

## BAB II LANDASAN TEORI

### 1.1. Tinjauan Pustaka

Tinjauan pustaka dari penelitian yang dilakukan sebelumnya digunakan untuk mendukung penelitian yang sedang dilakukan. Berikut tinjauan pustaka yang telah dilakukan sebelumnya oleh peneliti terdahulu dapat di lihat di Tabel 2.1:

Tabel 2. 1 Tinjauan Pustaka

1.	Penulis	Ai Rohmah, Falentino Sembiring, Adhitia Erfina
	Tahun	2021
	Judul	Implementasi algoritma k-means clustering analysis untuk menentukan hambatan pembelajaran daring
	Tujuan	Untuk memudahkan pihak sekolah maupun pemerintah dalam mengambil tindakan dalam upaya menunjang proses kegiatan belajar mengajar daring, maka perlunya peneliti memberikan sumbangsih pemikiran untuk menentukan tingkat hambatan pembelajaran daring, yang dibuat menjadi 2 cluster yaitu cluster rendah dan cluster tinggi
	Kesimpulan	Berdasarkan hasil dari penelitian dan pembahasan implementasi metode algoritma k-means clustering analysis untuk menentukan hambatan pembelajaran daring pada siswa dapat diambil kesimpulan sebagai berikut: 1. Dengan menggunakan metode algoritma k-means clustering analysis untuk menentukan hambatan pembelajaran daring pada siswa di SMK YASPIM Gegerbitung diperoleh hasil bahwa terdapat 9 kelas yang masuk kedalam kategori dengan hambatan pembelajaran dari tingkat rendah yaitu kelas 10 TKJ 1, 10 TKJ 2, 10 TKJ 3, 10 RPL, 10 TBSM 1, 10 TBSM 2, 11 RPL, 12 TKJ 1, dan 12 TKJ 2, dan 6 kelas masuk kategori dengan hambatan pembelajaran dari tingkat tinggi yaitu : 10 RPL, 11 TKJ 1, 11 TKJ 2, 11 TKJ 3, 11 TBSM dan 12 RPL. 2. Data yang digunakan untuk penelitian ini diambil

		dari dapodik dan e-rapor yang telah melalui tahap penyaringan (filtering) dan disortir menjadi data per kelas 3. Pada implementasi sistem k-means clustering analysis untuk menentukan hambatan pembelajaran daring telah berhasil dibuat, dengan menggunakan aplikasi pada Orange menunjukkan hasil sesuai dengan aplikasi yang telah dibuat.
2.	Penulis	Lina Listiani, Yoga Handoko Agustin, Mochammad Zaenal Ramdhani
	Tahun	2019
	Judul	Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokkan data penduduk
	Tujuan	pengelompokkan data penduduk untuk rekomendasi jenis pekerjaan berdasarkan faktor pendidikan, umur dan jenis kelamin menggunakan data mining algoritma k-means cluster.
	Kesimpulan	Setelah melakukan analisis, perancangan, implementasi dan pengujian yang telah dilakukan, maka diperoleh kesimpulan terhadap rekomendasi jenis pekerjaan berdasarkan data penduduk sebagai berikut : 1. Dihasilkan 4 kelompok jenis pekerjaan berdasarkan jenis kelamin, pendidikan, umur dan jenis pekerjaan dari data penduduk tahun 2017 untuk bahan rekomendasi jenis pekerjaan pada penduduk Kelurahan Bungursari yang belum memiliki pekerjaan. 2. Dari hasil pengelompokkan diketahui hasil untuk C1 adalah jenis pekerjaan Buruh harian lepas dengan jenis kelamin laki- laki dan perempuan, dan usia antara 21 sampai 35 tahun. C2 adalah jenis pekerjaan Wiraswasta dengan jenis kelamin laki- laki

		<p>dan perempuan, dan usia antara 36 sampai 49 tahun. C3 adalah jenis pekerjaan karyawan swasta dengan jenis kelamin laki- laki dan perempuan, dan usia antara 20 sampai 35 tahun. dan C4 adalah jenis pekerjaan karyawan honorer dengan jenis kelamin laki- laki dan perempuan, dan usia antara 36 sampai 49 tahun. 3. Nilai akurasi yang diperoleh yaitu sebesar 79% dan diklasifikasikan sebagai fair atau cukup pada klasifikasi nilai AUC.</p>
3.	Penulis	Dini Marlina, Nurelina Fauzer Putri, Andri Fernando, Aditya Ramadhan
	Tahun	2018
	Judul	Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak
	Tujuan	Penelitian ini dapat menjadi sebuah solusi agar memudahkan dalam mengetahui wilayah-wilayah yang memiliki tingkat anak cacat yang tinggi sehingga dapat dilakukan penanggulangan secara dini.
	Kesimpulan	K-Medoids mampu melakukan pengelompokan pada data sebaran anak cacat yang ada pada Provinsi Riau. Klaster yang dihasilkan pada penelitian ini adalah berjumlah tiga klaster. Klaster pertama mempunyai pola dengan jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 3, 2, 1, 2, dan 3. Hal ini menunjukkan bahwa pada klaster pertama memiliki data dengan nilai yang paling rendah dibandingkan klaster lainnya. Pada klaster kedua jumlah

		<p>penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 1, 3, 2, dan 3. Hal ini menunjukkan klaster kedua mempunyai data tingkat sebaran anak cacat yang lebih tinggi dibandingkan pada klaster yang pertama. Sedangkan pada klaster yang ketiga, Pola yang terbentuk pada klaster 3 adalah jumlah cacat (semua fisik) antara 015. Sementara jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 5, 3, 4, dan 5. Hal ini menunjukkan klaster ketiga mempunyai data dengan nilai cacat yang paling tinggi dibanding klaster lainnya. Adapun hasil validasi yang dihasilkan adalah 0.5009, ini menunjukkan bahwa Algoritma K-Medoids cukup baik dalam melakukan Pengelompokan pada data sebaran cacat pada Anak di wilayah Provinsi Riau.</p>
4.	Penulis	Abu Salam, Diyan Adiatma, Junta Zeniarja
	Tahun	2020
	Judul	Implementasi Algoritma K-Means dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS
	Tujuan	untuk memecahkan masalah tersebut serta menghasilkan sebuah rekomendasi kelompok penerima beasiswa PPA.
	Kesimpulan	Berdasarkan penelitian yang dilakukan oleh peneliti, maka didapatkan kesimpulan bahwa metode clustering dengan algoritma K-Means dapat

		<p>menghasilkan rekomendasi penerima beasiswa Peningkatan Prestasi Akademik (PPA) dengan melibatkan 7 atribut dan dengan jumlah dataset yang banyak. Penentuan centroid (titik pusat) sangat berpengaruh pada hasil cluster seperti pada hasil pengujian dengan data uji dan data keseluruhan menghasilkan hasil cluster yang berbeda. Pada penelitian yang dilakukan pengujian dengan seluruh data berjumlah 441 dataset, sebanyak 154 mahasiswa direkomendasikan mendapatkan beasiswa PPA sedangkan 287 lainnya tidak mendapatkan, dari 154 data hasil rekomendasi data riil pada tahun 2016 terdapat 113 data pendaftar beasiswa yang dinyatakan diterima dan telah menerima beasiswa PPA, dan setelah dilakukan proses evaluasi oleh biro kemahasiswaan beasiswa tersebut sudah dinyatakan tepat sasaran, sehingga dapat ditarik kesimpulan dengan dilakukan proses rekomendasi pendaftar beasiswa menggunakan metode cluster akan memudahkan dan mempersempit proses seleksi data dengan lebih cepat.</p>
5.	Penulis	Aldi Nurzahputra, Much Aziz Muslim, Miranita Khusniati
	Tahun	2017
	Judul	Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa
	Tujuan	menerapkan k-means dalam mengcluster kinerja dosen berdasarkan indeks kepuasan mahasiswa.

	Kesimpulan	Untuk melakukan penilaian kinerja dosen berdasarkan indeks kepuasan mahasiswa dapat menerapkan metode clustering K-Means. Data diperoleh dari mahasiswa melalui kuisioner sesuai aspek Reliability, Responsiveness, Assurance, dan Empathy. Data diolah untuk memperoleh nilai dari setiap dosen. Data kemudian diolah menggunakan Rapidminer untuk ditentukan nilai centroid dalam cluster baik dan cluster kurang dengan algoritma K-Means. Centroid data untuk cluster baik 17.099 dan cluster kurang 15.874. Sehingga diperoleh penilaian dosen berdasarkan indeks kepuasan mahasiswa dengan 5 dosen cluster baik dan 7 dosen cluster kurang.
--	------------	---

## 1.2.Implementasi

Implementasi berasal dari bahasa Inggris yaitu to implement yang berarti mengimplementasikan. Implementasi merupakan penyediaan sarana untuk melaksanakan sesuatu yang menimbulkan dampak atau akibat terhadap sesuatu. Sesuatu tersebut dilakukan untuk menimbulkan dampak atau akibat itu dapat berupa penerapan peraturan, Peraturan system dari pengolahan hingga hasil akhir yang dianggap permanen.

## 1.3.Kemiskinan

Secara etimologis, “kemiskinan” berasal dari kata “miskin” yang artinya tidak berharta benda dan serba kekurangan. Badan Pusat Statistik mendefinisikan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak lebih jauh disebutkan kemiskinan merupakan sebuah kondisi yang berada dibawah garis nilai standar kebutuhan minimum, baik untuk makanan dan non makanan yang disebut garis kemiskinan (poverty line) atau disebut juga batas kemiskinan (poverty threshold). Indikator kemiskinan yang digunakan secara umum

adalah tingkat upah, pendapatan, konsumsi, harapan hidup rata-rata, tingkat penyerapan anak usia sekolah dasar, proporsi pengeluaran pemerintah untuk pelayanan kebutuhan dasar masyarakat, pemenuhan bahan pangan (kalori/protein), air bersih, perkembangan penduduk, melek huruf, urbanisasi, pendapatan per kapita, dan distribusi pendapatan. Tolak ukur kemiskinan bukan hanya hidup dalam kekurangan pangan dan tingkat pendapatan yang rendah, akan tetapi melihat tingkat pembangunan perkotaan dan perdesaan (Debrina Vita Ferezagia, 2018).

Indeks Kedalaman Kemiskinan (Poverty Gap Index/P1) adalah Ukuran rata - rata kesenjangan pengeluaran masing - masing penduduk miskin terhadap garis kemiskinan. Nilai total Indeks Kesenjangan Kemiskinan menunjukkan biaya pengentasan kemiskinan dengan menetapkan target transfer yang ideal bagi masyarakat miskin dalam hal biaya transaksi dan tanpa insentif. Semakin rendah nilai Depth of Poverty Index, semakin besar potensi ekonomi dari Anti-Poverty Fund, yang mengidentifikasi karakteristik masyarakat miskin dan tujuan dari bantuan dan program mereka. Penurunan nilai Indeks Kesenjangan Kemiskinan menunjukkan bahwa rata-rata pengeluaran masyarakat miskin cenderung mendekati garis kemiskinan, dan ketimpangan pengeluaran masyarakat miskin juga semakin menurun (BPS, 2018).

#### **1.4.Data Mining**

Data mining merupakan kiasan dari bahasa inggris, *mine.jikamine* berarti menambang sumber daya yang tersembunyi di dalam tanah, maka Data Mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligence*), machine learning, statistik dan basis Data. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data mining

adalah langkah analisis terhadap proses penemuan pengetahuan didalam basis data atau *knowledge discovery in databases* yang disingkat KDD. Data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data yang masih berkembang(Dody W Sitohang, Alex Rikki, 2019).

### **1.5.Clustering**

Clustering atau klusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam cluster-cluster. Cluster adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam cluster yang sama dan dissimilar terhadap objek-objek yang berbeda cluster. Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain (Gifthera Dwilestari, Mulyawan, Martanto, Irfan Ali, 2021).

Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised), maksudnya metode ini diterapkan tanpa adanya latihan (training)dan tanpa ada guru serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering (Yuni Radana Sembiring, Saifullah, Riki Winanjaya, 2021).



## 1.6.Python

Python merupakan salah satu bahasa pemrograman dan mendukung pengembangan FOSS dan Linux. Tidak hanya itu, dalam kemajuan teknologi lainnya, Python hadir sebagai salah satu keterampilan yang lebih mumpuni yang perlu kita pelajari. Python biasanya sudah dibangun ke dalam distribusi tertentu seperti Ubuntu, Fedora, dan Slackware. Python sendiri merupakan bahasa pemrograman yang mendukung paradigma pemrograman berorientasi objek atau scripting. Python ada di bidang sistem operasi, virtualisasi, jaringan komputer, grafik komputer, kecerdasan buatan, teknologi web, game, dan banyak lagi. Bahasa pemrograman ini dibuat oleh Guido van Rossum dari Amsterdam, Belanda. Alasan asli untuk membuat bahasa pemrograman ini adalah bahasa skrip tingkat tinggi dari sistem operasi terdistribusi Amoeba. Bahasa pemrograman ini semakin banyak digunakan oleh para insinyur di seluruh dunia untuk membuat perangkat lunak, dan beberapa perusahaan menggunakan Python sebagai produsen perangkat lunak komersial (Suhesti, Tyan, 2014).

## 1.7.Algoritma K – Means

K-Means merupakan Algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah ditetapkan diawal. Algoritma K-Means sederhana untuk diimplementasikan dan dijalankan, relative cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang data mining (Yuni Radana Sembiring, Saifullah, Riki Winanjaya, 2021). Berikut ini adalah langkah-langkah dalam algoritma K-Means:

- a) Tentukan jumlah cluster (k) pada data set.
- b) Tentukan nilai pusat (centroid). Penentuan nilai centroid pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan

(1)

berikut ini:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^n X_k j^2 \quad (1)$$

Keterangan:

$V_{ij}$  = centroid rata-rata cluster ke-I untuk variabel ke-j

$N_i$  = Jumlah anggota cluster ke-i

i, k = indeks dari cluster

j = indeks dari variabel

$X_{kj}$  = nilai data ke-k variabel ke-j untuk cluster tersebut

c) Pada masing-masing record, hitung jarak terdekat dengan centroid. Jarak centroid yang digunakan adalah Euclidean Distance, dengan

d) rumus seperti pada persamaan (2)

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

Keterangan :

De = Euclidean Distance

i = Banyaknya objek <sup>2</sup>

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

e) Kelompokkan objek berdasarkan jarak ke centroid terdekat. Ulangi langkah ke-2, lakukan iterasi hingga centroid bernilai optimal.