MATHunesa
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115

Volume 13 No 02
Tahun 2025

PENERAPAN *K-NEAREST NEIGHBORS (K-NN)* DALAM MENILAI POTENSI *DROP OUT* MAHASISWA: STUDI PADA ASPEK AKADEMIK, SOSIAL, DAN EKONOMI

Vanya Tania
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: vanya.23067@mhs.unesa.ac.id*

Intan Nur Khasanah
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: intan.23091@mhs.unesa.ac.id

Syah Albani
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: syah.23012@mhs.unesa.ac.id

Nur Azizah
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: nur.23059@mhs.unesa.ac.id

Nurul Jihan
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: nurul.23046@mhs.unesa.ac.id

Abstrak

Perguruan tinggi merupakan lembaga pendidikan tertinggi yang berperan penting dalam mempersiapkan mahasiswa untuk menghadapi tantangan global. Namun tidak sedikit mahasiswa yang mengalami kegagalan akademik dalam menyelesaikan perkuliahannya (*drop out*). Tingginya tingkat *drop out* dapat diminimalisasikan dengan pengambilan keputusan yang tepat dalam upaya mencegah mahasiswa *drop out*. Untuk membantu pengambilan keputusan tersebut, dibutuhkan suatu teknologi seperti *Educational Data Mining (EDM)* yang ditujukan untuk lebih memahami pola data mahasiswa. Salah satu metode dalam *EDM* yang dapat digunakan adalah klasifikasi data dengan konsep *K-Nearest Neighbor (K-NN)*. Dalam penelitian ini, K-NN bekerja dengan menganalisis kedekatan data-data mahasiswa untuk memprediksi kemungkinan seorang mahasiswa lulus tepat waktu atau berisiko *drop out*. Data yang digunakan mencangkup aspek kehidupan mahasiswa, seperti aspek akademik, sosial, dan ekonomi. Berdasarkan hasil analisis maka dapat disimpulkan penggunaan K-NN mempunyai kinerja yang baik dan cukup akurat namun dapat ditingkatkan lagi dengan menambahkan teknik resampling untuk mengatasi ketidakseimbangan kelas yang mungkin dapat terjadi dari pengaruh banyaknya data. Namun, pendekatan K-NN dapat membantu perguruan tinggi merancang kebijakan karena memiliki kesederhanaan dalam pengimplementasiannya, sehingga mudah untuk diterapkan dalam bidang akademik.
Kata Kunci: *Educational Data Mining (EDM*), *K-Nearest Neighbors (KNN), Drop Out,* Perguruan Tinggi.

Abstract

College is the highest educational institution that is important in preparing students to face global challenges. However, not a few students experience academic failure in completing their lectures (drop out). The high dropout rate can be minimized by making the right decisions to prevent students from dropping out. To help make these decisions, a technology such as Educational Data Mining (EDM) is needed to better understand student data patterns. One method in EDM that can be used is data classification with the concept of K-Nearest Neighbor (K-NN). In this research, K-NN works by analyzing the closeness of student data to predict the likelihood of a student graduating on time or at risk of dropout. The data used covers aspects of student life, such as academic, social, and economic aspects. Based on the results of the analysis, it can be concluded that the use of K-NN has good performance and is quite accurate but can be improved by adding resampling techniques to overcome class imbalances that may occur from the influence of large amounts of data. However, the K-NN approach can help universities design policies because it is simple to implement, making it easy to apply in the academic field.
Keywords: *Educational Data Mining (EDM), K-Nearest Neighbors (KNN), Drop Out, College.*

PENDAHULUAN

Perguruan tinggi merupakan jenjang terakhir pendidikan formal bagi mahasiswa sebelum memasuki dunia kerja. Tingkat persaingan antar perguruan tinggi juga semakin kompetitif, seiring dengan tuntutan kualitas yang diharapkan oleh masyarakat. Dalam upaya memenuhi ekspektasi tersebut, perguruan tinggi dituntut untuk tidak hanya menyediakan pendidikan yang bermutu, tetapi juga menghasilkan lulusan yang kompeten dan siap bersaing di dunia kerja. Salah satu indikator yang mempengaruhi kualitas sebuah perguruan tinggi adalah akreditas. Akreditas mencerminkan keberhasilan lulusan dari perguruan tinggi tersebut dalam meraih kesuksesan di dunia kerja maupun di masyarakat. Namun terdapat sebuah kasus dimana tidak sedikit mahasiswa yang mengalami kegagalan akademik dalam menyelesaikan perkuliahannya (*drop out*). Kegagalan ini menjadi salah satu faktor utama penurunan akreditas perguruan tinggi.

Tingginya tingkat *drop out* dapat diminimalisasikan dengan pengambilan keputusan yang tepat dalam upaya mencegah mahasiswa *drop out*. Keputusan yang tepat harus melalui proses analisis yang mendalaman terhadap faktor-faktor apa saja yang mempengaruhi, seperti riwayat pendidikan, kondisi ekonomi, hingga keadaan lingkungan sosial tempat tinggal mereka. Tidak berhenti disitu saja, *drop out* juga berdampak negatif bagi keberlangsungan hidup mahasiswa, seperti hilangnya peluang kerja, ketidakstabilan finansial, dan kesulitan beradaptasi di lingkungan masyarakat. Oleh karena itu, dengan meminimalisasikan angka *drop out* bukan hanya untuk menjaga akreditas perguruan tinggi, tetapi juga untuk menjaga dan memastikan masa depan mahasiswa tersebut terjamin.

Dalam hal ini, berbagai solusi inovatif diberikan untuk meminimalisasikan angka *drop out*. Salah satunya dengan memanfaatkan *Educational Data Mining (EDM)* yang ditujukan untuk lebih memahami pola data mahasiswa. *EDM* dapat diimplementasikan dalam perguruan tinggi untuk memprediksi apakah mahasiswa tersebut beresiko *drop out* ataupun mengalami kesulitan dalam menyelesaikan perkuliahannya. Dengan melakukan teknik klasifikasi, data yang telah ada digolongkan menjadi dua macam data, yaitu data *training* dan data *testing*. Data *training* berfungsi untuk melatih model *machine learning* dalam mengenali pola data dan hubungan antara variabel bebas dan variabel terikat. Dalam pemrosesan data, *machine learning* menganalisis seluruh data yang menjadi faktor penyebab suatu hasil tertentu, sehingga menemukan hubungan pada setiap pola data. Setelah *machine learning* terlatih dengan data training, maka akan diuji dengan data testing yang berfungsi untuk mengukur sejauh mana kemampuan *machine learning* dalam memahami dan memprediksi data yang belum diproses sebelumnya. Adanya data *testing* untuk mengevaluasi performa *machine learning* dalam mengolah data yang telah dilatih dengan data training.

Dalam proses pengolahan data, dibutuhkan landasan algoritma dalam menentukan langkah-langkah agar pemrosesan data diolah secara sistematis dan menghasilkan informasi yang relevan. Pada penelitian ini, algoritma yang digunakan adalah *K-Nearest Neighbor (K-NN)*. Prinsip kerja K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga *(neighbor)* terdekatnya dalam data training. Algoritma K-NN digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma K-NN memiliki kesederhanaan dalam pengimplementasiannya, sehingga mudah untuk diterapkan dalam bidang akademik. Selain itu, proses pengolahan data dengan algoritma ini dapat mudah menangkap pola dalam semua jenis data. K-NN tidak memerlukan proses pelatihan model yang kompleks, semua perhitungan dilakukan pada saat prediksi, sehingga tidak memerlukan pemodelan statistik yang rumit.

Kajian Teori

Klasifikasi adalah salah satu metode dalam *machine learning* yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan fitur-fitur yang dimiliki. Dalam klasifikasi menggunakan dataset untuk membangun hubungan antara fitur dengan label. Beberapa algoritma yang digunakan dalam klasifikasi antara lain *Decision Tree, Support Vector Machine (SVM), Naive Bayes, dan K-Nearest Neighbors (K-NN).*

K-Nearest Neighbors (K-NN) adalah salah satu metode klasifikasi yang bekerja dengan mencari

tetangga terdekat dari data yang ingin diklasifikasikan. Langkah- langkah dalam algoritma K-NN:

1. Menentukan nilai K yaitu jumlah tetangga terdekat yang akan dipertimbangkan.
2. Menghitung jarak antara data uji dengan semua data dataset menggunakan matrik.
3. Memilih K tetangga terdekat berdasarkan nilai jarak terkecil.
4. Melakukan *voting* terhadap kelas dari tetangga.
5. Menentukan kelas mayoritas sebagai hasil prediksi.

KNN memiliki kelebihan memudahkan dalam implementasi untuk menangani data nonlinear. Namun, kelemahannya adalah kompleksitas komputasi yang tinggi ketika jumlah data sangat besar serta sensitivitas terhadap skala data dan pemilihan nilai K yang optimal.

Prediksi Mahasiswa *Drop Out*

Prediksi Mahasiswa *Drop Out* adalah suatu permasalahan dalam dunia pendidikan, di mana mahasiswa tidak menyelesaikan pendidikannya sesuai dengan waktu yang telah ditentukan. Penyebab utama dari *drop out* bisa berasal dari beberapa faktor, antara lain:

- Faktor akademik
  Nilai rendah, kesulitan dalam memahami materi, serta ketidakhadiran dalam perkuliahan.
- Faktor ekonomi
  Kesulitan dalam finansial contohnya dalam membayar biaya pendidikan.
- Faktor sosial
  Kurangnya dukungan dari lingkungan keluarga, teman, ataupun sekitar.
- Faktor psikologis
  Depresi, rendahnya motivasi dalam belajar atau bahkan masalah dalam kesehatan mental.

Dengan menggunakan algoritma K-NN, prediksi mahasiswa yang berisiko *drop out* dapat dilakukan dengan menganalisis data riwayat mahasiswa, seperti nilai akademik, kehadiran, dan kondisi sosial ekonomi. Model ini akan membantu pihak universitas dalam mengidentifikasi mahasiswa yang berisiko *drop out*.

Penerapan KNN dalam Prediksi *Drop Out* dalam implementasinya mempunyai langkah-langkah untuk memprediksi drop out yaitu:

1. Pengumpulan data
   Data mahasiswa akan dikumpulkan, termasuk nilai akademik, riwayat kehadiran, kondisi finansial, dan faktor lainnya.
2. *Preprocessing* data
   Melakukan standarisasi data agar hasil perhitungan jarak lebih akurat.
3. Pemilihan fitur
   Menentukan fitur yang relevan untuk meningkatkan akurasi model.
4. Penentuan parameter K
   Menentukan nilai K yang optimal dengan menggunakan metode validasi silang.
5. Evaluasi model
   Menggunakan matriks seperti akurasi, presisi, *recall,* dan *F1 -score* untuk menilai performa model.

METODE

Metode yang digunakan pada penelitian ini adalah klasifikasi dengan algoritma K-NN. Metode tersebut merupakan metode *machine learning* yang mengelompokkan data berdasarkan jarak data baru kedalam beberapa data tetangga *(neighbor)* terdekat. Algoritma K-NN mempunyai beberapa langkah, yaitu :

1. Mengunduh dataset
2. Melakukan *preprocessing* data seperti pengecekan data apakah ada *missing value*, data yang kosong, data yang duplikat, mengubah target yang *multi classifications* menjadi *binary classification* dan harus dalam bentuk numerik agar memudahkan proses pengolahan data.
3. Membagi data menjadi data latih dan data uji dengan perbandingan 80 : 20.
4. Melatih model K-NN dengan melatih data x *training* dan y *training*.
5. Menentukan nilai K menggunakan metode *cross validation* dengan rumus perhitungan manual sebagai berikut:

$$CV\ (k)\ = \frac{1}{k}\sum_{i=0}^{k} a\ (i)$$

6. Melakukan evaluasi.

Untuk mengevaluasi kinerja model digunakan metrik, seperti *precision, recall, F1-score, dan confusion matrix*. Precision digunakan untuk mengukur akurasi prediksi positif, hal ini sangat berguna pada saat kesalahan positif yang berisiko tinggi. *Recall* digunakan untuk menilai kemampuan model dalam mendeteksi seluruh kasus positif, terutama ketika prioritasnya adalah menangkap semua kasus positif. *F1-Score* digunakan untuk memberi keseimbangan antara *precision* dan *recall* pada saat data tidak seimbang. Sementara itu, *confusion matrix* digunakan untuk menampilkan jumlah prediksi benar dan salah pada setiap kelas, membantu menganalisis kekuatan serta kelemahan model. Dengan mengikuti langkah-langkah tersebut, algoritma K-NN dapat digunakan untuk mengklasifikasikan mahasiswa yang lulus tepat waktu atau *drop out* berdasarkan pengaruhnya.

HASIL DAN PEMBAHASAN

Pada penelitian ini, data yang digunakan adalah *Predict Students' Dropout and Academic Success*, dengan rincian data sebagai berikut :

Fitur	Tipe
Marital Status	Integer
Application mode	Integer
Application order	Integer
Course	Integer
Daytime/evening attendance	Integer
Previous qualification	Integer
Previous qualification (grade)	Continuous
Nacionality	Integer
Mother's qualification	Integer
Father's qualification	Integer
Mother's occupation	Integer
Father's occupation	Integer
Admission grade	Continuous
Displaced	Integer
Educational special needs	Integer
Debtor	Integer
Tuition fees up to date	Integer
Gender	Integer
Scholarship holder	Integer
Age at enrollment	Integer
International	Integer
Curricular units 1st sem (credited)	Integer
Curricular units 1st sem (enrolled)	Integer
Curricular units 1st sem (evaluations)	Integer
Curricular units 1st sem (approved)	Integer
Curricular units 1st sem (grade)	Integer
Curricular units 1st sem (without evaluations)	Integer
Curricular units 2nd sem (credited)	Integer
Curricular units 2nd sem (evaluations)	Integer
Curricular units 2nd sem (approved)	Integer
Curricular units 2nd sem (grade)	Integer
Curricular units 2nd sem (without evaluations)	Integer
Unemployment rate	Continuous
Inflation rate	Continuous
GDP	Continuous

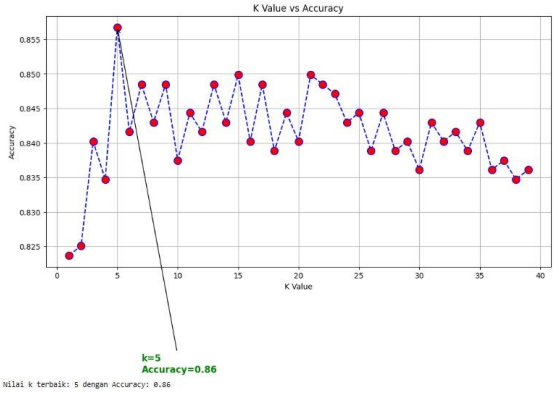
Gambar 1. Tabel Fitur Tipe Data *Predict Students' Dropout and Academic Success*

Dengan spesifikasi target sebagai berikut,
Drop Out : 1421
Graduate : 2209
Enrolled : 794

Data diambil dari
https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success

Berdasarkan analisis data, diperlukan melakukan beberapa langkah sebelum diolah menggunakan algoritma K-NN. Pada tahap *preprocessing* data, tipe target yang berbentuk kategori harus diubah menjadi bentuk integer agar mempermudah perhitungan jarak antar data. Langkah selanjutnya adalah data klasifikasi *multiclass* yang mempunyai tiga target (*drop out, graduate, enrolled*) harus diubah menjadi dua target (*drop out* dan *non-drop out*) agar model lebih fokus dalam memprediksi mahasiswa yang berpotensi *drop out* dan hasil yang didapatkan lebih akurat. Selain itu diperlukan standarisasi data pada fitur agar seluruh data memiliki jarak yang seimbang dan akurat.

Pada tahap pengujian model klasifikasi, algoritma K-NN digunakan untuk memprediksi apakah seorang mahasiswa akan mengalami *drop out* atau *non-drop out*. Namun algoritma K-NN memiliki kelemahan dalam menentukan nilai K dengan dataset yang memiliki banyak fitur dan data, seperti pada penelitian ini. Oleh karena itu, penulis menggunakan *cross validation* dalam memilih nilai K terbaik, penulis melakukan pengujian dengan nilai K antara 1 − 40 hingga diperoleh nilai K terbaik. Pemilihan K sangat diperhatikan agar tidak mengalami *overfitting* yaitu nilai K yang terlalu kecil menyebabkan model terlalu spesifik terhadap data *training* dan *underfitting* yaitu nilai K terlalu besar hingga mendekati jumlah data latih.



Gambar 2. Grafik Hubungan antara Nilai K dan Akurasi Model

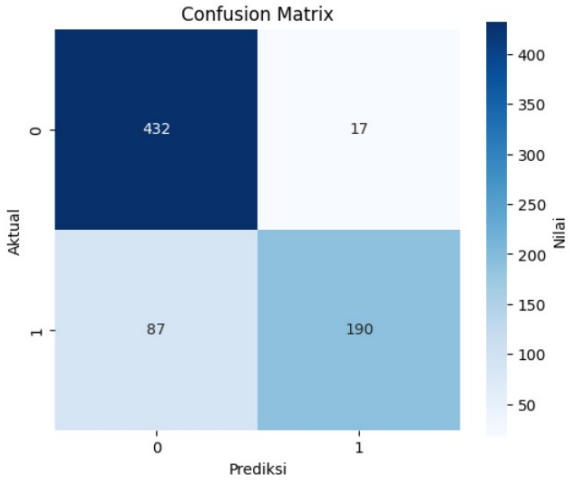
Dari grafik pada Gambar.2 dapat dilihat bahwa akurasi model cenderung stabil dan meningkat pada nilai K tertentu. Pada nilai K = 5 (nilai terbaik yang ditemukan dalam pengujian), model menunjukkan akurasi tertinggi. Oleh karena itu, nilai K = 5 dipilih sebagai nilai optimal untuk model K-NN data tersebut. Setelah memilih nilai K terbaik, penulis mengevaluasi kinerja model dengan menggunakan *classification report,* yang memberikan metrik evaluasi seperti *precision, recall, dan f1-score* untuk setiap kelas. Berikut adalah hasil *classification report* untuk model K-NN yang diuji :

	presicion	re-call	f1-score	support
Non drop out	0.83	0.96	0.89	449
drop out	0.92	0.69	0.79	277

Gambar 3. Tabel Hasil *Classification Report*

Dari Gambar 3 diatas dapat disimpulkan bahwa:

- *Precision* untuk kelas *non drop out* (0) adalah 0.83, yang menunjukkan bahwa 83% dari prediksi model untuk kelas ini benar-benar *non drop out.*
- *Recall* untuk kelas *non drop out* (0) mencapai 0.96, yang berarti model berhasil mendeteksi 96% mahasiswa yang sebenarnya *non drop out.*
- *F1-Score* untuk kelas *non drop out* (0) adalah 0.89, yang menunjukkan keseimbangan yang baik antara *precision* dan *recall*
- Untuk kelas *drop out* (1), *precision* adalah 0.92, *recall* adalah 0.69, dan *F1-Score* 0.79. Model menunjukkan kesulitan dalam mendeteksi mahasiswa *drop out* yang tercermin dari nilai *recall* yang lebih rendah.
- Sebagai evaluasi lebih lanjut, penulis memvisualisasikan *confusion matrix* untuk memahami distribusi prediksi model dalam membedakan antara kelas *drop out dan non-drop out*. Sebagai berikut:



Gambar 4. Visualisasi dari *Matrik Confusion*

Dari gambar diatas menunjukan seberapa baik model dapat mengklasifikasikan data ke dalam kelas yang benar:

- *True Negatives (TN)*: jumlah prediksi yang benar ketika model memprediksi kelas negatif (0). Pada matriks ini, TN adalah 432 (bagian kiri atas).
- *False Positives (FP):* jumlah prediksi yang salah ketika model memprediksi kelas positif (1), tetapi seharusnya negatif (0). Pada matriks ini, FP adalah 17 (bagian kanan atas).
- *False Negatives (FN):* jumlah prediksi yang salah ketika model memprediksi kelas negatif (0), tetapi seharusnya positif (1). Pada matriks ini, FN adalah 87 (bagian kiri bawah).
- *True Positives (TP):* jumlah prediksi yang benar ketika model memprediksi kelas positif (1). Pada matriks ini, TP adalah 190 (bagian kanan bawah).

Dari hasil tersebut, jumlah *false negatives* memiliki nilai yang cukup tinggi. *False negatives* adalah jumlah kasus positif aktual yang gagal diidentifikasi oleh model sebagai positif. Artinya ketika sesuatu yang seharusnya terdeteksi sebagai positif justru diprediksi sebagai negatif oleh model. Pada penelitian ini nilai *false negatives* sebesar 190 orang yang memiliki implikasi dalam hasil penelitian, yaitu mahasiswa berisiko tidak terdeteksi. Semakin tinggi nilai *false negatives* maka semakin sulit terdeteksi untuk *drop out* atau yang seharusnya *drop out* tetapi salah diprediksi sebagai *non drop out*. Untuk meminimalisasi kan agar data tersebut dapat terdeteksi dengan cara optimalkan nilai *recall* karena

false negatives berhubungan dengan *recall*, salah satu caranya dengan menambahkan teknik *resampling* yang dapat membantu mengurangi ketidakseimbangan data dan mengoptimalkan nilai *recall*.

UCAPAN TERIMA KASIH

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat, tauhid, dan hidayah-Nya kami dapat menyelesaikan artikel ilmiah "Penerapan *K-Nearest Neighbors* (KNN) dalam Menilai Potensi Drop Out Mahasiswa: Studi pada Aspek Akademik, Sosial, dan Ekonomi" hingga selesai. Tak lupa kami menghaturkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasihat, dan pemikiran, terutama kepada :

1. Bapak Dimas Avian Maulana, S.Si., M.Si selaku dosen pembimbing yang telah memberikan wawasan, bimbingan, dan dukungan selama proses penulisan artikel ilmiah.
2. Teman-teman kelompok yang sudah bekerjasama dalam penulisan artikel ilmiah selama 2 bulan.
3. Seluruh pihak yang sudah terlibat secara langsung maupun tidak langsung dalam penulisan artikel ilmiah ini.

Mohon maaf jika terdapat kesalahan dalam penulisan artikel ilmiah ini. Kami berharap artikel ilmiah ini dapat menjadi langkah awal menuju pengetahuan yang lebih dalam dan memberikan manfaat yang lebih luas bagi kita semua.

PENUTUP

SIMPULAN

Dalam penelitian ini, algoritma K-NN digunakan untuk memprediksi status mahasiswa, yang awalnya terdiri dari tiga kelas: *graduate, enrolled,* dan *drop out*. Data target kemudian diubah menjadi dua kelas: *drop out* dan *non-drop out* untuk menyederhanakan proses klasifikasi. Model ini menunjukkan kinerja yang baik dalam memprediksi mahasiswa *non drop out*, dengan *precision* sebesar 0.83, *recall* 0.96, dan *F1-score* 0.89. Hal ini menunjukkan bahwa model K-NN cukup efektif dalam mendeteksi mahasiswa yang tidak berisiko *drop out.* Namun, untuk kelas *drop out,* meskipun *precision*-nya cukup tinggi (0.92), nilai *recall* yang rendah (0.69) menunjukkan bahwa model masih kesulitan dalam mendeteksi mahasiswa yang berisiko *drop out*, yang kemungkinan besar disebabkan oleh ketidakseimbangan jumlah data antara kelas *drop out* dan *non-drop out.*

Dengan demikian, meskipun model K-NN menunjukkan kinerja yang baik secara keseluruhan, ada kelemahan dalam mendeteksi mahasiswa yang berisiko *drop out*. Hal ini mengindikasikan perlunya perbaikan lebih lanjut, seperti penggunaan teknik *resampling* untuk mengatasi ketidakseimbangan kelas, atau penyesuaian bobot kelas.

SARAN

Berdasarkan hasil analisis, terdapat beberapa saran untuk meningkatkan artikel ilmiah mengenai penggunaan algoritma KNN dalam memprediksi status mahasiswa *drop out* dan *non-drop out*, sebagai berikut:

1. Ada bagian analisis hasil, tambahkan pembahasan lebih mendalam mengenai ketidakseimbangan kelas antara *drop out* dan *non drop out*. Misalnya, menjelaskan mengapa *recall* untuk kelas *drop out* relatif rendah meskipun *precision* tinggi, dan bagaimana teknik seperti oversampling, undersampling, atau penyesuaian bobot kelas dapat membantu mengatasi masalah tersebut.
2. Melakukan penelitian dengan algoritma lain: mengingat kelemahan model K-NN dalam mendeteksi mahasiswa yang berisiko *drop out*, untuk selanjutnya penulis di harapkan bisa menggunakan algoritma lain yang lebih efektif untuk data tidak seimbang, seperti *Support Vector Machines (SVM)* atau *Random Forest.*
3. Memanfaatkan data yang lebih lengkap serta mengoptimalkan parameter pada metode prediksi untuk meningkatkan akurasi hasil analisis.

Diharapkan dengan menerapkan saran-saran di atas, model prediksi dapat lebih akurat dalam mendeteksi mahasiswa berisiko *drop out*. Penggunaan algoritma lain dan optimasi data serta parameter diharapkan meningkatkan kinerja model,

sehingga penelitian berikutnya dapat menghasilkan hasil yang lebih akurat

DAFTAR PUSTAKA

Shiddiqi, M.F.A dan Bhakti, H.D. (2024). " Perbandingan Algoritme K-nearest Neighbor dan Naive Bayes dalam Klasifikasi Kelulusan Mahasiswa ". Jurnal Mahasiswa Teknik Informatika, vol. 8, no. 5.

Novitasari, F. (2023). " Sistem Klasifikasi Penyakit Jantung Menggunakan teknik Pendekatan Smote pada Algoritma Modified K-Nearest Neighbor ". Building of Informatics technology and Science, vol. 5, no. 1.

Cahyanti, F.L.D, et al. (2023). "Klasifikasi Data Mining dengan Algoritma Machine Learning untuk Prediksi Penyakit Liver". Technologia, vol. 14, no. 2.

Naufal, M.F & Kusuma, S.F. (2023). " Analisis Perbandingan Algoritma Machine Learning dan Deep Learning untuk Klasifikasi Citra Sistem Isyarat Bahasa Indonesia (SIBI) ". Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 10, no. 4.

Sinaga, D., Solaiman, E.J., & Kaunang, F.J. (2021). " Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out di Universitas Advent Indonesia ". Jurnal TeIKa, vol. 11, no. 2.

Anwar, M.T., Lucky Heriyanto, L., & Fanini, F. (2021). " Model Prediksi Dropout Mahasiswa Menggunakan Teknik data Mining ". Jurnal Informatika Upgris, vol. 7, no. 1.

Sidik, A.D.W.M, et al. (2020). "Gambaran Umum Metode Klasifikasi Data Mining". Jurnal Teknik Elektro, vol. 2, no. 2.

Ratniasih, N.L. (2019). " Penerapan algoritma K-Nearest Neighbour (K-NN) untuk Penentuan Mahasiswa Berpotensi Drop Out". Jurnal Teknologi Informasi dan Komputer, vol. 5, no. 3.

Ramayasa, I.P. (2018). " Perancangan Sistem Klasifikasi Mahasiswa Drop Out Menggunakan Algoritma K-Nearest Neighbor ". Seminar Nasional Sistem Informasi dan Teknologi Informasi.

Yeyen Dwi Atma, Y.D. & Setyanto, A. (2018). " Perbandingan Algoritma C4.5 dan K-NN dalam Identifikasi Mahasiswa Berpotensi Drop Out ". METIK JURNAL, vol. 2, no. 2.

Hendrian, S.(2018). "Algoritma Klasifikasi data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan". Faktor Exacta, vol. 11, no. 3.

Saputra, Y. & Primadasa, Y. (2018). "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor". Techno.COM, vol. 17, no. 4.

Chandani, V., Purwanto & Wahono, R.S. (2015). " Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film". Journal of Intelligent Systems, vol. 1, no. 1.

Nurhayati, S., Kusrini & Luthfi, E.T. (2015). " Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine". Jurnal Ilmiah SISFOTENIKA, vol. 5, no.1.

Ratnawati, D.E., Marji & Muflikhah, L. (2014). " Pengembangan Metode Klasifikasi Berdasarkan K-Means dan LVQ". Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 1, no. 1.