

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Pustaka

Pada penelitian ini peneliti telah melakukan studi literatur terhadap beberapa jurnal penelitian terbaru sebelumnya. Jurnal tersebut dipilih berdasarkan topik yang sesuai dan berkaitan dengan tema penelitian yang diteliti yaitu mengenai analisis data pengiriman barang menggunakan metode *K-Means Clustering*. Berikut penjelasan lebih lanjut.

Penelitian pertama membahas tentang Penerapan *Data Mining* untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman di PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar [3]. Dimana data minat konsumen terhadap jasa pengiriman JNE dengan menerapkan *data mining*, menggunakan *K-Means Clustering*. Data yang digunakan adalah data pada bulan September, Oktober, dan November 2019 dengan cabang di 27 kota/kabupaten atau kecamatan. Hasil yang diperoleh berupa *cluster* tinggi sebanyak 12 kota/kabupaten atau kecamatan dan *cluster* rendah sebanyak 15 kota/kabupaten atau kecamatan. Data hasil penelitian ini digunakan untuk pertimbangan pengambilan keputusan dalam menentukan minat konsumen PT JNE [3].

Penelitian kedua adalah penelitian mengenai analisis perbandingan pengelompokan *export* dan *import* di Indonesia berdasarkan hitungan bulan menggunakan K-Means [12]. Penelitian ini bertujuan untuk mengelompokkan ekspor dan impor di Indonesia berdasarkan bulan untuk mendapatkan perbandingan tingkat ekspor dan impor, apakah tinggi atau rendah. Pada penelitian ini

digunakan *data mining K-Means* yang diimplementasikan menggunakan *RapidMiner*. Data yang digunakan bersumber dari Badan Pusat Statistik Indonesia dari Januari – Desember 2018. Menghasilkan kluster ekspor tinggi terdiri dari 5 bulan dan kategori rendah 7 bulan. Sedangkan kluster impor tinggi terdiri dari 7 bulan dan kategori rendah terdiri dari 5 bulan [12].

Penelitian ketiga, yaitu analisis *clustering* dengan metode *K-Means* untuk pengelompokan penjualan produk pada swalayan Fadhila [13]. Yang menjelaskan tentang *software* aplikasi Tanagra pada *data mining* yang dapat digunakan untuk mengakses beberapa metode *data mining* yang ada. Aplikasi ini juga menggunakan *dataset input*, yaitu menggunakan data barang di swalayan Fadhila Bengkulu. Dalam penerapannya menggunakan *clustering* dengan algoritma *K-Means*, dan didapatkan dua jenis kelompok data, yaitu data penjualan rendah dan tinggi. Kesimpulannya jika hasilnya ini membantu untuk mengetahui jenis barang yang laris terjual atau tidak, sehingga barang yang ada di gudang tidak menumpuk [13].

Penelitian keempat metode *clustering K-Means* dalam pengelompokan penjualan produk [7]. Tujuannya untuk merancang aplikasi dan menganalisis penjualan produk pada swalayan MM.TIKA kota Bengkulu menggunakan *K-Means*. Sistem tersebut memberikan kemudahan untuk menganalisis pengelompokan penjualan produk pada swalayan MM.TIKA, serta menentukan dan mengklasifikasi barang yang laku atau kurang laku. Mereka melakukan pengumpulan data adalah melalui observasi dan wawancara. Setelah mendapatkan hasilnya, pemilik dapat mencari alternatif produk lain untuk menggantikan produk yang kurang laku [7].

Penelitian kelima merupakan penerapan *data mining* pada *export* buah menurut negara tujuan menggunakan *K-Means Clustering* [14]. Mereka menggunakan data dari berbagai dokumen keterangan ekspor impor yang dihasilkan oleh Direktorat Jenderal Bea dan Cukai dari 2002 – 2015 sebanyak 11 data. Variabel yang mereka gunakan adalah jumlah

ekspor berat bersih (netto) dan nilai *Free On Board* (FOB). Kemudian data diolah dengan *clustering*, yang terbagi menjadi *cluster* tingkat ekspor tinggi, sedang, dan rendah. Dengan Centroid data yang mereka gunakan untuk *cluster* tingkat ekspor tinggi 904.276,5, Centroid data sedang 265.501 dan rendah 34.280,1. Hasilnya, indeks ekspor buah dengan 2 negara *cluster* tingkat ekspor tinggi, yaitu India dan Pakistan, 3 negara *cluster* tingkat ekspor sedang, yaitu Singapura, Bangladesh dan negara lainnya, dan 6 negara *cluster* tingkat ekspor rendah, yaitu Hongkong, Tiongkok, Malaysia, Nepal, Vietnam dan Iran. Ini berguna sebagai masukan kepada pemerintah, tentang negara mana saja yang menjadi prioritas tertinggi pada kegiatan ekspor berdasarkan *cluster* yang telah dilakukan [14].

Penelitian keenam tentang *clustering K-Means* menggunakan data ekspor studi kasus PT. GAIKINDO [15]. Untuk mengetahui data mobil terlaris dan tidak laris untuk mengatur volume pengiriman mobil ke luar negeri menggunakan metode *clustering* dengan algoritma *K-Means*. Mereka terlebih dahulu mengelompokkan data dengan cara memasukkan data *brand* dan tipe mobil yang dikirim pada tiap negara. Hasilnya menunjukkan jika terdapat 3 *cluster* yang terbentuk dengan kategori sangat diminati, cukup diminati, dan sedikit diminati [15].

Penelitian ketujuh, *K-Means Clustering* untuk mengelompokkan data pelabuhan, serta bongkar muat barang di Indonesia [16]. Pada penelitian ini dilakukan proses *clustering* jumlah pelabuhan di daerah untuk dilakukan analisa dari pembagian data pelabuhan pada tahun 2016 hingga 2017, apakah itu di usahakan atau tidak dalam beberapa tahun terakhir dengan memanfaatkan algoritma *K-Means Clustering*. Hasilnya adalah perkembangan pelabuhan ditentukan oleh jumlah bongkar/muat barang baik dalam negeri ataupun luar negeri, pertumbuhan pelabuhan yang diusahakan dibagi menjadi beberapa *cluster* mulai dari sangat rendah hingga sangat tinggi, serta kesimpulan bahwa perkembangan pelabuhan tidak hanya ditinjau dari bongkar/muat barang saja, tetapi ada beberapa

faktor yang mempengaruhi, seperti batasan geografis, masalah tenaga kerja, kurangnya keamanan, dan kurangnya prasarana pelabuhan [16].

Penelitian kedelapan ini mengenai klustering data penjualan toko perlengkapan *outdoor* menggunakan metode K-Means [5]. Untuk membantu toko Genta Corp, sebuah toko retail di kota Bogor yang menjual peralatan outdoor untuk mengelompokkan data penjualan guna memaksimalkan manajemen stok dan menyusun strategi pemasaran. Variabel yang digunakan, diantaranya : kode barang, data barang masuk, data barang keluar, dan stok barang. Pengolahan data dilakukan secara manual dengan menggunakan algoritma *K-Means* dan software Rapid Miner. Hasilnya didapatkan tiga cluster, yaitu 2 jenis barang terlaris, 8 jenis barang cukup laris, dan 18 jenis barang kurang laris [5].

Penelitian yang terakhir ini membahas tentang penerapan *K-Means clustering* data COVID-19 di Provinsi Jakarta [17]. Pada penelitian yang membahas tentang pengelompokan yang dilakukan dengan parameternya pasien positif dan meninggal dari tiap kelurahan di Provinsi Jakarta. Pengolahan data dengan menggunakan *K-Means* untuk *clusterisasi* dan Google Colaboratory. Dari hasil analisis yang dilakukan terdapat persebaran data pada bagian *cluster* 1 sebanyak 173 dan *cluster* 2 sebanyak 18. Hasilnya memberikan informasi daerah dengan jumlah kasus positif, serta tingkat kesembuhan tertinggi maupun terendah yang dapat digunakan sebagai evaluasi dalam proses penanganan virus COVID-19 [17].

Berdasarkan kajian penelitian terdahulu, maka dapat disimpulkan bahwa metode *K-Means Clustering* dapat dilakukan untuk mengetahui hasil dari *clustering* sebuah *dataset*. Oleh karena itu, penulis menerapkan metode *K-Means Clustering* pada penelitian Analisis Data Pengiriman Barang Ekspedisi untuk Mengelompokkan Potensi Barang Pengiriman dengan menggunakan Google Colaboratory (Google Colab), studi kasus TIKI KC Purbalingga.

Tabel 2.1. Penelitian Terdahulu

No	Peneliti	Judul	Metode	Hasil
1	Rudi Wardana Nasution, Suhada, Ika Okta Kirana, Indra Gunawan, Ika Purnama Sari	Penerapan <i>Data Mining</i> untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman di PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar [3].	<i>Data Mining</i> menggunakan <i>K-Means Clustering</i>	<i>Cluster</i> tinggi sebanyak 12 kota/kabupaten/kecamatan dan <i>cluster</i> rendah sebanyak 15 kota/kabupaten/kecamatan. Data hasil penelitian ini digunakan untuk pertimbangan pengambilan keputusan dalam menentukan minat konsumen PT JNE [3].
2	Boby Iskandar, S Saifullah, Eka Irawan, Irfan Sudahri Damanik, Ilham Syahputra Saragih	Analisis perbandingan pengelompokan <i>export</i> dan <i>import</i> di Indonesia berdasarkan hitungan bulan menggunakan <i>K-Means</i> [12].	<i>K-Means</i>	Menghasilkan kluster ekspor tinggi terdiri dari 5 bulan dan kategori rendah 7 bulan. Sedangkan kluster impor tinggi terdiri dari 7 bulan dan kategori rendah terdiri dari 5 bulan [12].

No	Peneliti	Judul	Metode	Hasil
3	Benri Melpa Metisen, Herlina Latipa Sari	<i>Clustering</i> dengan metode <i>K-Means</i> untuk pengelompokan penjualan produk pada swalayan Fadhila [13].	<i>K-Means</i>	Kesimpulannya jika hasilnya ini membantu untuk mengetahui jenis barang yang laris terjual atau tidak, sehingga barang yang ada di gudang tidak menumpuk [13].
4	Yulia Darmi, Agus Setiawan	<i>Clustering K-Means</i> dalam pengelompokan penjualan produk [7].	<i>K-Means</i>	Sistem tersebut memberikan kemudahan untuk menganalisis pengelompokan penjualan produk pada swalayan MM.TIKA, serta menentukan dan mengklasifikasi barang yang laku atau kurang laku.
5	Agus Perdana Windarto	Penerapan <i>data mining</i> pada <i>export</i> buah menurut negara tujuan menggunakan <i>K-Means Clustering</i> [14].	<i>K-Means</i>	Hasilnya, indeks ekspor buah dengan 2 negara cluster tingkat ekspor tinggi, yaitu India dan Pakistan, 3 negara <i>cluster</i> tingkat ekspor sedang, yaitu Singapura, Bangladesh dan negara lainnya, dan 6 negara <i>cluster</i> tingkat ekspor rendah, yaitu

No	Peneliti	Judul	Metode	Hasil
				Hongkong, Tiongkok, Malaysia, Nepal, Vietnam dan Iran. Ini berguna sebagai masukan kepada pemerintah, tentang negara mana saja yang menjadi prioritas tertinggi pada kegiatan ekspor berdasarkan <i>cluster</i> yang telah dilakukan [14].
6	Wakhid Afifi, Dhiya'an Ramadhanty Nastiti, dan Qurrotul Aini	<i>Clustering K-Means</i> menggunakan data ekspor studi kasus PT. GAIKINDO [15].	<i>K-Means</i>	Hasilnya menunjukkan jika terdapat 3 <i>cluster</i> yang terbentuk dengan kategori sangat diminati, cukup diminati, dan sedikit diminati [15].
7	Abdi Hazman, Dr.Purba Daru Kusuma, S.Si., M.T., Casi Setianingsih, S.T., M.T.	<i>K-Means Clustering</i> untuk mengelompokkan data pelabuhan, serta bongkar muat barang di Indonesia [16].	<i>K-Means</i>	perkembangan pelabuhan ditentukan oleh jumlah bongkar/muat barang baik dalam negeri ataupun luar negeri, pertumbuhan pelabuhan yang diusahakan dibagi menjadi beberapa <i>cluster</i> mulai dari sangat rendah hingga sangat tinggi, serta kesimpulan bahwa

No	Peneliti	Judul	Metode	Hasil
				perkembangan pelabuhan tidak hanya ditinjau dari bongkar/muat barang saja, tetapi ada beberapa faktor yang mempengaruhi, seperti batasan geografis, masalah tenaga kerja, kurangnya keamanan, dan kurangnya prasarana pelabuhan [16].
8	Fintri Indriyani, Eni Irfiani	klastering data penjualan toko perlengkapan <i>outdoor</i> menggunakan metode K-Means [5].	<i>K-Means</i>	Menghasilkan rekomendasi barang yang laris, kurang laris, dan cukup laris, sehingga data dijadikan rujukan bagi manajemen untuk mengatur stok barang agar toko tidak mengecewakan pelanggan karena barang yang ingin dibeli tidak tersedia.
9	Meida Cahyo Untoro, Leslie Anggraini, Maria Andini, Hesti Retnosari, M. Anas Nasrulloh	Penerapan <i>K-Means clustering</i> data COVID-19 di Provinsi Jakarta [17].	<i>K-Means</i>	Informasi daerah dengan jumlah kasus positif, serta tingkat kesembuhan tertinggi maupun terendah yang dapat digunakan sebagai evaluasi dalam proses penanganan

No	Peneliti	Judul	Metode	Hasil
				virus COVID-19. Persebaran data terbagi menjadi 2 <i>cluster</i> , yaitu <i>cluster</i> 1 sebanyak 173 data dan <i>cluster</i> 2 sebanyak 18 data. [17].

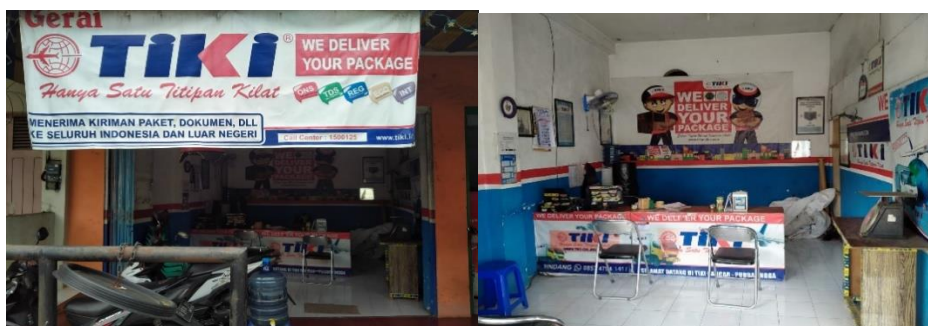
Berdasarkan penelitian sebelumnya tepatnya yang telah dijelaskan pada kajian pustaka, maka penelitian ini penulis akan melakukan analisis data pengiriman barang ekspedisi untuk mengelompokkan potensi barang pengiriman menggunakan metode *K-Means Clustering* studi kasus TIKI Kantor Cabang Purbalingga, yang memanfaatkan data pengiriman periode 2018 – 2021.

2.2 Dasar Teori

Berikut adalah kajian mengenai beberapa teori yang digunakan pada penelitian ini :

2.2.1 Ekspedisi (TIKI)

PT. Citra Van Titipan Kilat atau TIKI merupakan salah satu perusahaan jasa pengiriman terkemuka, yang didirikan di Jakarta pada 1 September 1970[1]. Kini TIKI telah berkembang pesat dengan sebaran jaringan operasional di 65 kota besar di Indonesia, dengan lebih dari 500 kantor perwakilan, lebih dari 3700 gerai, dan 6000 lebih karyawan (<https://tiki.id/id/tentang-tiki>)[2]. Salah satunya adalah TIKI Kantor Cabang (KC) Purbalingga yang beralamat di Jl. Letnan Yusuf Kalimanah RT 004/02, Babakan, Purbalingga Utara SPBU Puji Utami, yang didirikan oleh pasangan yang bernama Mardiono dan Misri. Sejalan dengan visi dan misi yang dimiliki TIKI, TIKI KC Purbalingga ini memiliki visi menjadi yang terbaik dalam melayani kebutuhan pelanggan dalam jasa pengiriman. Sementara misinya, yaitu bekerja secara jujur, bertanggung jawab, dan menjunjung tinggi profesionalisme. Sama halnya dengan kantor cabang lainnya, TIKI KC Purbalingga ini juga melayani pengiriman paket dan dokumen[1].



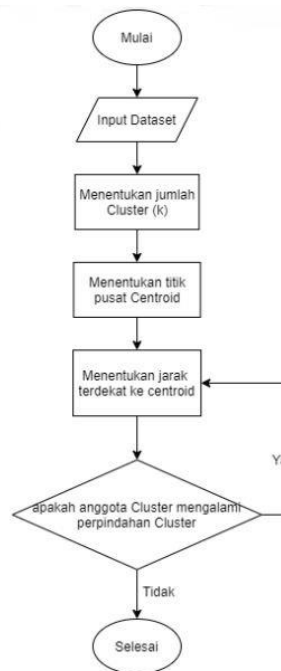
Gambar 2.1. TIKI KC Purbalingga

2.2.2 *Data Mining*

Data mining atau *pattern recognition* merupakan sebuah metode pengolahan data untuk menemukan adanya pola tersembunyi dari sebuah data. Hasil pengolahan data menggunakan *data mining* dapat dimanfaatkan untuk pengambilan keputusan di masa depan. Berkat kemampuannya dalam mengolah data dalam skala besar, menjadikan *data mining* berperan penting dalam bidang industri, keuangan, cuaca, ilmu pengetahuan, dan juga teknologi. Ada berbagai metode dalam *data mining*, diantaranya *clustering*, klasifikasi, regresi, seleksi variabel, dan market basket analisis [3][4].

2.2.3 **K-Means**

K-Means adalah metode dalam *data mining*[5]. Pada *K-Means* analisis *cluster* dilakukan untuk membagi objek menjadi k *cluster* kemudian diamati dimana setiap objek *cluster* diperoleh dari rata – rata terdekat[6]. *K-Means* juga termasuk salah satu metode pengelompokan data *nonhierarki* (sekatan) yang berusaha membagi data (partisi) ke dalam bentuk dua atau lebih kelompok (*cluster*) dengan masing – masing karakteristik tiap kelompoknya[7]. Algoritma *K-Means* sangatlah sederhana dan cukup cepat dalam pengimplementasiannya, sehingga menjadi salah satu yang paling sering digunakan. Bahkan berhasil diimplementasikan pada berbagai bidang, seperti *computer vision*, *statistic to market segmentation*[8][9]. Cara kerjanya dengan memposisikan titik sebagai titik tengah (*centroid*) tiap *cluster*. Kemudian mengaitkan setiap titik kumpulan data ke *centroid* terdekat dengan menggunakan *Euclidian*, dan dilakukan berulang hingga tidak ada perubahan pada *centroid* [8]. Tahapan *K-Means Clustering* dijelaskan pada *flowchart* berikut ini.



Gambar 2.1 Flowchart algoritma *K-Means Clustering*[10]

2.2.4 Clustering

Clustering bertujuan untuk mengelompokkan setiap objek yang memiliki kesamaan, sedangkan objek yang berbeda akan dikelompokkan pada grup yang berbeda juga. *Clustering* juga merupakan teknik yang banyak digunakan untuk pengolahan data dan juga dalam mengelompokkan sebuah objek yang memiliki kesamaan. Prinsipnya untuk mengelompokkan objek adalah dengan memaksimalkan kemiripan dari intra kelas dan meminimalkan kemiripan antar kelas. Penilaian kesamaan dan ketidaksamaan dari nilai – nilai atribut yang menggambarkan objek dan ukuran yang berbeda[8], [11]. Proses *clustering* dimulai dengan identifikasi data yang akan di *cluster*, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan di *cluster* dan m adalah jumlah variabel. Pada awal iterasi nantinya, pusat *cluster* ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Lalu menghitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan perhitungan jarak data ke- i (x_i) pada pusat *cluster* ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidian, seperti persamaan berikut :

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (1)$$

Sebuah data akan menjadi bagian dari anggota *cluster* ke-k jika jarak datanya ke pusat *cluster* ke-k bernilai paling kecil dibandingkan dengan jarak ke pusat *cluster* lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) yang mengelompokkan data – data yang menjadi anggota setiap *cluster*.

$$\text{Min } \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (2)$$

Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata – rata dari data yang menjadi anggota pada *cluster* tersebut dengan menggunakan rumus:

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p} \quad (3)$$

Dimana x_{ij} elemen dari *cluster* ke-k
 p = banyaknya anggota *cluster* ke-k

2.2.5 Python

Python pertama kali dikenalkan oleh Guido Van Rossum (GvR) tahun 1991, untuk menangani kesalahan, serta mengutamakan sintaksis yang mudah dibaca dan dipahami. *Python* juga merupakan bahasa pemrograman yang mendukung multi paradigm pemrograman dan termasuk bahasa nterpretative multifungsi. Yang dilengkapi dengan fungsionalitas pustaka standar besar dan komprehensif atau dapat diterapkan di berbagai platform. Perkembangan pengguna *python* sangatlah tinggi, terhitung pada akhir 2018 saja sudah mencapai 60% baik dari kalangan developer dan juga data scientist. Ini menunjukkan bahwa bahasa ini banyak digunakan oleh *Data Science*, *Machine Learning*, dan *Internet of Things* [10].

2.2.6 Google Collaboratory

Merupakan layanan *cloud* yang mirip dengan *Jupyter notebook*, yang menjadi berbeda adalah *google collaboratory* merupakan layanan *cloud*. Banyak digunakan oleh mahasiswa sarjana ataupun pasca sarjana karena penggunaannya

gratis, dan banyak digunakan untuk pembelajaran *machine learning*. Tentunya jika dibandingkan dengan *Jupiter notebook*, *google collaboratory* lebih canggih karena menggunakan layanan *cloud* [10].