Terakreditasi SINTA Peringkat 4
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 200/M/KPT/2020
Masa berlaku mulai Volume 9 Nomor 1 Tahun 2019 sampai Volume 13 Nomor 2 Tahun 2023
Terbit online pada laman: https://ojs.stmikplk.ac.id/index.php/saintekom/index

SAIN TEKOM
ISSN 2088-1770 (Print)
ISSN 2503-3247 (Online)
Vol. 13 No.1 (2023) 22-31

JURNAL
SAINTEKOM
Sains, Teknologi, Komputer, dan Manajemen

Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes

Abdul Rahman
Informatika, Fakultas Teknik dan Komputer, Universitas Baturaja
Jl. Ratu Penghulu No.2301, Karang Sari, Tj. Baru, Kec. Baturaja Timur, Kab. Ogan Komering Ulu, Sumatera Selatan
Email: abdulrahman@ft.unbara.ac.id

ABSTRACT

Getting good academic performance is the goal of the learning process carried out by the education office under the auspices of the Ministry of Education and supervised by the government. Governments that want to be successful in educating students should pay attention to their new generation because they are the future successors of the nation. Students of all levels are the benchmark for a country's success. Therefore, it is necessary to know the student's academic performance from an early age, in order to get special treatment related to the student's learning achievement. In this study, the academic achievement of students from various levels of education such as elementary, middle, and high schools was tried to be determined by applying various data mining classification methods such as Decision Tree and Naive Bayes. This data grouping is open where the data can be accessed easily and can be used in future research. The data grouping is divided into 3 categories, namely Low (L), Medium (M) and High (H). From the results of the dataset trial, it shows that the highest classification accuracy is Decision Tree of 83.89% and Naive Bayes of 85.97%. Thus the Naïve Bayes method is more accurate in grouping students' academic performance data. The results of this study will be used to create a student grouping system based on student learning performance using the appropriate algorithm. So that his contribution later when creating the application system for the formation of study groups can use the Naïve Bayes method.

Keywords : academic; classification; data mining; decision tree; naive bayes

ABSTRAK

Mendapatkan performa akademik yang baik adalah tujuan dari proses pembelajaran yang diselenggarakan oleh dinas pendidikan dibawah naungan kementerian pendidikan dan diawasi oleh pemerintah. Pemerintah yang ingin sukses dalam mendidik siswa harus memperhatikan generasi barunya karena mereka adalah masa depan penerus bangsa. Peserta didik dari segala jenjang adalah tolak ukur kesuksesan sebuah negara. Oleh karena itu perlu untuk mengetahui performa akademik siswa sejak dini, agar bisa mendapatkan penanganan khusus terkait dengan prestasi belajar siswa tersebut. Dalam penelitian ini, prestasi akademik siswa yang berasal dari berbagai jenjang pendidikan seperti SD, SMP, dan SMA dicoba ditentukan dengan menerapkan berbagai metode klasifikasi *data mining* seperti Decision Tree dan Naive Bayes. Pengelompokan data ini bersifat terbuka dimana data bisa diakses dengan mudah dan dapat digunakan dalam penelitian selanjutnya. Pengelompokan data dibagi menjadi 3 kategori yaitu Rendah(L), Sedang(M) dan Tinggi(H). Dari hasil uji coba *dataset* menunjukkan bahwa akurasi hasil klasifikasi tertinggi adalah Decision Tree Sebesar 83.89% dan Naive Bayes sebesar 85.97%. Dengan demikian metode Naïve Bayes lebih akurat dalam mengelompokkan data-data performa akademik siswa. Hasil dari penelitian ini nantinya digunakan untuk membuat sistem pengelompokan mahasiswa berdasarkan performa belajar siswa dengan menggunakan algoritma yang sesuai. Sehingga kontribusi nya nanti pada saat pembuatan sistem aplikasi pembentukan kelompok belajar bisa menggunakan metode Naive Bayes.

Kata kunci *:* akademik; klasifikasi; *data mining*; *decision tree; naive bayes*

Diterima Redaksi: 30-11-2022 | Selesai Revisi: 7-3-2023 | Diterbitkan Online: 31-3-2023

1. PENDAHULUAN

Setiap siswa ingin sukses, begitu juga dengan pemerintah berusaha meningkatkan keberhasilan siswa untuk menjamin masa depan yang cerah bagi negara mereka. Jadi, peningkatan prestasi akademik siswa sangat penting dengan membawa sistem pendidikan ke tingkat yang lebih baik. Keberhasilan siswa berubah sesuai dengan beberapa kondisi. Jika kondisi ini dapat ditentukan, prestasi siswa dapat meningkat. Kondisi yang mempengaruhi keberhasilan siswa seperti "sikap siswa dan orang tuanya", "nilai siswa", "kesulitan mata pelajaran", dan lain-lain. Untuk meningkatkan prestasi akademik siswa, pertama-tama kita perlu menentukan faktor mana yang paling efektif atas keberhasilan akademik (Nida Uzel et al., 2018). Untuk tujuan ini, berbagai metode penambangan data telah diterapkan pada beberapa kumpulan data yang disiapkan sebelumnya. Saat ini, kumpulan data baru terus dibuat, dan teknik *data mining* diterapkan untuk mendapatkan hasil yang lebih baik.

Data mining adalah prosedur untuk mengekstraksi informasi yang berarti dari kumpulan data dengan menggunakan berbagai teknik pembelajaran mesin (Abdul Rahman et al., 2021). Baru-baru ini, *data mining* pendidikan yang merupakan bidang yang menggunakan teknik *data mining* untuk memprediksi kinerja siswa. *Data mining* pendidikan juga menentukan faktor-faktor yang mempengaruhi prestasi siswa, membuat kesimpulan dengan menggunakan faktor-faktor tersebut, dan memberikan perbaikan bagi sistem pendidikan (Rahman et al., 2019).

Dataset yang digunakan dalam penelitian ini adalah *dataset* pendidikan *Experience Application Programming Interface* (xAPI) yang dihasilkan dari sistem *e-learning* oleh (Aljarah, 2022). *Dataset* tersebut mencakup informasi tentang 480 siswa dari berbagai tingkatan seperti tingkat Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), dan Sekolah Menengah Atas (SMA) yang berisi 16 atribut. Atribut ini dibagi menjadi tiga kategori yaitu sebagai "demografi", "akademik", dan "informasi perilaku". 305 siswa laki-laki dan 175 siswa perempuan yang berasal dari 14 negara yang berbeda (misalnya Kuwait, Yordania, Irak dan lain-lain) ada di *dataset*. Menurut 16 atribut, penelitian sebelumnya (Nida Uzel et al., 2018)

DOI: https://doi.org/10.33020/saintekom.v13i1.349

mengklasifikasikan tingkat keberhasilan siswa menjadi tiga kategori sebagai "rendah", "sedang", atau "tinggi". Pada penelitian sebelumnya (Nida Uzel et al., 2018) menggunakan algoritma apriori. Namun, dalam penelitian ini kami menggunakan *naive bayes* dan *decision tree*.

Penelitian yang dilakukan oleh (Alturki et al., 2021) menjelaskan bahwa salah satu tujuan utama lembaga pendidikan tinggi adalah untuk memberikan pendidikan berkualitas tinggi kepada siswa mereka dan mengurangi angka putus sekolah. Hal ini dapat dicapai dengan memprediksi prestasi akademik siswa sejak dini menggunakan *Educational Data mining* (EDM) . Penelitian ini bertujuan untuk memprediksi nilai akhir siswa dan mengidentifikasi siswa honorer pada tahap awal. Penelitian EDM telah muncul sebagai bidang penelitian yang menarik, yang dapat membuka pengetahuan berharga dari *database* pendidikan untuk berbagai tujuan, seperti mengidentifikasi siswa putus sekolah dan siswa yang membutuhkan perhatian khusus dan menemukan siswa kehormatan untuk mengalokasikan beasiswa (Tanoli et al., 2021). Hasilnya adalah prediksi kinerja akademik dapat membantu guru dan siswa dalam banyak hal (Rahman et al., 2019). Hal ini juga memungkinkan penemuan awal siswa berprestasi (Rahman & Budiyanto, 2019). Dengan demikian, peluang yang memang layak dapat ditawarkan misalnya beasiswa, magang, dan workshop. Hal ini juga dapat membantu mengidentifikasi siswa yang memerlukan perhatian khusus untuk mengambil intervensi yang tepat pada tahap sedini mungkin. Selain itu, instruktur dapat mengetahui kemampuan setiap siswa dan menyesuaikan tugas mengajar berdasarkan kebutuhan siswa (Alturki et al., 2021).

2. METODE

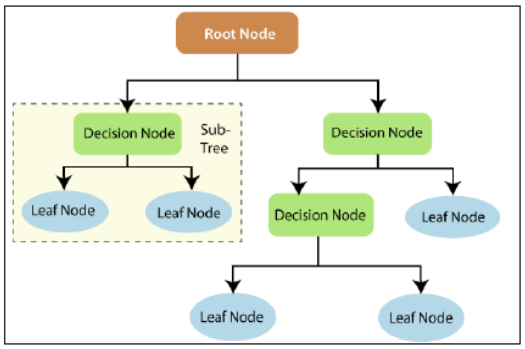
Penelitian klasifikasi performa akademik siswa menggunakan metode decision tree dan naive bayes ini menggunakan dua buah metode yang sudah populer dikalangan para peneliti yaitu metode klasifikasi *Decision Tree* (DT) dan *naive bayes*.

2.1 Decision Tree

Menurut (Charbuty & Abdulazeez, 2021) Algoritma klasifikasi dalam pembelajaran mesin berisi beberapa algoritma, dan dalam artikel ini difokuskan pada algoritma DT secara

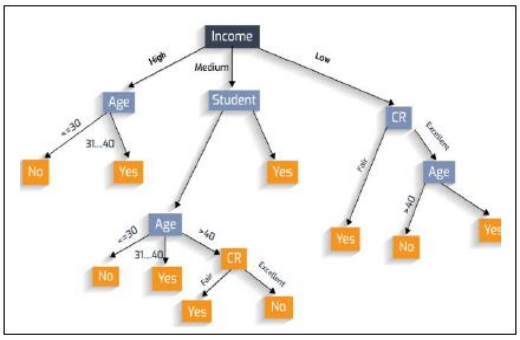
DOI: https://doi.org/10.33020/saintekom.v13i1.349

umum seperti pada Gambar 1 mengilustrasikan struktur DT.



Gambar 1. Ilustrasi struktur *decision tree* (Charbuty & Abdulazeez, 2021)

Gambar 1 menjelaskan bahwa DT adalah salah satu metode ampuh yang biasa digunakan diberbagai bidang, seperti pembelajaran mesin, pemrosesan gambar, dan identifikasi pola. DT adalah model berurutan yang menyatukan serangkaian tes dasar secara efisien dan kohesif dimana fitur numerik dibandingkan dengan nilai ambang batas disetiap pengujian (Charbuty & Abdulazeez, 2021) (Tangirala, 2020) (Karacan et al., 2020) (Jena & Dehuri, 2020). Aturan konseptual jauh lebih mudah untuk dibangun dari pada bobot numerik dalam jaringan saraf koneksi antar node (Charbuty & Abdulazeez, 2021). Selain itu, DT adalah model klasifikasi yang biasanya digunakan dalam *Data mining* (Zhou et al., 2021). *Node* dan cabang terdiri dari setiap pohon. Setiap *node* mewakili fitur dalam kategori yang akan diklasifikasikan dan setiap subset mendefinisikan nilai yang dapat diambil oleh *node* (Zhou et al., 2021). Karena analisisnya yang sederhana (Asruddin et al., 2020) dan presisinya pada berbagai bentuk data, DT telah menemukan banyak bidang implementasi (Zhou et al., 2021), Gambar 2 menunjukkan contoh DT dengan menggunakan algoritma C4.5.



Gambar 2. Contoh *Decision Tree*

Gambar 2 menjelaskan bahwa terlihat contoh dari model DT dengan algoritma C4.5 yang digambarkan dalam bentuk cabang dari masing-masing atribut yaitu Age, Student, CR. Dari masing-masing statement akan menghasilkan algoritma baru sesuai data yang kita inginkan.

2.2 Naïve Bayes

Metode klasifikasi *naive bayes* adalah salah satu algoritma dalam teknologi klasifikasi yang mudah diimplementasikan dan cepat dalam kecepatan pemrosesan (Nugroho et al.,

DOI: https://doi.org/10.33020/saintekom.v13i1.349

2021). Metode *Naïve Bayes* menggunakan model statistik untuk melakukan proses klasifikasi data. Metode ini menghitung nilai probabilitas data uji berdasarkan data kasus yang sudah pernah terjadi (Nugroho et al., 2021). Berikut adalah simulasi perhitungan pada probabilitas *naïve bayes* yang bisa dilihat pada Persamaan 1 berikut:

$$\boldsymbol{P(A|B) = \frac{P(B|A)P(B)}{P(A)}} \qquad (1)$$

P (A|B) adalah peluang A jika diketahui keadaan B, P(B|A) adalah peluang *evidence* B jika diketahui hipotesis A, P(B) adalah probabilitas B tanpa melihat *evidence* apapun, P(A) adalah peluang *evidence* A.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset dari (Aljarah, 2022) dengan statistik pada Tabel 3 sebagai berikut :

Table 3. Dataset Xapi-Edu-Data

Gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	…	Student Absence Days	Class
M	KW	KuwaIT	MiddleSchool	G-07	A	Math	…	Under-7	H
M	KW	KuwaIT	MiddleSchool	G-07	B	Math	…	Under-7	M
M	KW	KuwaIT	lowerlevel	G-04	A	IT	…	Above-7	L
M	lebanon	lebanon	MiddleSchool	G-08	A	Math	…	Above-7	L
F	KW	KuwaIT	MiddleSchool	G-08	A	Math	…	Above-7	H
F	KW	KuwaIT	MiddleSchool	G-06	A	IT	…	Under-7	M
M	KW	KuwaIT	MiddleSchool	G-07	B	IT	…	Above-7	M
M	KW	KuwaIT	MiddleSchool	G-07	A	Math	…	Above-7	M
F	KW	KuwaIT	MiddleSchool	G-07	A	IT	…	Under-7	M
M	KW	KuwaIT	MiddleSchool	G-07	B	IT	…	Under-7	H
F	KW	KuwaIT	MiddleSchool	G-07	A	IT	…	Above-7	M
F	KW	KuwaIT	MiddleSchool	G-07	B	IT	…	Under-7	M
M	KW	KuwaIT	MiddleSchool	G-07	A	IT	…	Under-7	M
M	KW	KuwaIT	MiddleSchool	G-07	A	IT	…	Above-7	L
M	KW	KuwaIT	MiddleSchool	G-07	B	IT	…	Above-7	L
M	KW	KuwaIT	MiddleSchool	G-07	A	IT	…	Above-7	L
M	KW	KuwaIT	MiddleSchool	G-07	B	IT	…	Under-7	M
M	KW	KuwaIT	MiddleSchool	G-08	A	Arabic	…	Above-7	L
M	KW	KuwaIT	MiddleSchool	G-08	A	Science	…	Under-7	M
…	…	…	…	…	…	…	…	…	…
F	Jordan	Jordan	MiddleSchool	G-08	A	History	…	Above-7	L
F	Jordan	Jordan	MiddleSchool	G-08	A	History	…	Above-7	L

(Sumber: Aljarah, 2022)

DOI: https://doi.org/10.33020/saintekom.v13i1.349

Kumpulan data pendidikan yang dikumpulkan dari *Learning Management System* (LMS) yang disebut Kalboard 360. Kalboard 360 adalah LMS multi-agen, yang telah dirancang untuk memfasilitasi pembelajaran melalui penggunaan teknologi terdepan. Sistem tersebut memberi pengguna akses sinkron ke sumber daya pendidikan dari perangkat apa pun dengan koneksi Internet. Data dikumpulkan menggunakan alat pelacak aktivitas pelajar yang disebut xAPI.

xAPI adalah komponen *Total Learning Architecture* (TLA) yang memungkinkan untuk memantau kemajuan pembelajaran dan tindakan pelajar seperti membaca artikel atau menonton video pelatihan. xAPI membantu penyedia aktivitas pembelajaran untuk menentukan pelajar, aktivitas, dan objek yang menggambarkan pengalaman belajar.

Dataset terdiri dari 480 catatan siswa dan 16 fitur. Ciri-ciri tersebut diklasifikasikan ke dalam tiga kategori utama: (1) Ciri-ciri demografis seperti jenis kelamin dan kebangsaan. (2) Fitur latar belakang akademik seperti tahap pendidikan, Tingkat kelas dan bagian. (3) Fitur perilaku seperti mengangkat tangan di kelas, membuka sumber, menjawab survei oleh orang tua, dan kepuasan sekolah.

Dataset terdiri dari 305 laki-laki dan 175 perempuan. Mahasiswa tersebut berasal dari berbagai daerah seperti 179 siswa dari Kuwait, 172 mahasiswa dari Yordania, 28 mahasiswa dari Palestina, 22 mahasiswa dari Irak, 17 mahasiswa dari Lebanon, 12 mahasiswa dari Tunis, 11 mahasiswa dari Arab Saudi, sembilan mahasiswa dari Mesir, tujuh mahasiswa dari Syria, enam mahasiswa dari USA, Iran dan Libya, empat mahasiswa dari Maroko dan satu mahasiswa dari Venezuela.

Dataset dikumpulkan melalui dua semester pendidikan: 245 catatan siswa dikumpulkan selama semester pertama dan 235 catatan siswa dikumpulkan selama semester kedua. Kumpulan data juga mencakup fitur kehadiran sekolah seperti siswa diklasifikasikan ke dalam dua kategori berdasarkan hari ketidakhadiran mereka 191 siswa melebihi 7 hari ketidakhadiran dan 289 siswa dengan hari ketidakhadiran di bawah 7.

Dataset ini juga mencakup kategori fitur baru, fitur ini adalah proses persalinan orang tua dalam proses

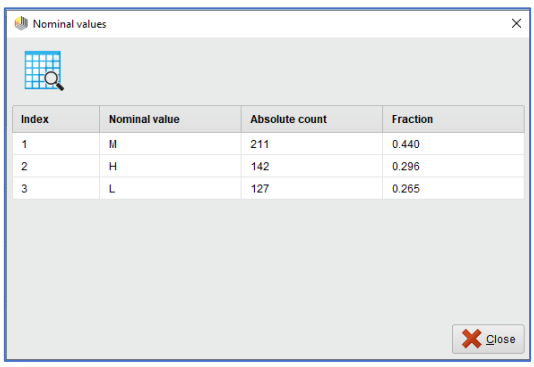
DOI: https://doi.org/10.33020/saintekom.v13i1.349

pendidikan. Fitur partisipasi orang tua memiliki dua sub fitur yaitu survei jawaban orang tua dan kepuasan sekolah orang tua. Ada 270 orang tua menjawab survei dan 210 tidak, 292 orang tua puas dari sekolah dan 188 tidak.

Terdapat *dataset* 480 siswa yang ada pada *dataset* diambil 50 data yang digunakan sebagai data latihan, dimana data ini nanti akan dijadikan patokan seberapa pintar metode klasifikasi dalam mengelompokkan data yang ada.

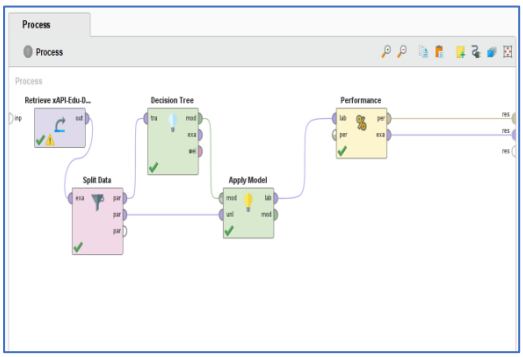
3.1 Metode *Decision Tree*

Langkah awal dalam menggunakan metode ini kita harus memberikan label dari masing-masing data yang ada, disini terbagi menjadi 3 buah kelas data yaitu Rendah (L), Sedang (M) dan Tinggi (H), seperti pada ditampilkan pada Gambar 4.



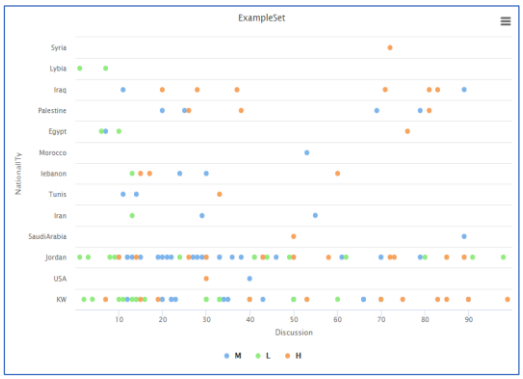
Gambar 3. Detail kelas *dataset*

Setelah data mempunyai label yang jelas baru kita akan mengelompokan data berdasarkan dengan kelas. Untuk membuat data bisa terbagi menjadi *dataset* dan data *testing* dalam penelitian ini menggunakan operator split pada Rapid Miner, dengan ketentuan data *training* sebesar 70%, dan data *testing* 30%. Hasilnya terlihat pada Gambar 5.



Gambar 4. Desain Process Klasifikasi Data Menggunakan Metode *Decision Tree*

Hasil akurasi klasifikasi data *performance* akademik siswa adalah sebesar 83.89%. Berikut adalah tampilan sebaran kelompok dalam berdiskusi berdasarkan negara bagian seperti pada Gambar 6.

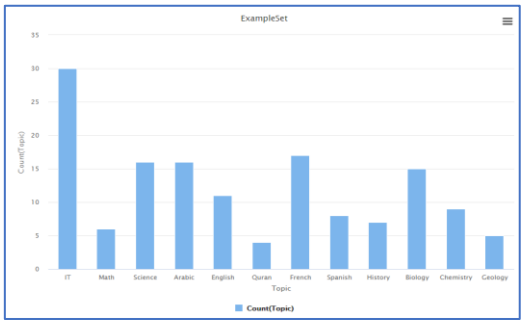


Gambar 5. Keaktifan dalam berdiskusi berdasarkan negara bagian

Dalam hal peminatan dalam memilih mata pelajaran yang disukai, mata pelajaran *Information Technology*

DOI: https://doi.org/10.33020/saintekom.v13i1.349

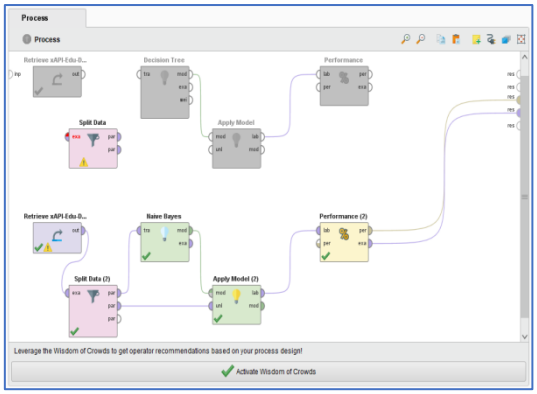
(IT) menduduki peringkat pertama. Seperti pada Gambar 7 berikut.



Gambar 6. Topik Belajar Yang Paling Diminati

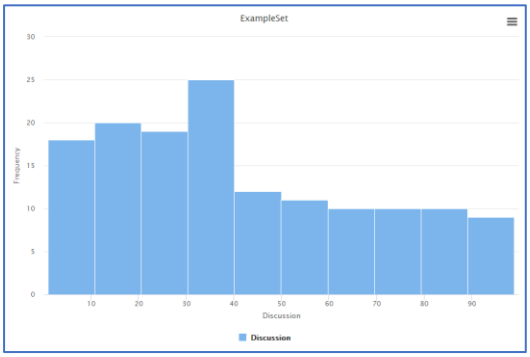
3.2 Metode Naïve Bayes

Setelah dilakukan analisa perhitungan dengan menggunakan tools Rapid Miner dengan data yang sama dengan model desain yang sama *naïve bayes* tingkat akurasi nya sebesar 85.97%. Seperti pada desain proses Gambar 8 berikut :



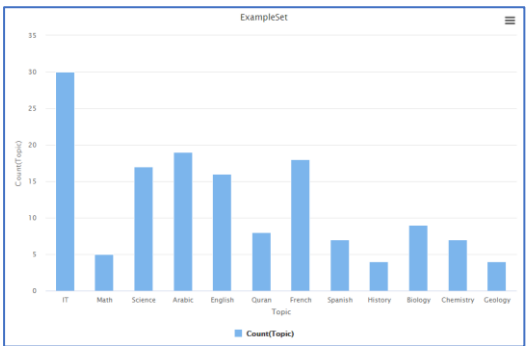
Gambar 7. Desain Proses Klasifikasi Dengan Metode *Naive Bayes*

Untuk sebaran kelompok dalam berdiskusi berdasarkan negara bagian seperti Gambar 9.



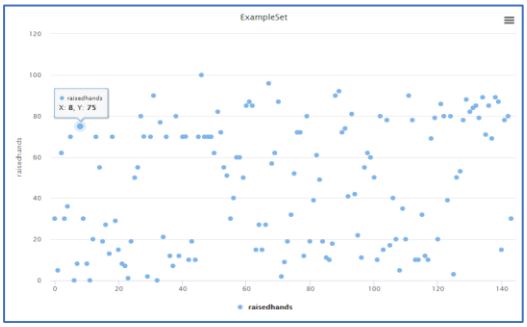
Gambar 8. Kelompok Diskusi Dengan *Naive Bayes*

Pada data yang terkategori peminatan dalam memilih mata pelajaran yang disukai, mata pelajaran IT menduduki peringkat pertama seperti Gambar 10 berikut.



Gambar 9. Topik Belajar Yang Paling Diminati Dengan Metode *Naive Bayes*

Dalam hal keaktifan bertanya *naïve bayes* juga lebih unggul seperti Gambar 11.



Gambar 10. Keaktifan Dalam Bertanya Dan Berdiskusi

DOI: https://doi.org/10.33020/saintekom.v13i1.349

4. KESIMPULAN

Setelah dilakukan uji data dari *dataset* yang digunakan dalam proses perbandingan metode *data mining* dan menggunakan *dataset* dari Kaggle metode dengan akurasi tertinggi adalah *naïve bayes* dengan nilai 85.97%, untuk metode DT nilai akurasinya sebesar 83.89%. Data presentasi hasil diketahui bahwa Naïve Bayes mendapatkan hasil yang lebih baik dibandingkan dengan Decision Tree, ini berarti bahwa untuk pembuatan sistem aplikasi pengelompokan belajar siswa dapat menggunakan algoritma Naïve Bayes agar mendapatkan akurasi hasil yang maksimal.

Setelah dilakukan uji coba dataset dengan menggunakan metode data mining diatas, diharapkan pada penelitian selanjutnya bisa menggunakan metode *data mining* yang lainnya, sehingga nilai akurasi nya bisa lebih tinggi.

DAFTAR PUSTAKA

Abdul Rahman, Destiarini, & Kuswanto, J. (2021). Fuzzy Logic Recommended Student Learning Levels. *Jurnal Informatika Polinema*, *7*(2). https://doi.org/10.33795/jip.v7i2.531

Aljarah, I. (2022). *Students' Academic Performance Dataset*. Kaggle. Retrieved April 15, 2022, from https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

Alturki, S., Alturki, N., & Stuckenschmidt, H. (2021). Using Educational Data Mining To Predict Students' Academic Performance For Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, *20*. https://doi.org/10.28945/4835

Asruddin, Rahman, A., & Rambe, J. K. (2020). Analisa SWOT Pengembangan Media Belajar Sejarah Di Sekolah Menengah Pertama Kelas IX Semester Ganjil. *INTECH*, *1*(1), 8–16.

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01). https://doi.org/10.38094/jastt20165

Hameed, I. A. (2016). A simplified implementation of interval type-2 fuzzy system and its application in students' academic evaluation. *2016 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2019*, 650–656. https://doi.org/10.1109/FUZZ-IEEE.2016.7737748

Jena, M., & Dehuri, S. (2020). Decision tree for classification and regression: A state-of-the art review. In *Informatica (Slovenia)* (Vol. 44, Issue 4).

DOI: https://doi.org/10.33020/saintekom.v13i1.349

https://doi.org/10.31449/INF.V44I4.3023

Karacan, I., Sennaroglu, B., & Vayvay, O. (2020). Analysis of life expectancy across countries using a decision tree. *Eastern Mediterranean Health Journal*, *26*(2). https://doi.org/10.26719/2020.26.2.143

Nida Uzel, V., Sevgi Turgut, S., & Ayşe Özel, S. (2018). Prediction of Students' Academic Success Using Data Mining Methods. *Proceedings - 2018 Innovations in Intelligent Systems and Applications Conference, ASYU 2018*. https://doi.org/10.1109/ASYU.2018.8554006

Nugroho, F. A., Solikin, A. F., Anggraini, M. D., & Kusrini, K. (2021). Sistem Pakar Diagnosa Virus Corona Dengan Metode Naïve Bayes. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *9*(1). https://doi.org/10.30646/tikomsin.v9i1.553

Rahman, A., & Budiyanto, U. (2019). Case based reasoning adaptive e-learning system based on visual-auditory-kinesthetic learning styles. In Irawan, H. Irawan, M. A. Riyadi, & M. Facta (Eds.), *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 177–182). IEEE. https://doi.org/10.23919/EECSI48112.2019.8976921

Rahman, A., Mutiarawan, R. A., Darmawan, A., Rianto, Y., & Syafrullah, M. (2019). Prediction of students academic success using case based reasoning. In Irawan, H. Irawan, M. A. Riyadi, & M. Facta (Eds.), *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 171–176). IEEE. https://doi.org/10.23919/EECSI48112.2019.8977104

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. *International Journal of Advanced Computer Science and Applications*, *2*. https://doi.org/10.14569/ijacsa.2020.0110277

Tanoli, Z., Seemab, U., Scherer, A., Wennerberg, K., Tang, J., & Vähä-Koskela, M. (2021). Exploration of databases and methods supporting drug repurposing: A comprehensive survey. In *Briefings in Bioinformatics* (Vol. 22, Issue 2). https://doi.org/10.1093/bib/bbaa003

Zhou, H. F., Zhang, J. W., Zhou, Y. Q., Guo, X. J., & Ma, Y. M. (2021). A feature selection algorithm of decision tree based on feature weight. *Expert Systems with Applications*, *164*. https://doi.org/10.1016/j.eswa.2020.113842

DOI: https://doi.org/10.33020/saintekom.v13i1.349