JURNAL DISPROTEK
ISSN: 2088-6500 (p); 2548-4168 (e)
Vol 15, No. 1, Januari 2024, hlm. 1-8
DOI: 10.34001/jdpt
https://ejournal.unisnu.ac.id/JDPT

PENERAPAN TEKNIK DATA MINING UNTUK MENENTUKAN POLA KINERJA SISWA SD SANTA MARIA REMBANG DALAM MENGHADAPI UJIAN KELULUSAN

APPLICATION OF DATA MINING TECHNIQUES TO DETERMINE PERFORMANCE PATTERNS OF STUDENTS AT SD SANTA MARIA REMBANG IN FACING THE GRADUATION EXAM

Sholihul Ibad[1]

[1] Institut Teknologi dan Bisnis Tuban
Email : [1*]sholihulibad80@gmail.com
**Penulis Korespondensi*

Abstrak – SD Santa Maria merupakan salah satu Sekolah Dasar yang berada di Kabupaten Rembang. Saat menghadapi ujian kelulusan, SD Santa Maria memiliki masalah kinerja atau kemampuan siswanya menurun, khuhusnya kegagalan pada ujian matematika dan bahasa. Tujuan dari penelitian ini adalah mengelompokkan siswa berdasarkan kinerja siswa dan mengetahui pola dari kinerja siswanya dalam menghadapi ujian kelulusan berdasarkan kelompok yang terbentuk. Metode yang diusulkan terfokus pada metode *clustering* dan klasifikasi. Metode *clustering* dengan menggunakan algoritma K-Means bertujuan untuk mencari jumlah kelompok *(cluster)* yang paling optimal dan mengelompokkan siswa pada *cluster* yang sesuai. Dengan demikian, metode klasifikasi yang diusulkan menggunakan algoritma *Decision Tree* untuk menemukan pola dari siswa berdasarkan kelompok *(cluster)* yang terbentuk. Hasil penelitian ini menghasilkan k=2 sebagai kelompok *(cluster)* yang paling optimal dan mengelompokkan siswa menjadi 2 kelompok (*cluster* 0 dan *cluster* 1). Hasil *cluster* yang memuat *cluster* sebagai atribut baru ini lah yang digunakan untuk metode klasifikasi dengan algoritma *Decision Tree*. *Cluster* dijadikan sebagai label dan pola yang dihasilkan berbentuk *decision tree* (pohon keputusan) atau peraturan *if-then*.

Kata kunci: Kinerja Siswa; Data Mining; Clustering; Klasifikasi;

Abstract *- SD Santa Maria is one of the elementary school in Rembang Regency. SD Santa Maria has problems with the performance or reduced ability of students in the matriculation examination, especially failures in mathematics and language tests. The purpose of this study is to group students on the basis of student performance and on the basis of the formed groups to determine the pattern of student performance when approaching the matriculation exam. The proposed method focuses on clustering and classification methods. The K-Means clustering method aims to find the best number of groups (clusters) and groups of students in the respective clusters. Therefore, the proposed classification method uses a decision tree algorithm to find patterns of students based on the created groups (clusters). The results of this research showed that k = 2 as the best group (cluster) and the grouping of students into 2 groups (cluster 0 and cluster 1). The clustering results with clusters as a new attribute are used for the classification method of the decision tree algorithm. Clusters are used as labels and the resulting patterns are in the form of decision trees or rules.*

Keywords*: Student Performance; Data Mining; Clustering; Classification;*

1. PENDAHULUAN

Didalam dunia pendidikan tentunya ada standarisasi dalam mengukur kelulusan siswa ataupun peserta didiknya, standarisasi merupakan salah satu bentuk kesesuaian dan kepadanan mengikuti suatu pedoman, hal ini dapat diartikan bahwa pendidikan juga mempunya kesesuaian dan standar kelulusan siswa tersebut. Standarisasi dunia pendidikan Sekolah Dasar (SD) adalah paduan nilai ujian nasional dan nilai rapor sekolah yang menunjukan kemampuan siswa itu di sekolah. Siswa mempunyai kewajiban untuk belajar, dalam hal ini harus

lebih giat lagi dalam belajar karena akan menghadapi Ujian Nasional (UN). Sebagai salah satu syarat untuk bisa melanjutkan jenjang pendidikan ke tingkat selanjutnya dan seperti yang kita ketahui, bahwa akhir-akhir ini standar kelulusan Ujian Nasional (UN) di indonesia semakin tinggi. Oleh karena itu, sekolah seharusnya mengetahui apa yang menjadi faktor-faktor yang menentukan tingkat kelulusan siswanya. Hal penting dalam menentukan tingkat kelulusan siswanya adalah kinerja atau kemampuan siswa saat ujian. Untuk mengetahui faktor-faktor internal dan eksternal yang mempengaruhi kinerja siswa tersebut merupakan hal yang penting agar persiapan saat ujian dapat semaksimal mungkin.

SD Santa Maria merupakan salah satu Sekolah Dasar yang ada di Kabupaten Rembang. Saat menghadapi ujian kelulusan, SD Santa Maria memiliki masalah kinerja atau kemampuan siswanya menurun, khuhusnya kegagalan pada ujian matematika dan bahasa. Dengan kegagalan tersebut, kepala sekolah SD Santa Maria ingin mengelompokkan siswa berdasarkan kinerja siswa dan mengetahui pola dari kinerja siswanya dalam menghadapi ujian kelulusan berdasarkan kelompok yang terbentuk. Data mining dapat diaplikasikan pada bidang lembaga atau institusi pendidikan dan sering disebut juga dengan Educational Data Mining (EDM) [1]. Data mining dapat juga didefenisikan menggali data dari banyaknya informasi yang akan dicari sehingga data yang perlu diketahui akan lebih mudah dicari dengan adanya sistem pola yang dibuat berdasarkan titik terdekat dengan informasi yang sering di perlukan. Pemilihan suatu metode ataupun algoritma dari data mining ini juga harus tepat karena sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan [2]. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan aturan atau pola dari ratusan atau ribuan data dari sebuah relasasibasis data yang sangat besar [3].

Dari hal itulah, penulis memberikan solusi berupa penerapan teknik data mining. Data mining yang digunakan terfokus pada metode *clustering* dan klasifikasi. Metode *clustering* dengan menggunakan algoritma K-Means bertujuan untuk mencari jumlah kelompok *(cluster)* yang paling optimal dan mengelompokkan siswa pada *cluster* yang sesuai. Dengan demikian, metode klasifikasi yang diusulkan menggunakan algoritma *Decision Tree* untuk menemukan pola dari siswa berdasarkan kelompok *(cluster)* yang terbentuk.

2. METODE PENELITIAN

Pada metode penelitian menyampaikan langkah-langkah yang dilakukan penulis untuk mengumpulkan data dan informasi. Langkah-langkah yang dilakukan penulis sebagai berikut:

2.1. PENELITIAN YANG BERHUBUNGAN

Terdapat beberapa penelitian yang telah dilakukan sebelumnya tentang prediksi kinerja siswa, salah satunya penelitian yang dilakukan oleh Amjad Abu Saa yang berjudul "Educational Data Mining & Students Performance Prediction". Penelitian ini mempelajari dan menganalisis data pendidikan terutama kinerja siswa. Educational Data Mining (EDM) adalah bidang studi yang berkaitan dengan penggalian data pendidikan untuk mengetahui pola dan pengetahuan yang menarik dalam organisasi pendidikan. Studi ini sama-sama memperhatikan subjek ini, khususnya, kinerja siswa. Studi ini mengeksplorasi beberapa faktor yang secara teoritis dianggap mempengaruhi kinerja siswa di pendidikan tinggi, dan menemukan model kualitatif yang paling baik mengklasifikasikan dan memprediksi kinerja siswa berdasarkan faktor pribadi dan sosial yang terkait [4].

Penelitian selanjutnya dilakukan oleh Cortez, dkk yang berjudul "Using data mining to predict secondary school student performance". Penelitian sini membahas tingkat pendidikan penduduk Portugis telah meningkat dalam beberapa dekade terakhir, statistik menjaga Portugal di ujung ekor Eropa karena tingkat kegagalan siswa yang tinggi. Secara khusus, kurangnya keberhasilan dalam kelas inti Matematika dan bahasa Portugis sangat serius. Di sisi lain, bidang Business Intelligence (BI)/Data Mining (DM), yang bertujuan mengekstraksi pengetahuan tingkat tinggi dari data mentah, menawarkan alat otomatis yang menarik yang dapat membantu domain pendidikan. Penelitian ini bermaksud untuk mendekati prestasi siswa di pendidikan menengah menggunakan teknik BI/DM metode Decision Tree. Hasilnya menunjukkan bahwa akurasi prediksi yang baik dapat dicapai [5].

Penelitian selanjutnya dilakukan oleh Widyahastuti, dkk yang berjudul "Student Performance Prediction Using Data Mining Techniques". Penelitian ini membahas kurangnya sistem yang ada untuk menganalisis dan menilai kinerja dan kemajuan siswa tidak diperhatikan. Ada 2 alasan mengapa hal ini sering terjadi. Pertama, sistem yang ada tidak akurat untuk memprediksi kinerja siswa. Kedua, karena kurangnya pertimbangan terhadap beberapa faktor vital yang mempengaruhi kinerja siswa. Memprediksi kinerja siswa adalah tugas yang lebih menantang sebagai akibat dari sejumlah besar informasi dalam database akademik. Sistem yang diusulkan ini dapat membantu untuk memprediksi kinerja siswa secara lebih akurat. Untuk pendekatan data mining yang sesuai akan diterapkan. Dalam pendekatan ini, langkah preprocessing akan diterapkan pada dataset mentah sehingga algoritma penambangan akan diterapkan dengan benar. Prediksi tentang kinerja siswa dapat membantunya untuk meningkatkan kinerjanya [6].

Penelitian selanjutnya dilakukan oleh Kumar,dkk yang berjudul "Mining Educational Data to Analyze Students Performance". Penelitian ini membahas salah satu cara untuk mencapai tingkat kualitas tertinggi dalam

sistem pendidikan tinggi adalah dengan menemukan pengetahuan untuk prediksi tentang pendaftaran siswa dalam kursus tertentu, keterasingan model pengajaran kelas tradisional, deteksi cara tidak adil yang digunakan dalam ujian online, deteksi nilai abnormal dalam hasil. lembar siswa, prediksi tentang kinerja siswa dan sebagainya. Pengetahuan tersembunyi di antara kumpulan data pendidikan dan dapat diekstraksi melalui teknik penambangan data. Penelitian ini dirancang untuk membenarkan kemampuan teknik data mining dalam konteks pendidikan tinggi dengan menawarkan model data mining untuk sistem pendidikan tinggi di universitas. Dalam penelitian ini, tugas klasifikasi digunakan untuk mengevaluasi kinerja siswa dan karena ada banyak pendekatan yang digunakan untuk klasifikasi data, metode pohon keputusan digunakan di sini. Dengan tugas ini kami mengekstrak pengetahuan yang menggambarkan kinerja siswa dalam ujian akhir semester. Ini membantu lebih awal dalam mengidentifikasi anak putus sekolah dan siswa yang membutuhkan perhatian khusus dan memungkinkan guru untuk memberikan nasihat/konseling yang tepat [7].

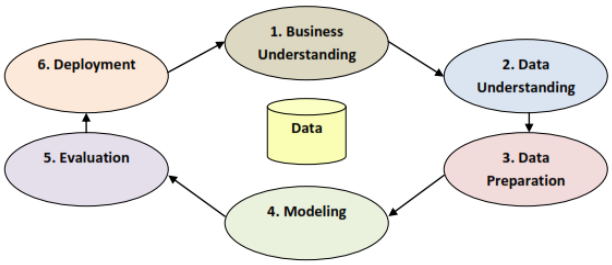
Penelitian selanjutnya dilakukan oleh Rahma,dkk yang berjudul "Implementation of C4 . 5 Algorithm to Predict the Ability of Junior High School Student to Face National Exam". Penelitian ini membahas sekolah membutuhkan sebuah data mining dengan metode klasifikasi yang dapat membantu memprediksi kesiapan siswa menghadapi ujian nasional. Metode tersebut mengolah data yang memiliki perbedaan atribut kemudian disusun ke dalam kategori yang sesuai. Data tersebut di prediksi menggunakan algoritma Decision Tree, yang merupakan salah satu machine learning menggunakan perhitungan probabilitas. Penggunaan algoritma ini didukung dengan simulasi yang dilakukan menggunakan aplikasi RapidMiner dan mendapatkan nilai akurasi sebesar 99,48%. Dari hasil simulasi tersebut, diolah kembali menjadi sebuah aplikasi yang ditujukan untuk membantu pihak sekolah. Untuk menguji kegunaan dari aplikasi tersebut dilakukan penyebaran kuesioner berjumlah 10 soal ke 20 guru dan memperoleh hasil index 83,3% yang berarti memuaskan [8].

Penelitian selanjutnya dilakukan oleh Fajar,dkk yang berjudul "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa". Dalam penelitian ini, data mahasiswa yang diambil oleh penulis adalah nomor induk mahasiswa, jalur masuk perguruan tinggi, pendapatan orang tua dan indeks prestasi komulatif. Penggalian informasi pada sebuah data berukuran besar tidak dapat dilakukan dengan mudah dan hal ini bisa dilakukan dengan teknologi data mining. Data mining yang disebut juga dengan Knowledge Discovery in Database adalah sebuah proses secara otomatis atas pencarian data didalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi hubungan (association) atau pengelompokan (clustering). Dengan menggunakan metode k-means clustering, peneliti mencoba untuk mengekstrak pengetahuan yang bisa menggambarkan kinerja prestasi akademik mahasiswa pada akhir semester dan hasil dari penelitian tersebut menunjukkan bahwa dari semua jumlah cluster yang dimasukkan, untuk cluster yang berjumlah 3 memiliki nilai silhouette coefficient yang paling mendekati nilai Si= 1, yaitu dengan nilai 0,108690751. Selain itu pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik mahasiswa dan nilai akademis mahasiswa yang masuk melalui jalur reguler & jalur prestasi akademik mempunyai nilai IPK rata-rata tertinggi. Sehingga, pihak fakultas dapat mempertimbangkan untuk lebih memprioritaskan penerimaan mahasiswa baru melalui jalur reguler & prestasi akademik. [9]

2.2. METODE PENELITIAN

Metode yang digunakan penulis dalam penelitian ini adalah metode *CRISP-DM (CRoss-Industry Standard Process for Data Mining)*. Tahapan dalam data mining terbagi dalam beberapa langkah yang disebut CRISP-DM *(CRoss-Industry Standard Process for Data Mining)* [10] yaitu antara lain adalah:

1. *Business Understanding/Organizational Understanding* (Pemahaman bisnis/ organisasi): Tahap pemahaman sistem yang berjalan dan kebutuhan apa yang dibutuhkan dalam menyelesaikan masalah yang timbul didalamnya.
2. Data *Understanding* (Pemahaman data): Tahap pemahaman dan pengumpulan data yang dibutuhkan untuk sebelum dilakukan persiapan untuk analisa. Pada tahap ini data yang dikumpulkan harus merupakan data yang tepat digunakan untuk proses penelitian dan mewakili masalah yang akan dipecahkan serta sesuai dengan kebutuhan dan kepentingan.
3. Data *Preparation* (Persiapan data): Tahap persiapan dan seleksi data yang telah dikumpulkan dan diubah menjadi bentuk yang dapat diolah dalam model yang ditentukan selanjutnya.
4. *Modeling* (Pemodelan): Proses analisa dan pemodelan data yang telah disiapkan dimana dalam ini dilakukan penerapan atau penghitungan berdasarkan algoritma atau metode yang ditentukan untuk mendapatkan hasil yang diinginkan sesuai dengan kebutuhan pengguna dan melakukan representasi pemecahan masalah.
5. *Evaluation* (Evaluasi): Melakukan analisa dan evaluasi dari hasil model yang telah dibuat apakasudah sesuai standar dan telah memecahkan masalah atau memenuhi kebutuhan dari pengguna.
6. *Deployment* (Penerapan): Tahap penerapan hasil dari model yang telah dievaluasi dan dianalisa untuk kemudian dijadikan bentuk yang dapat diolah kembali.



Gambar 1. Tahapan Metode CRISP-DM (Sumber: [10])

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada jurnal ini meliputi:

3.1. Business Understanding

Saat menghadapi ujian kelulusan, SD Santa Maria memiliki masalah kinerja atau kemampuan siswanya menurun, khuhusnya kegagalan pada ujian matematika dan bahasa. Dari masalah tersebut, penelitian ini bertujuan untuk:

1. Mencari jumlah kelompok (cluster) yang paling optimal dan mengelompokkan siswa pada cluster yang sesuai
2. Menemukan pola dari siswa berdasarkan kelompok (cluster) yang terbentuk

3.2. Data Understanding

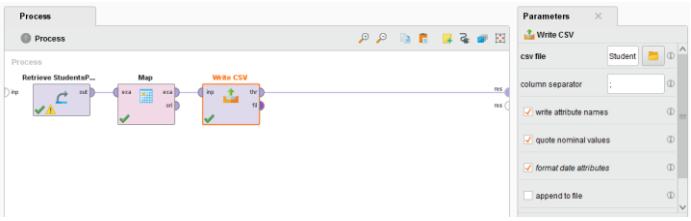
Untuk keperluan pengelompokan tersebut Kepala Sekolah SD Santa Maria mendapatkan 549 data yang dipilih secara random dari nilai berbagai mata pelajaran (StudentsPerformance.csv) dan atribut yang dipilih pada tabel 1.

Tabel 1. Atribut Dataset StudentsPerformance.csv (Sumber: [6])

Atribut	**Nilai**					**Keterangan**
Gender	0 =Female	1=Male				Jenis Kelamin
Race/Ethnicity	1=Group A	2=Group B	3=Group C	4=Group D	5 Group E	Ras/Etnis
Parental Level Of Education	1=Junior High School	2=High School	3=Associate's Degree	4=Bachelor's Degree	5=Master's Degree	Tingkat Pendidikan Orang Tua
Test Preparation Course	0=None	1= Completed				Kursus Persiapan Ujian
Math Score	Skala 0-100					Skor Menghitung
Reading Score	Skala 0-100					Skor Membaca
Writing Score	Skala 0-100					Skor Menulis
Gender	0=Female	1=Male				Jenis Kelamin

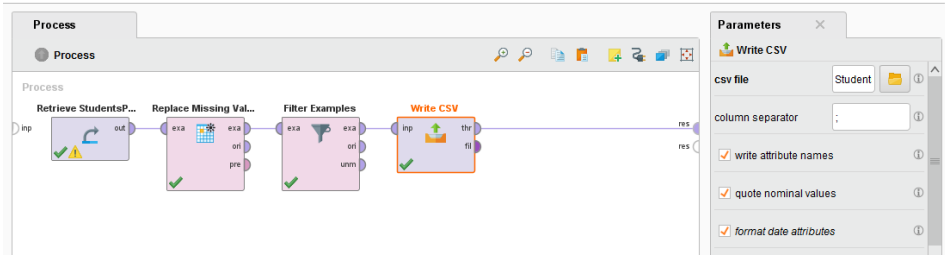
3.3. Data Preparation

Untuk keperluan pengelompokan (cluster), dataset StudentsPerformance.csv yang type datanya nominal diubah menjadi numerik dengan menggunakan operator Map dan menggunakan operator Write CSV untuk menyimpan data yang sudah diubah menjadi numerik (StudentsPerformance-numeric.csv).



Gambar 2. Penggunakan Operator Map dan Write CSV

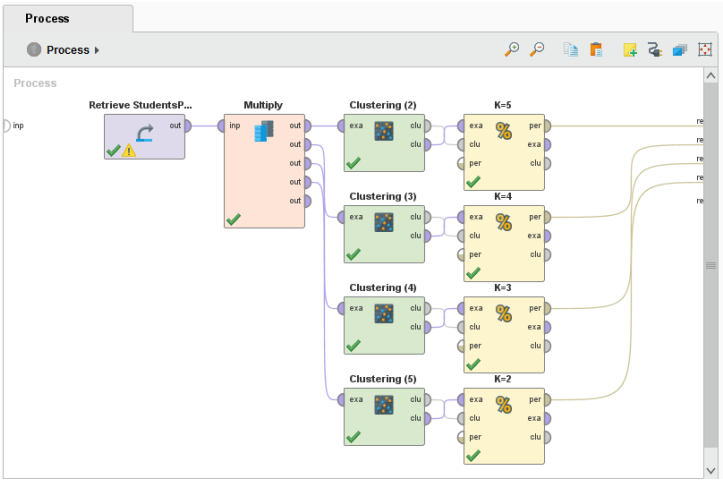
Dataset StudentsPerformance-numeric.csv ini lah yang dipakai untuk clustering, akan tetapi data StudentsPerformance-numeric.csv memiliki masalah yaitu ada *Missing* (data yang kosong) dan *Noisy* (nilainya tidak wajar). Maka dari itu, data harus dibersihkan terlebih dahulu menggunakan operator Replace Missing Value digunakan untuk mengganti data yang kosong dengan nilai rata-rata dan operator Filter Examples untuk menghapus nilai yang tidak wajar tersebut. Setelah datanya bersih, simpan data menggunakan operator Write CSV (StudentsPerformance-clean.csv).



Gambar 3. Proses *Cleaning* Data *Missing* dan *Noisy*

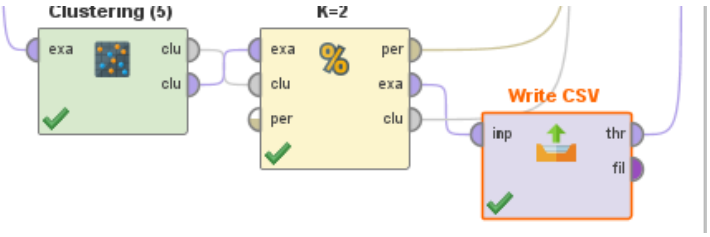
3.4. Modeling

Setelah datanya bersih, dilakukan proses modeling menggunkan 5 peran data mining. Tujuan 1 menggunkan Metode clustering dengan menggunakan algoritma K-Means dengan menentukan nilai k paling optimal dengan melakukan ujicoba pengubahan nilai k (k=5, k=4, k=3, k=2) selanjutnya melakukan pengukuran *performance* dengan menggunakan *Cluster Distance Performance*, untuk mendapatkan nilai *Davies Bouldin Index* (DBI). Nilai DBI semakin rendah berarti cluster yang kita bentuk semakin baik



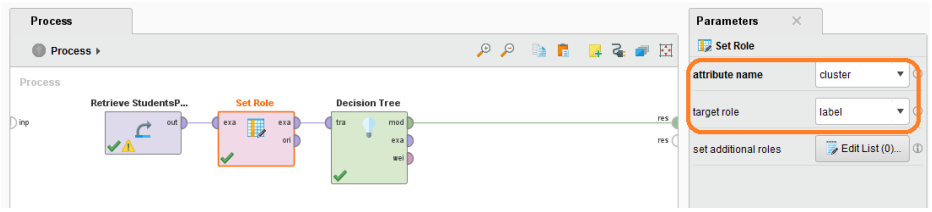
Gambar 4. Implementasi K-Means dan *Cluster Distance Performance* pada RapidMiner

Setelah mendapatkan nilai k yang paling optimal (k=2), simpan data hasil cluster yang memuat cluster sebagai atribut baru dengan operator Write CSV (StudentsPerformance-cluster.csv).



Gambar 5. Proses Simpan Hasil Cluster

Hasil cluster yang memuat cluster sebagai atribut baru ini lah yang digunakan untuk menemukan pola dari kinerja siswa berdasarkan kelompok yang terbentuk (tujuan 2). Dengan demikian, metode klasifikasi yang diusulkan menggunakan algoritma Decision Tree. Cluster dijadikan label dan pola yang dihasilkan bisa berbentuk decision tree (pohon keputusan) atau if-then.



Gambar 6. Implementasi Decision Tree pada RapidMiner

3.5. Evaluation

Evaluasi tujuan 1 berupa nilai *Davies Bouldin Index* (DBI) sebagai berikut:

Tabel 2. Nilai DBI

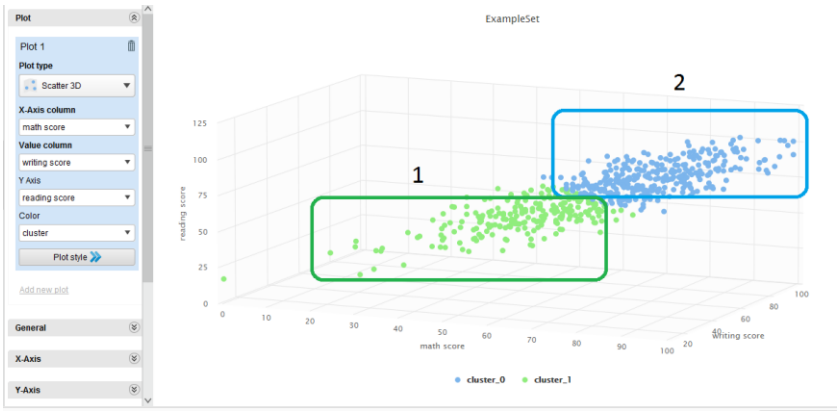
	K=5	K=4	K=3	K=2
DBI	-0.939	-0.889	-0.834	-0.792

Cluster yang baik adalah k=2, selanjutnya mengidentifikasi dua cluster dengan melihat Tabel Centroid.

Attribute	cluster_0	cluster_1
gender	0.442	0.593
race/ethnicity	3.305	2.925
parental level of education	2.526	2.292
test preparation course	0.449	0.217
math score	74.944	53.137
reading score	78.209	54.938
writing score	77.383	53.221

Gambar 7. Tabel Centroid pada RapidMiner

Dari gambar 7 Cluster_0 memiliki kinerja tinggi dalam ujian (warna biru) dan Cluster_1 memiliki kinerja rendah dalam ujian (warna merah). Hasil clustering juga dapat dilihat pada grafik berikut.



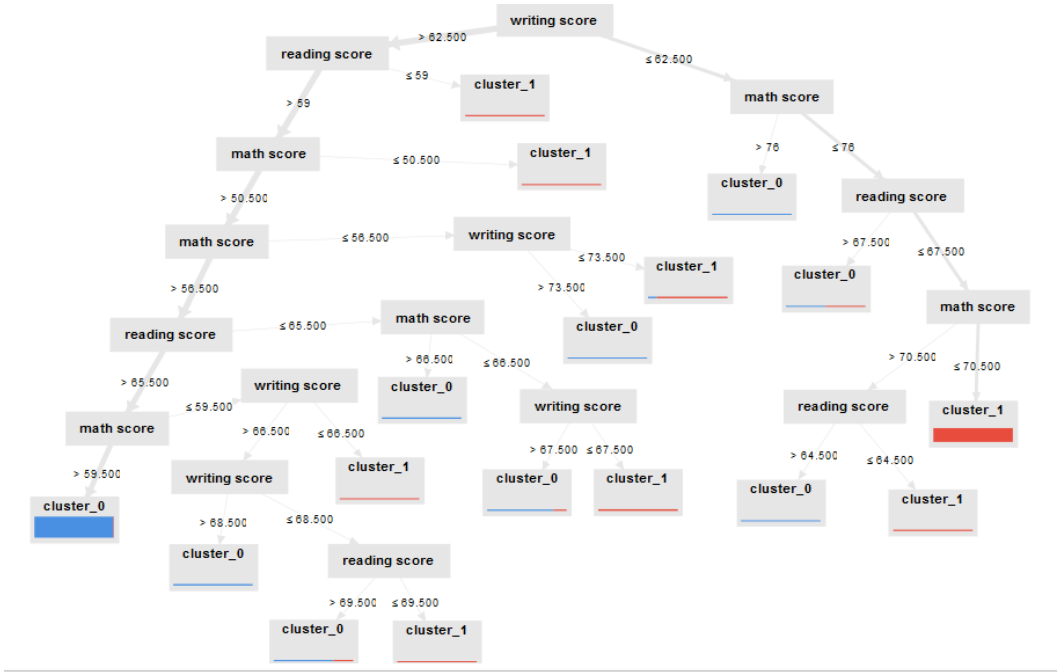
Gambar 8. Grafik Hasil Clustering

Gambar 8 menunjukkan grafik tiga dimensi yang menunjukkan hubungan antara atribut *math score, reading score* dan *writing score*. Warna menunjukkan cluster yang terbentuk, cluster_1 memiliki *math score, reading score* dan *writing score* rendah (warna hijau) sedangkan cluster_0 memiliki *math score, reading score* dan *writing score* tinggi (warna biru).

Evaluasi tujuan 2 berupa pola if-then (gambar 9) atau berupa pohon keputusan (gambar 10).

```
writing score > 62.500
|   reading score > 59
|   |   math score > 50.500
|   |   |   math score > 56.500
|   |   |   |   reading score > 65.500
|   |   |   |   |   math score > 59.500: cluster_0 (cluster_0=289, cluster_1=2)
|   |   |   |   |   math score ≤ 59.500
|   |   |   |   |   |   writing score > 66.500
|   |   |   |   |   |   |   writing score > 68.500: cluster_0 (cluster_0=8, cluster_1=0)
|   |   |   |   |   |   |   writing score ≤ 68.500
|   |   |   |   |   |   |   |   reading score > 69.500: cluster_0 (cluster_0=3, cluster_1=1)
|   |   |   |   |   |   |   |   reading score ≤ 69.500: cluster_1 (cluster_0=0, cluster_1=2)
|   |   |   |   |   |   writing score ≤ 66.500: cluster_1 (cluster_0=0, cluster_1=2)
|   |   |   |   reading score ≤ 65.500
|   |   |   |   |   math score > 66.500: cluster_0 (cluster_0=8, cluster_1=0)
|   |   |   |   |   math score ≤ 66.500
|   |   |   |   |   |   writing score > 67.500: cluster_0 (cluster_0=5, cluster_1=1)
|   |   |   |   |   |   writing score ≤ 67.500: cluster_1 (cluster_0=0, cluster_1=10)
|   |   |   math score ≤ 56.500
|   |   |   |   writing score > 73.500: cluster_0 (cluster_0=2, cluster_1=0)
|   |   |   |   writing score ≤ 73.500: cluster_1 (cluster_0=1, cluster_1=8)
|   |   math score ≤ 50.500: cluster_1 (cluster_0=0, cluster_1=3)
|   reading score ≤ 59: cluster_1 (cluster_0=0, cluster_1=5)
writing score ≤ 62.500
|   math score > 76: cluster_0 (cluster_0=2, cluster_1=0)
|   math score ≤ 76
|   |   reading score > 67.500: cluster_0 (cluster_0=1, cluster_1=1)
|   |   reading score ≤ 67.500
|   |   |   math score > 70.500
|   |   |   |   reading score > 64.500: cluster_0 (cluster_0=2, cluster_1=0)
|   |   |   |   reading score ≤ 64.500: cluster_1 (cluster_0=0, cluster_1=5)
|   |   |   math score ≤ 70.500: cluster_1 (cluster_0=0, cluster_1=186)
```

Gambar 9. Pola If-Then



Gambar 10. Pola Pohon Keputusan (Decision Tree)

Dari gambar 9 dan 10, atribut (faktor) yang paling berpengaruh adalah *writing score*, *reading score* dan *math score*, sedangkan atribut (faktor) yang tidak berpengaruh adalah *gender*, *race/ethnicity*, *parental level of education* dan *test preparation course*.

3.6. Deployment

Deployment dari evaluasi tujuan 1 yaitu karena kelompok cluster_1 memiliki kinerja rendah dalam ujian maka Kepala Sekolah SD Santa Maria harus melakukan peningkatan bimbingan belajar yang ketat pada siswa yang termasuk pada kelompok cluster_1.

Deployment dari evaluasi tujuan 2 yaitu Kepala Sekolah SD Santa Maria bisa menambah jam belajar dan membuka les khusus *writing score*, *reading score* dan *math score* untuk siswa yang akan ujian. Kepala Sekolah SD Santa Maria juga bisa menugaskan guru yang kompeten dibidang *writing score*, *reading score* dan *math score* untuk mengajar siswa yang akan ujian.

4. KESIMPULAN

Hasil dari implementasi teknik data mining dengan menggunakan aplikasi Rapidminer, dilakukan pengelompokan menggunakan metode clustering dengan algoritma K-Means yang menghasilkan k=2 sebagai kelompok *(cluster)* yang paling optimal dan mengelompokkan siswa menjadi 2 kelompok (*cluster* 0 dan *cluster* 1). Hasil cluster yang memuat cluster sebagai atribut baru ini lah yang digunakan untuk metode klasifikasi dengan algoritma *Decision Tree*. *Cluster* dijadikan sebagai label dan pola yang dihasilkan berbentuk *decision tree* (pohon keputusan) atau peraturan *if-then*.

DAFTAR PUSTAKA

[1] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, 2019, doi: 10.1109/TSMCC.2010.2053532.

[2] M. Sadikin, R. Rosnelly, and T. Surya Gunawan, "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1100–1109, 2020, doi: 10.30865/mib.v4i4.2434.

[3] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.

[4] A. Abu, "Educational Data Mining & Students' Performance Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 212–220, 2019, doi: 10.14569/ijacsa.2016.070531.

[5] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," *15th Eur. Concurr. Eng. Conf. 2008, ECEC 2008 - 5th Futur. Bus. Technol. Conf. FUBUTEC 2008*, vol. 2003, no. 2000, pp. 5–12, 2018.

[6] F. Widyahastuti and V. U. Tjhin, "Student Performance Prediction Using Data Mining Techniques," no. May, pp. 90–95, 2020, doi: 10.5220/0009017000900095.

[7] B. Kumar and S. Pal, "Mining Educational Data to Analyze Students Performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, 2019, doi: 10.14569/ijacsa.2011.020609.

[8] N. Z. Rahma and A. Setyono, "Implementation of C4 . 5 Algorithm to Predict the Ability of Junior High School Student to Face National Exam," vol. 8, no. 1, pp. 35–46, 2019.

[9] F. N. R. F. J. Aziz, B. D. Setiawan, and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 6, pp. 2243–2251, 2019.

[10] M. North, *Data Mining for the Masses*. 2012.