Siswa	Produk Online Shop	Provinsi	Produk Swalayan	Siswa	Pasien	Jilbab	Tipe Motor
Metode	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>	<i>Clustering</i>
Algoritme	<i>K Means</i>	<i>K Means</i>	<i>K Means</i>	<i>K Means</i>	<i>K Means</i>	<i>K Means</i>	<i>K Means</i>				
Tools	-	-	-	PHP, MySQL	RapidMiner	RapidMiner	Tanagra, SPSS, XLMiner	-	RapidMiner	Sublime Text Editor	RapidMiner

Perbedaan penelitian penulis dengan penelitian sebelumnya, terletak pada objek penelitian. Objek penelitian penulis berupa produk otomotif. Penelitian sebelumnya menggunakan orang, kelas, kebutuhan primer, dan wilayah sebagai objek penelitian. Untuk aspek pembanding lainnya tidak memiliki perbedaan yang signifikan.