

BAB II

LANDASAN TEORI

2.1 Tinjauan Penelitian terdahulu

Berdasarkan hasil *studi literatur* terkait kinerja algoritma K-Means dan K-Medoids pada pengujian *dataset* menggunakan software *Rapidminer* ada beberapa jurnal yang mendekati namun berbeda dalam penggunaan metode dan analisa pengujiannya, kemudian dengan melakukan *literature review* sangat bermanfaat sekali agar penelitian selanjutnya memiliki keterbaruan.

Tools Rapidminer memiliki definisi diantaranya :

Rapidminer merupakan sebuah aplikasi atau *Software* perangkat lunak yang berfungsi sebagai alat pembelajaran ilmu *data mining* (Khalimy , 2020). *Platfrom* dikembangkan oleh suatu perusahaan yang bertujuan untuk bisnis komersial, penelitian, pendidikan, pelatihan, serta semua langkah dalam pembelajaran yang menyangkut suatu data yang besar.

2.2 Literatur Riview

Manfaat dari *literature review* yang dilakukan diantaranya adalah sebagai berikut ;

1. Memperdalam pengetahuan tentang bidang yang di teliti (*Buku Textbook*)
2. Mengetahui hasil penelitian yang berhubungan dan yang sudah pernah dilaksanakan (*related research*) (*Paper*)
3. Mengetahui perkembangan ilmu bidang yang kita pilih (*state-of-the-art research*) (*Paper*)
4. Memperjelas masalah penelitian (*research problems*) (*Paper*)
5. Mengetahui metode-metode terkini yang diusulkan para peneliti demi menyelesaikan masalah penelitian (*state-of-the-art methods*) (*Paper*)

Berdasarkan manfaat tersebut akan dilakukan *literature review* terhadap penelitian yang sedang dilakukan berkaitan dengan “ *Perbandingan Algoritma K-Means dan K-Medoids untuk pengelompokkan Data Covid-19 Di Indonesia*” Beberapa *literature review* yang tercapai mengenai penelitian terdekat maupun terdahulu adalah sebagai berikut ;

Table 1. Tabel Literature Sebelumnya

No.	Penulis/Tahun	Judul	Sumber	Hasil
1.	(Herviany et al. 2021)	Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat	Institute of Research and Publication Indonesia. 1 Iss. 1 April 2021, pp: 34-40 P-ISSN: 0000 0000 E-ISSN: 0000 0000	Pengelompokan metode K-Means lebih Optimal dibandingkan dengan menggunakan metode K-Medoids pada data kejadian tanah longsor Provinsi Jawa Barat pada tahun 2019.
2.	(Sindi et al. 2020)	ANALISIS ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA	JURNAL DASI ISSN: 1411- (Jurnal Teknologi Informasi) Vol.4, No.1, 2020	Algoritma K-Medoids dapat melakukan pengelompokan data covid-19 mana saja wilayah yang terinfeksi di wilayah masing-masing dengan pengklasteran terbaik dilakukan dengan 3 cluster. Dari 34 record diperoleh 1 record pada cluster pertama, 2 record pada cluster kedua, 31 record pada cluster ketiga.
3.	(Rahayu, Prianto, and Novia 2021)	PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK	Jurnal Teknik Informatika, Vol. 13, No. 2, April 2021	Hasil perbandingan yang didapatkan pada kedua metode tersebut untuk nilai accuracy algoritma K-means sebesar 72% dan algoritma Naïve Bayes

		MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN PADA PT. PERTAMINA (PERSERO)		sebesar 95%, untuk nilai precision yang didapatkan oleh algoritma K-means sebesar 72% dan algoritma Naïve Bayes sebesar 100%, dan untuk nilai recall yang didapatkan oleh algoritma K-means sebesar 100% dan algoritma Naïve Bayes sebesar 95%.
4.	(Dahlia et al. 1045)	PENERAPAN DATA MINING TERHADAP DATA COVID-19 MENGGUNAKAN ALGORITMA KLASIFIKASI	Jurnal Informatika, Vol. 21, No.1, Juni 2021	Dari ketiga algoritma, didapatkan akurasi tertinggi dari algoritma Naïve Bayes dengan hasil 80,79% dengan hasil isolated dan memiliki AUC 0.881 yang dikategorikan sebagai good clasification atau klasifikasi yang baik.
5.	(Kamila, Khairunnisa, and Mustakim 2019)	Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau	Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, Vol. 5, No. 1, Februari 2019, Hal. 119-125	Perbandingan antar kedua algoritma clustering, K-Medoids dan K-Means tidak menunjukkan perbedaan yang signifikan mengenai pengelompokan data. Adapun dalam pengolahannya. apabila makin tinggi iterasi dan

				pengelompokkan yang ditentukan, maka pengolahan data akan semakin lama.
--	--	--	--	---

6.	(Euis Saraswati, Yuyun Umaidah, and Apriade Voutama 2021)	Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19	Generation Journal /Vol.5 No.2 / e-ISSN: 2549-2233 / p-ISSN: 2580-4952	Klasifikasi menggunakan ANN memperoleh hasil accuracy sebesar 88,62%, precision 91,5% dan recall 95,73% dengan persentase 80:20 (80% data latih dan 20% data uji) dengan menggunakan 1 hidden layer yang berjumlah 10 node dan learning rate sebesar 0.02.
7.	(Aditya, Jovian, and Sari 2020)	Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019	Jurnal Media Informatika BUDIDARMA Volume 4, Nomor 1, Januari 2020	Hasil cluster dengan algoritma K-Means didapatkan untuk cluster 1 terdapat 14 provinsi, cluster 2 terdapat 5 provinsi, dan cluster 3 terdapat 15 provinsi dengan tingkatan cluster 1 adalah cluster dengan nilai ujian nasional Tinggi, cluster 2 adalah cluster dengan nilai ujian nasional Rendah dan cluster 3 adalah cluster dengan nilai ujian nasional Sedang.
8.	(Pulungan, Suhada, and Suhendro 2019)	Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan	KOMIK (), 3(1), pp. 329–334. doi:	Penerapan data mining dengan menggunakan metode K-Medoids. sebagai perhitungan perbandingannya, Dengan menggunakan 2 cluster yaitu, cluster tinggi dan cluster rendah di peroleh

		Utama	10.30865/komik.v3i1.160 9. 2019.	cluster tinggi sebanyak 3 lapangan pekerjaan utama. Dan cluster rendah sebanyak 14 lapangan
9.	(Silitonga et al. 2019)	Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi	Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019 ISBN:, pp. 313–318. 2019.	Berdasarkan penelitian yang telah dilakukan maka. Data yang diolah menjadi 2 cluster yaitu cluster tingkat memilah sampah rendah (C1) dan cluster tingkat memilah sampah tinggi (C2). Dimana hasil penelitian ini menyimpulkan dari 33 provinsi di indonesia bahwa cluster tingkat perilaku memilah sampah rendah (C1) diperoleh 22 Provinsi dan Cluster rendah sebanyak 12 Provinsi
10.	(Kaligis and Yulianto 2022)	Analisis Perbandingan Algoritma K-Means, K-Medoids, Dan X-	Jurnal Penerapan Teknologi Informasi dan Komunikasi ISSN 2828-7940, e-ISSN	Davies bouldin index pada setiap algoritma yang diterapkan pada penelitian ini, algoritma k-means menjadi metode yang terbaik dalam menentukan

		Means Untuk Pengelompokan Kinerja Pegawai	2829-1727 Volume 01 Nomor 03 Tahun 2022	data kinerja pegawai ke dalam kriteria cluster yang dibutuhkan, karena memiliki nilai DBI paling terkecil, di mana cluster yang dihasilkan dalam metode k-means kriteria nilai tinggi (C1) berjumlah 3 pegawai, sedangkan nilai rendah (C0) berjumlah 25 pegawai.
11.	(Susanti and Widodo 2017)	Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015	Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai Islami) Vol.1, No.1, Juli 2017, Hal. 116-122 p-ISSN: 2580-4596; e-ISSN: 2580-460X	Dengan menggunakan perbandingan rasio simpangan baku dari metode K-Means dan K-Medoids Clustering, maka diperoleh metode terbaik untuk pengklasteran terhadap kelayakan Puskesmas di DIY tahun 2015 adalah metode K-Means Clustering dengan nilai rasio simpangan baku yang lebih kecil dibandingkan dengan metode K-Medoids Clustering, yaitu sebesar

				54,69%.
12.	(Goldblatt 1994)	ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI MAHASISWA NON AKTIF	Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012) Jurnal Vol 2, No 1 (2012)	Berdasarkan hasil perhitungan AUC dari masing-masing algoritma, dapat diketahui bahwa dalam penelitian ini algoritma logistic regression (0,933), decision tree (0,963) , naïve bayes (0,976) dan neural network (0,976) sehingga keempat algoritma tersebut termasuk dalam klasifikasi paling baik.

Tabel 1. merupakan me-review jurnal-jurnal terkait penelitian sebelumnya, sehingga penulis mengambil referesnsi dari jurnal tersebut untuk dijadikan bahan acuan dalam penelitian yang berjudul **“Perbandingan Algoritma K-Means dan K-Medoids untuk pengelompokkan Data Covid-19 Di Indonesia”**. Dari beberapa jurnal yang penulis riview salah satunya mengambil Judul, Peneliti/Tahun, Metode, State of The art dari jurnal tersebut yang berkaitan dengan clustering data mining kinerja algoritma dengan memanfaatkan algoritma *K-Means* dan *K-Medoids* untuk dijadikan *literetur riview*, dengan itu peneliti lebih mudah dalam mencari referensi mulai dari penerapan metode penelitian, tujuan penelitian, masalah penelitian, batasan penelitian untuk dijadikan penelitian terkait kinerja Algoritma K-Means dan K-Medoids untuk pengelompokkan Data Covid-19 Di Indonesia.

2.4 Teori Pendukung Penelitian

Rapidminer

Rapidminer adalah perangkat lunak yang bersifat open source yang memiliki solusi terkait melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik .Sistem Operasi yang dibutuhkan *software RapidMiner* (Sudirman, Windarto, and Wanto 2018) diantaranya ;

1. *Microsoft Windows (x86-32) Windows XP, Windows Server 2003, Windows Vista, Windows Server 2008, Windows 7.*
2. *Microsoft Windows (x64) Windows XP untuk x64, Windows Server 2003 untuk x64, Windows Vista untuk x64, Windows Server 2008 pada x64, Windows 7 untuk x64.*
3. *Unix Sistem 32 atau 64 bit.*
4. *Linux Sistem 32 atau 64 bit.*
5. *Apple Macintosh sistem 32 atau 64 bit.*

2.4.1 Dataset

Dataset adalah kumpulan data, Tabel dan juga di dalamnya terdapat relasi antar data Tabel (data relation) atau lebih mudahnya di dalam satu dataset bisa terdapat banyak data Tabel yang berelasi.

2.4.2 Data Mining

Data mining merupakan suatu istilah yang digunakan saat menguraikan penemuan pengetahuan dalam database. *Data mining* juga disebut sebagai serangkaian proses akan menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Windarto et al. 2020). Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau

tersembunyi di suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

Data Mining dilakukan dengan *tool* khusus, yang mengeksekusi operasi Data Mining yang telah didefinisikan berdasarkan model analisis. Data mining melakukan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi di sejumlah data besar yang disimpan ketika menjalankan bisnis perusahaan.

2.4.3 Arsitektur Data Mining

Pencarian pengetahuan yang menarik dari data berukuran besar yang disimpan dalam basis data, data *warehouse* atau tempat penyimpanan informasi lainnya. Demikian arsitektur *system* data mining memiliki komponen-komponen utama diantaranya ;

1. *Database, data warehouse, world wide web*, atau tempat penyimpanan informasi lainnya, seperti bisa berbentuk satu atau banyak database, data *warehouse, spreadsheet*, ataupun tempat penyimpanan informasi lainnya. *Data Cleaning, Data Integration* dan *Data Selection* dapat dijalankan di data tersebut.

2. *Database dan data warehouse server* komponen ini bertanggung jawab dalam pengambilan data yang relevan, berdasarkan permintaan pengguna.
3. *Knowledge Based* komponen ini merupakan domain *knowledge* yang digunakan agar memandu pencarian atau mengevaluasi pola-pola yang dihasilkan.
4. *Data mining engine* bagian ini merupakan komponen penting dalam arsitektur sistem data mining. Komponen ini terdiri dari modul-modul fungsional seperti karakterisasi, asosiasi, klasifikasi, dan analisis cluster.

Graphical user interface (GUI) modul ini berkomunikasi dengan pengguna dan data mining. Melalui komponen ini, pengguna berinteraksi dengan sistem menggunakan *query*.