

Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering

Siti Nur Arofah¹, Fitri Marisa²

¹sitinurafah1295@gmail.com, ²fitrimarisa@widyagama.ac.id

Program Studi Teknik Informatika, Universitas Widyagama Malang

Abstract— *Mathematics is one of the knowledge, which can improve thinking ability, and argue. Interests are a condition in which a person has an attention to something and is accompanied by the desire to know and study as well as prove it further. Problems in this research how to know the interest of students on the subjects of mathematics by using data mining techniques using Clustering method using K-Means Algorithm. Tests in this study, researchers used data Task Value, Daily Deuteronomic Value (UH) and Values Final Test Semester (UAS) students to know students' interest in mathematics subjects. The result of clustering using k-means algorithm shows that 45 students have high interest in mathematics subjects, 48 students have moderate interest and 29 students have low interest in mathematics.*

Intisari— Matematika merupakan salah satu disiplin ilmu, yang dapat meningkatkan kemampuan berpikir, dan berargumentasi. Minat merupakan suatu keadaan di mana seseorang mempunyai perhatian terhadap sesuatu dan disertai keinginan untuk mengetahui dan mempelajari maupun membuktikannya lebih lanjut. Permasalahan pada penelitian ini bagaimana mengetahui minat siswa pada mata pelajaran matematika dengan menggunakan Teknik data mining dengan metode Clustering menggunakan Algoritma K-Means. Pengujian pada penelitian ini, peneliti menggunakan data Nilai Tugas, Nilai Ulangan Harian (UH) dan Nilai Ujian Akhir Semester (UAS) siswa untuk mengetahui minat siswa terhadap mata pelajaran matematika. Hasil clustering dengan menggunakan algoritma k-means memperlihatkan bahawa 45 siswa memiliki minat tinggi pada mata pelajaran matematika, 48 siswa memiliki minat sedang dan 29 siswa memiliki minat rendah pada mata pelajaran matematika.

Kata Kunci— Matematika, Nilai Tugas, Nilai Ulangan Harian (UH) dan Nilai Ujian Akhir Semester (UAS), Clustering, K-Means

I. PENDAHULUAN

A. Latar Belakang

Kemampuan matematika yang diperlukan untuk menguasai dan menciptakan teknologi di masa depan[1],

menjadikan penguasaan matematika yang kuat perlu dibina sejak dini. Masyarakat beranggapan bahwa Matematika menjadi tolok ukur kecerdasan dan kepandaian anak dalam belajar[2]. Menurut mereka matematika adalah mata pelajaran yang paling baik, berkualitas, dan paling tinggi kedudukannya dibandingkan dengan pelajaran lain[3]. Namun demikian Matematika masih dipandang sebagai salah satu bidang studi yang tidak disenangi atau bahkan paling dibenci dan masih melekat pada kebanyakan siswa yang mempelajarinya[3].

Hal itulah yang mempengaruhi minat seorang siswa dengan matematika berbeda dengan siswa yang lainnya. Minat belajar menurut “Djamarah” dalam cenderung menghasilkan prestasi yang tinggi, sebaliknya minat belajar yang kurang akan menghasilkan prestasi yang rendah. Hal itulah yang menyebabkan minat setiap siswa berbeda dengan siswa lainnya[4]. Dengan permasalahan tersebut maka peneliti menggunakan Teknik data Mining dengan metode Clustering menggunakan algoritma k-means untuk mengelompokkan minat setiap siswa yang memiliki minat tinggi, sedang atau bahkan rendah pada mata pelajaran matematika[5].

Dalam metode clustering mempartisi data ke dalam kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data obyek. Tujuan dari pengelompokan sekumpulan data obyek kedalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan.

Hasil clustering minat siswa pada mata pelajaran matematika dengan menggunakan algoritma k-means di dapatkan 3 cluster. Cluster 0 dengan minat siswa sedang sebanyak (37%) atau sebanyak 45 siswa dari 122 siswa. Cluster 1 dengan minat siswa rendah sebanyak (24%) atau sebanyak 29 siswa dari 122 siswa. Cluster 2 dengan minat siswa tinggi sebanyak (39%) atau sebanyak 48 siswa dari 122 siswa[6].

Dengan demikian minat siswa pada mata pelajaran matematika lebih banyak daripada siswa yang kurang memiliki minat pada mata pelajaran matematika.

II. TINJAUAN PUSTAKA

Penelitian ini bertujuan untuk mengelompokkan data siswa yang memiliki minat dengan mata pelajaran matematika menggunakan algoritma *K-means Clustering*[7]. Untuk mencapai tujuan tersebut, penulis akan menggunakan data siswa kelas VII dari sebuah sekolah swasta di Kota Malang.

A. Algoritma K-Means

Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat kluster (centroid) terdekat dengan data. K-Means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam cluster yang lain [7].

Cara kerja algoritma K-means yaitu sebagai berikut:

- 1) Tentukan nilai *k* sebagai jumlah kluster yang ingin dibentuk.
- 2) Inisialisasi *k* sebagai centroid yang dapat dibangkitkan secara random.
- 3) Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan Euclidean Distance yaitu sebagai berikut:

$$d(P,Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Gambar 1. Persamaan Euclidean

- 4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
- 5) Tentukan posisi centroid baru (*k*)
- 6) kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

III. LANGKAH-LANGKAH PENGELOMPOKKAN DATA

1. Jenis Data

Jenis data yang digunakan adalah data kuantitatif. Data kuantitatif adalah data yang dapat dihitung, berupa angka atau nominal. Data hasil ujian siswa adalah data kuantitatif karena berupa angka atau nominal yang dapat di hitung.

2. Sumber Data

Sumber data yang digunakan dalam penelitian adalah Data Sekunder. Data sekunder merupakan sumber data penelitian yang diperoleh secara tidak langsung melalui media perantara diperoleh dan dicatat oleh pihak lain. Sumber data dalam penelitian ini diambil dari kelas VII dari sebuah sekolah swasta di Kota Malang. Data tersebut merupakan data dari siswa kelas VII semester ganjil tahun ajaran 2017-2018 dengan banyak data 122 data dengan field yaitu Nama Siswa, Nilai Tugas, Nilai Ulangan Harian (UH) dan Nilai Ujian Akhir Semester (UAS). Berikut adalah contoh data siswa kelas VII

Tabel 1
Data Siswa

No	Nama	Nilai Tugas	Nilai UH	Nilai UAS
1	Abhista Irsa Ramadhan	100	99	97
2	Adilla Ghoniyya Ashari	93	93	74
3	Adinda Azkia Putri Lukmana	96	100	87
4	Aflahchintya Azka Ardhana	82	86	56
5	Ahmad Arifin Ilham	88	78	65
6	Ahmad Bayu Nugroho	78	84	66
7	Ahmad Mahbub Thoriq Y.	81	81	57
8	Ahmad Zahran N.	90	76	67
9	Aisyah Putri Kamila	74	73	59
10	Alfiona Absharina Putri	85	88	76
11	Alfi Zahro Firdausy Shomad	88	83	75
12	Alif Rayhan Maron	69	66	43
13	Almira Amelia Giantama	83	98	67
14	Amrita Salsabila A.	99	78	63
15	Ananda Putri Aliyah	96	79	70
16	Anindya Maura Natania	73	75	70

3. Pengumpulan Data

Berdasarkan sumber data yang digunakan pada penelitian ini, maka metode pengumpulan data yang penulis gunakan adalah sebagai berikut,

- a. Dokumentasi, adalah metode pengumpulan data dengan mengumpulkan catatan-catatan atau dokumen-dokumen. Data ini adalah data dari hasil ujian siswa.
- b. Studi Pustaka, adalah mempelajari karya ilmiah, buku ilmiah, dan sumber ilmiah lainnya yang sesuai dengan penelitian dan memiliki hubungan dengan masalah yang diteliti. Refrensi ilmiah yang penulis gunakan adalah sumber-sumber yang terdapat dalam daftar pustaka

B. Aplikasi Weka

WEKA (Waikato Environment for Knowledge Analysis) merupakan aplikasi data mining open source berbasis Java menurut "Sulianta" dalam. WEKA merupakan sebuah sistem data mining yang dikembangkan oleh Universitas Waikato di Selandia Baru yang mengimplementasikan algoritma data mining. WEKA adalah sebuah koleksi mesin pembelajaran algoritma untuk tugas-tugas data mining. Algoritmanya dapat

diterapkan secara langsung ke data set atau dipanggil dari kode Java sendiri.



Gambar 2. Tampilan awal weka tools

Dari data yang di eksekusi, didapatkan 3 kelompok atau cluster dengan hasil sebagai berikut:

- ❖ Siswa dengan nilai = 93 untuk cluster 0, sebanyak 45 Siswa dari 122 Siswa (37%)
- ❖ Siswa dengan nilai = 79 untuk cluster 1, sebanyak 29 Siswa dari 122 Siswa (24%)
- ❖ Siswa dengan nilai = 87 untuk cluster 2, sebanyak 48 Siswa dari 122 Siswa (39%)

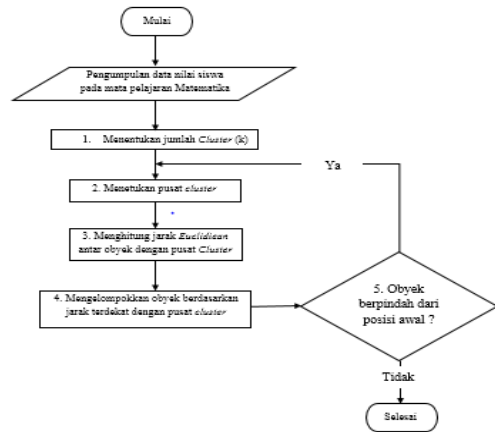
IV. ANALISIS

A. Analisis

Untuk mengetahui minat setiap siswa tersebut maka peneliti melakukan penelitian dengan menggunakan perhitungan nilai siswa menggunakan metode *Clustering* dengan algoritma *K-means*. Dalam penelitian ini data yang digunakan peneliti adalah data nilai siswa kelas VIII semester ganjil tahun ajaran 2017/2018 dengan banyak data 122 data siswa. Data tersebut di peroleh dari Nilai Tugas, Nilai Ulangan Harian dan Nilai Ujian Akhir Semester. Dengan adanya penelitian inidiharapkan guru akan mengetahui minat setiap siswa.

B. Perancangan

Pada perancangan penelitian ini metode *cluster* dengan algoritma *k-means* akan digambarkan dengan flowchart yang akan di jabarkan dengan langkah-langkah perhitungannya.



Gambar 3. Flowchart Analisis Cluster

Dari gambar flowchart di atas, dapat diketahui pada tahap ini dilakukan proses pengolahan data menggunakan algoritma *K-means Clustering*. Tahapan proses algoritma *K-means* adalah sebagai berikut:

- Tentukan nilai k dari jumlah *cluster* yang ingin dibentuk. Cluster yang akan dibentuk adalah 3 cluster.
- Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal yang terbentuk yakni cluster nilai tertinggi, cluster nilai sedang dan cluster nilai rendah. Titik pusat dari setiap cluster dapat dilihat pada tabel 2.

Tabel 2.

Titik Pusat Awal Setelah Cluster

Titik Pusat Cluster	Nama	Nilai Tugas	Nilai UH	Nilai UAS
Cluster 0	Nanda Lidya Cinta Aulia Sari	98	94	89
Cluster 1	Alif Rayhan Maron	69	66	43
Cluster 2	Ayshaqirana	70	84	93

- Setelah menentukan titik pusat awal cluster langkah selanjutnya yaitu hitung jarak centroid setiap data ke masing-masing centroid menggunakan persamaan euclidean distance. Sebagai contoh akan dihitung jarak dari data siswa pertama ke pusat cluster pertama:

$$D(0,0) = \sqrt{(100 - 98)^2 + (99 - 94)^2 + (97 - 89)^2} = 9,6437(2)$$

$$D(0,1) = \sqrt{(100 - 69)^2 + (99 - 66)^2 + (97 - 43)^2} = 70.4699 (2)$$

$$D(0,3) = \sqrt{(100 - 70)^2 + (99 - 84)^2 + (97 - 93)^2} = 33.7787(2)$$

- Dari hasil perhitungan diatas di dapatkan hasil bahwa jarak data siswa pertama dengan pusat cluster pertama adalah 0, jarak data siswa pusat cluster kedua adalah 41.7732 dan jarak data siswa pusat cluster ketiga

adalah 70.4699. Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data siswa pertama yang paling dekat dengan cluster 1, sehingga data siswa data mahasiswa pertama dimasukkan ke dalam cluster 1.

- Setelah semua data ditempatkan ke dalam cluster terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata yang ada pada cluster tersebut. Dalam hal ini jika centroid yang baru tidak ada perubahan dari centroid yang lama maka hentikan iterasi, jika ada perubahan maka lanjutkan iterasi berikutnya.
- Pengelompokan hasil cluster pada iterasi pertama belum sama, dalam hal ini bangkitkan ulang centroid baru dengan rumus:

$$C = \frac{\sum m}{n} \quad (3)$$

Sumber: Larose (2015) dalam [4]

Dimana :

C : centroid data

m : anggota data yang termasuk kedalam centroid tertentu

n : jumlah data yang menjadi anggota centroid tertentu

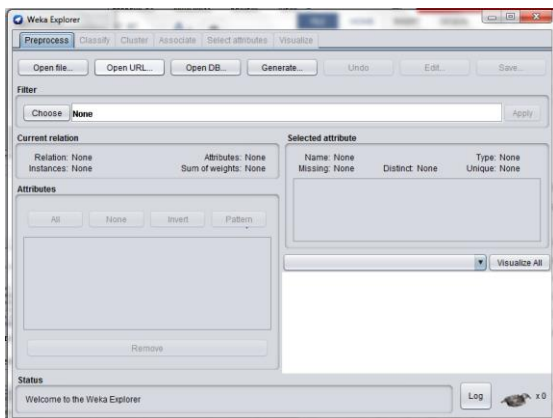
V. HASIL DAN PEMBAHASAN

A. Implementasi Clustering

Penelitian ini menggunakan program aplikasi weka 3.8 untuk mengelompokkan minat siswa melalui Nilai Tugas, Ulangan Harian (UH) dan Nilai Ujaian Akhir Semester (UAS) dengan menggunakan metode clustering, yakni algoritma *k-means*. Penelitian ini dilakukan pada nilai semester ganjil siswa kelas delapan di sebuah sekolah swasta di Kota Malang tahun ajaran 2017/2018. Dalam penelitian ini titik pusat awal yang terbentuk yakni cluster nilai tertinggi, cluster nilai sedang dan cluster nilai rendah.

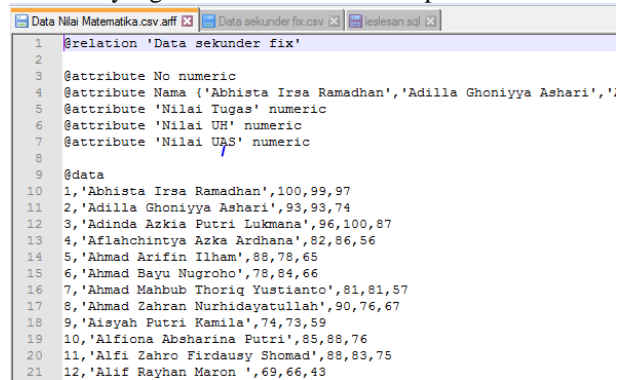
B. Proses Input dan Output

Pada saat membuka aplikasi WEKA ada 4 tombol yang dapat di gunakan untuk menjalankan aplikasi. Untuk menguji clustering ini menggunakan tombol explorer yang akan memunculkan tampilan sebagai berikut.



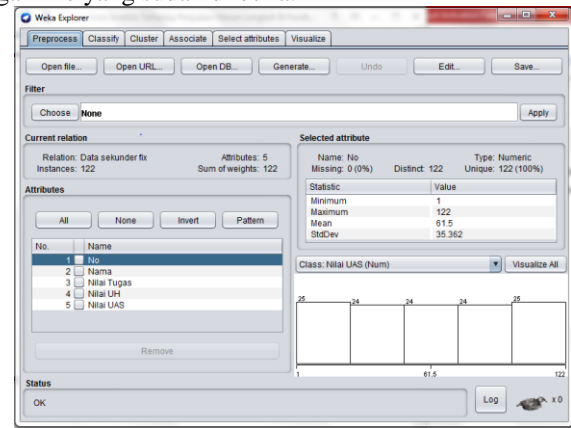
Gambar 4. Weka Explorer

Pada bagian atas window, tepatnya pada bawah judul bar.Terdapat deretan data, seperti Preprocess,Classify,Cluster,Associate,Select Attributes Visualize. Namun yang aktif hanya Preprocess ini dikarenakan sebelum menggunakan algoritma diatas pastikan sudah melakukan set file yang akan dieksekusi. Buka file yang akan dieksekusi oleh Weka, dalam hal ini file yang akan di eksekusi menggunakan ekstensi .csv dengan type ARFF. Berikut adalah gambaran file yang akan di eksekusi dalam penelitian ini



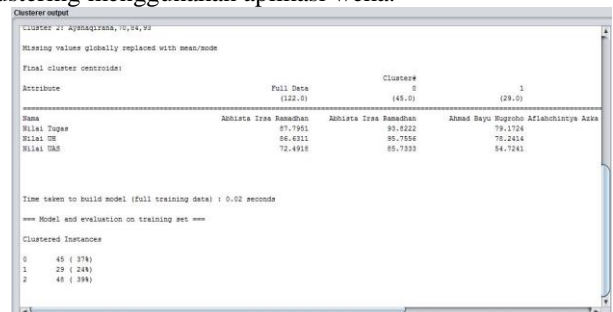
Gambar 5. Data dengan format .arff

Di bawah ini adalah tampilan Praprocess pada weka dengan file yang sudah di buka.

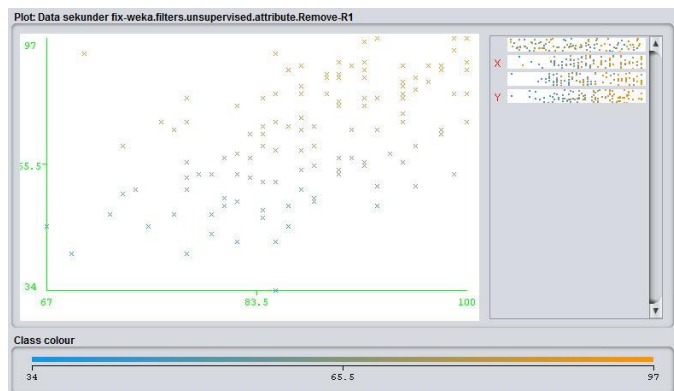


Gambar 6. File yang sudah di buka pada apk Weka

Setelah membuka file yang akan di eksekusi langkah berikutnya yaitu memilih algoritma yang akan di gunakan. Pada penelitian ini menggunakan *Algoritma Clustering* dengan *Cluster Simple K-Means*. Di bawah ini adalah hasil clustering menggunakan aplikasi weka.



Gambar 7. Hasi Clustering menggunakan Weka



Gambar 8. Hasil Grafik Clustering dengan Weka

Dari data yang di eksekusi, didapatkan 3 kelompok atau cluster dengan hasil sebagai berikut:

- ❖ Siswa dengan nilai = 93 untuk cluster 0, sebanyak 45 Siswa dari 122 Siswa (37%)
- ❖ Siswa dengan nilai = 79 untuk cluster 1, sebanyak 29 Siswa dari 122 Siswa (24%)
- ❖ Siswa dengan nilai = 87 untuk cluster 2, sebanyak 48 Siswa dari 122 Siswa (39%)

Maka cluster 0 dengan nilai tertinggi di miliki sebanyak 45 siswa, jumlah ini lebih sedikit daripada cluster 2 yang notabene memiliki minat mata pelajaran matematika yang sedang. Namun hal bagusnya yaitu cluster 0 tidak lebih banyak daripada cluster 1. Yaitu siswa yang memiliki minat mata pelajaran matematika lebih banyak daripada siswa yang kurang memiliki minat dengan matematika.

V. PENUTUP

A. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat di ketahui bahwa dengan kumpulan nilai ujian yang di lakukan siswa, dapat di ketahui seberapa minat siswa dengan mata pelajaran matematika. Hasil *clustering* dengan menggunakan algoritma *k-means* memperlihatkan bahawa 45 siswa memiliki minat tinggi dengan mata pelajaran matematika, 48 siswa memiliki minat sedang dan 29 siswa memiliki minat rendah pada mata pelajaran matematika.

B. Saran

Saran dari peneliti adalah perlu dicoba algoritma lain untuk melakukan pengelompokan minat siswa sehingga diharapkan pengelompokan yang dilakukan memiliki nilai akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] D. R. A, A. Vienda, S. Florencius, and S. Santi, "Peranan Teknologi Komunikasi Dalam Menciptakan Masyarakat Di Masa Depan," pp. 73–86.
- [2] T. Kinerja, P. Pada, R. S. P. M. C. Pekanbaru, D. Machasin, M. Si, and D. C. Amsal, "TOLAK UKUR KECERDASAN EMOSIONAL, KECERDASAN

- INTELEKTUAL, DAN DISIPLIN TERHADAP KINERJA PERAWAT PADA R.S PMC PEKANBARU." [3] P. Pendidikan, M. Pascasarjana, P. Alam, and S. U. Indonesia, "MEMBANGUN KARAKTER MELALUI PEMBELAJARAN MATEMATIKA," vol. 6, pp. 130–141.
- [4] M. Achdiyat, "Prestasi belajar matematika ditinjau dari kepercayaan diri dan keaktifan siswa," vol. 6, no. 1, pp. 50–61, 2016.
- [5] I. Method, K. C. Based, S. Value, W. Interface, C. Study, and I. U. M. M. Magelang, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface," vol. 18, no. 1, pp. 76–82, 2015.
- [6] F. L. Sibuea, A. Sapta, S. Informasi, and S. Royal, "SISWA BERPRESTASI MENGGUNAKAN METODE K-MEANS CLUSTERING," vol. IV, no. 1, 2017.
- [7] B. M. Metisen and H. L. Sari, "ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN," vol. 11, no. 2, pp. 110–118, 2015.

