



Analisis Perbandingan Kinerja Algoritma C45 Dan Naive Bayes Dalam Prediksi Penerimaan Mahasiswa Baru Di Universitas Pamulang

**Rahdat Jajuly^{1*}, Rio Maryono², Achmad Agun Gunawan³, Nandito Heriyana⁴, Aries
Saifudin⁵**

^{1,2,3,4,5}Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

E-mail : ^{1*}rahdat.j7@gmail.com, ²riomaryono747@gmail.com, ³gunawanacmad7@gmail.com,

⁴nanditoheriyana@gmail.com, ⁵aries.saifudin@unpam.ac.id

(* : coressponding author)

Abstrak - Penelitian ini bertujuan untuk mengadakan analisis perbandingan kinerja antara algoritma C45 dan Naive Bayes dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang. Data yang digunakan data training dan data testing. Hasil eksperimen menunjukkan bahwa algoritma C45 mencapai tingkat akurasi sebesar 88.74%, sedangkan Naive Bayes mencapai 87.24%. Temuan ini menunjukkan bahwa algoritma C45 lebih unggul dalam memprediksi jumlah calon mahasiswa baru yang akan mendaftar di universitas ini. Studi ini memberikan wawasan yang berharga bagi pengambil keputusan di universitas untuk meningkatkan efisiensi dan efektivitas proses penerimaan mahasiswa baru.

Kata Kunci : Algoritma C45, Naive Bayes, Prediksi Penerimaan Mahasiswa Baru, Universitas Pamulang, Data Mining, Klasifikasi, Tingkat Akurasi.

Abstract - This research aims to conduct analysis performance comparison between the C45 and Naive Bayes algorithms in predicting new student admissions at Pamulang University. The data used is training data and testing data. Experimental results show that the C45 algorithm achieves an accuracy level of 88.74%, while Naive Bayes reaches 87.24%. These findings indicate that the C45 algorithm is superior in predicting the number of prospective new students who will enroll at this university. This study provides valuable insights for university decision makers to improve the efficiency and effectiveness of the new student admissions process.

Keywords: C45 Algorithm, Naive Bayes, Prediction Of New Student Admissions, Pamulang University, Data Mining, Classification, Accuracy Level.

1. PENDAHULUAN

Universitas Pamulang merupakan salah satu Perguruan Tinggi Swasta yang telah berdiri tahun 15 Mei 2000 cukup lama dengan jumlah mahasiswa yang diterima di tiap tahun cukup banyak (Yahya et al., 2019). Setiap tahun ajaran baru, kegiatan penerimaan mahasiswa baru di kampus ini terus menghasilkan data mahasiswa yang semakin banyak. Salah satu indikator vital dalam operasional perguruan tinggi adalah proses penerimaan mahasiswa baru (PMB). Di era digital ini, penggunaan teknik data mining dalam analisis data PMB menjadi semakin penting untuk mendukung pengambilan keputusan yang tepat dan efisien.

Universitas Pamulang, sebagai salah satu institusi pendidikan tinggi di Indonesia, terus menghadapi tantangan dalam menarik minat calon mahasiswa baru. Masih ada kesenjangan antara jumlah pendaftar dan jumlah mahasiswa yang akhirnya mendaftar ulang. Faktor-faktor seperti preferensi masyarakat terhadap perguruan tinggi negeri dan swasta, serta persepsi terhadap kualitas pendidikan, menjadi pertimbangan utama dalam proses ini.

Dalam konteks ini, teknik klasifikasi dalam data mining menawarkan pendekatan yang kuat untuk memprediksi keberhasilan penerimaan mahasiswa baru. Dua algoritma yang umum digunakan dalam klasifikasi adalah C45 dan Naive Bayes. Kedua algoritma ini memiliki pendekatan yang berbeda namun potensial dalam memberikan hasil yang akurat.

Penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja antara algoritma C45 dan Naive Bayes dalam memprediksi jumlah mahasiswa baru yang akan mendaftar di Universitas Pamulang. Data yang digunakan dalam penelitian ini terdiri dari sejumlah besar data historis PMB yang mencakup periode beberapa tahun terakhir. Dengan menggunakan teknik ini, diharapkan dapat



memberikan wawasan yang mendalam bagi pengambil keputusan universitas dalam meningkatkan strategi penerimaan mahasiswa baru yang lebih efektif dan efisien.

2. METODOLOGI

Penelitian ini dilakukan untuk membandingkan kinerja algoritma C45 dan Naive Bayes dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang. Metodologi penelitian yang digunakan meliputi beberapa tahap,

sebagai berikut:

Langkah-langkah yang akan diambil dalam penelitian:

a. Pengumpulan Data:

Data yang digunakan dalam penelitian ini terdiri dari data historis penerimaan mahasiswa baru di Universitas Pamulang. Data tersebut mencakup atribut-atribut seperti:

- 1) Nama calon mahasiswa
- 2) Asal sekolah
- 3) Nilai ujian masuk
- 4) Program studi yang diminati
- 5) Status registrasi ulang (ya/tidak)

b. Pra-pemrosesan Data:

Data yang terkumpul lalu diproses untuk memastikan kualitas dan konsistensinya. Tahapan pra-pemrosesan meliputi:

- 1) Pembersihan data: Menghapus atau memperbaiki data yang hilang atau tidak valid.
- 2) Normalisasi data: Mengubah data ke dalam bentuk yang sesuai untuk diproses lebih lanjut.
- 3) Pembagian data: Data dibagi menjadi data training (2133 data) dan data testing (533 data).

c. Pembagian Data:

Dua algoritma klasifikasi yang digunakan dalam penelitian ini adalah C45 dan Naive Bayes.

1) Algoritma C45:

Algoritma ini membangun pohon keputusan berdasarkan gain informasi dari setiap atribut. Setiap cabang pohon keputusan menunjukkan hasil keputusan berdasarkan atribut-atribut tertentu.

2) Algoritma Naive Bayes:

Naive Bayes lebih cocok digunakan pada data berukuran besar dan mampu menangani data yang tidak lengkap serta tangguh terhadap atribut yang tidak relevan dalam dataset. (Metode et al., n.d.) Algoritma ini menggunakan teorema Bayes dengan asumsi independensi antar atribut. Setiap atribut dianggap berkontribusi secara independen terhadap probabilitas hasil keputusan

d. Evaluasi Kinerja:

Kinerja kedua algoritma dievaluasi menggunakan data testing. Beberapa metrik evaluasi yang digunakan meliputi:

- 1) Akurasi: Proporsi prediksi yang benar dari total prediksi.
- 2) Precision: Proporsi prediksi positif yang benar dari total prediksi positif.
- 3) Recall: Proporsi prediksi positif yang benar dari total kasus positif aktual.



4) F1 Score: Harmoni rata-rata dari precision dan recall.

e. Analisis Hasil:

Hasil evaluasi dari kedua algoritma dibandingkan untuk menentukan algoritma yang memiliki kinerja terbaik. Dalam penelitian ini, algoritma dengan akurasi tertinggi dianggap sebagai yang paling efektif untuk memprediksi penerimaan mahasiswa baru di Universitas Pamulang.

f. Analisis Hasil:

Berdasarkan hasil analisis, kesimpulan diambil mengenai algoritma yang paling sesuai untuk digunakan dalam prediksi penerimaan mahasiswa baru. Rekomendasi juga diberikan kepada pengelola universitas untuk meningkatkan proses penerimaan mahasiswa baru berdasarkan temuan penelitian ini.

3. HASIL DAN PEMBAHASAN

Penelitian ini berhasil membandingkan kinerja dua algoritma klasifikasi, yaitu C45 dan Naive Bayes, dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang. Data yang digunakan terdiri dari data historis penerimaan mahasiswa baru, dengan 2133 data sebagai data training dan 533 data sebagai data testing. Hasil evaluasi kinerja menunjukkan bahwa algoritma C45 mencapai tingkat akurasi sebesar 88.74%, precision sebesar 89.00%, recall sebesar 88.50%, dan F1 Score sebesar 88.75%. Sementara itu, algoritma Naive Bayes menunjukkan akurasi sebesar 87.24%, precision sebesar 87.50%, recall sebesar 87.00%, dan F1 Score sebesar 87.25%.

Hasil ini menunjukkan bahwa algoritma C45 memiliki kinerja yang lebih baik dibandingkan dengan Naive Bayes dalam memprediksi status registrasi ulang mahasiswa baru di Universitas Pamulang. Akurasi yang lebih tinggi pada C45 menunjukkan bahwa algoritma ini lebih tepat dalam memprediksi apakah seorang calon mahasiswa akan mendaftar ulang atau tidak. Precision yang lebih tinggi pada algoritma C45 (89.00%) menunjukkan bahwa algoritma ini lebih efektif dalam mengurangi jumlah false positives, yaitu kasus di mana prediksi menyatakan mahasiswa akan registrasi ulang padahal sebenarnya tidak. Sementara itu, recall yang lebih tinggi pada C45 (88.50%) menunjukkan kemampuannya dalam mengidentifikasi semua calon mahasiswa yang benar-benar akan registrasi ulang, mengurangi jumlah false negatives.

F1 Score memberikan gambaran yang seimbang antara precision dan recall. Algoritma C45 memiliki F1 Score yang lebih tinggi (88.75%) dibandingkan dengan Naive Bayes (87.25%), mengindikasikan bahwa C45 memberikan keseimbangan yang lebih baik antara prediksi yang tepat dan lengkap.

Keunggulan utama dari algoritma C45 adalah salah satu metode klasifikasi data menggunakan teknik keputusan yang terkenal dan disukai karena berbagai kelebihanannya. Kelebihan tersebut antara lain: mampu mengolah data numerik (kontinu) dan diskret, mampu menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan, dan tercepat dibandingkan dengan algoritma-algoritma lain yang menggunakan memori utama di komputer. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, yang ditemukan oleh J. Ross Quinlan pada tahun 1986. Algoritma ID3 adalah singkatan dari Iterative Dichotomizer 3 dan dikategorikan sebagai algoritma pembelajaran (Susila & Suradarma, 2018). Algoritma ini membangun pohon keputusan yang mudah diinterpretasikan, membantu pengambil keputusan di Universitas Pamulang memahami faktor-faktor yang mempengaruhi registrasi ulang mahasiswa baru. Namun, kelemahan dari C45 adalah kompleksitasnya yang lebih tinggi, membutuhkan lebih banyak sumber daya komputasi dan waktu pemrosesan, terutama untuk dataset yang sangat besar.

Naive Bayes unggul dalam kecepatan dan efisiensi komputasi. Algoritma ini sederhana dalam implementasinya karena mengasumsikan independensi antar atribut. Namun, asumsi independensi ini bisa menjadi kelemahan jika atribut-atribut tersebut tidak benar-benar independen, seperti yang terlihat pada akurasi yang sedikit lebih rendah dibandingkan C45.



Berdasarkan hasil penelitian ini, algoritma C45 lebih disarankan untuk digunakan dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang karena akurasi dan F1 Score yang lebih tinggi. Namun, Naive Bayes tetap dapat menjadi alternatif yang baik jika waktu dan sumber daya komputasi menjadi kendala, karena algoritma ini tetap memberikan akurasi yang cukup tinggi dengan efisiensi komputasi yang lebih baik.

Untuk meningkatkan keakuratan prediksi di masa mendatang, disarankan untuk melakukan analisis lebih lanjut dengan menambahkan lebih banyak variabel yang mungkin mempengaruhi keputusan registrasi ulang calon mahasiswa, serta mencoba kombinasi teknik data mining lainnya seperti ensemble methods yang dapat menggabungkan kelebihan dari beberapa algoritma.

4. KESIMPULAN

4.1 Kesimpulan

Penelitian ini membandingkan kinerja algoritma C45 dan Naive Bayes dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang. Berdasarkan hasil analisis dan evaluasi, beberapa kesimpulan utama yang dapat diambil adalah sebagai berikut:

Algoritma C45 menunjukkan akurasi yang lebih tinggi (88.74%) dibandingkan dengan Naive Bayes (87.24%). Hal ini mengindikasikan bahwa C45 lebih unggul dalam memprediksi apakah calon mahasiswa akan melanjutkan ke proses registrasi ulang atau tidak.

Algoritma C45 memiliki nilai precision (89.00%) dan recall (88.50%) yang lebih tinggi dibandingkan dengan Naive Bayes, yang masing-masing memiliki nilai precision (87.50%) dan recall (87.00%). Ini menunjukkan bahwa C45 lebih efektif dalam mengidentifikasi calon mahasiswa yang benar-benar akan mendaftar ulang, sekaligus mengurangi jumlah prediksi yang salah.

F1 Score dari algoritma C45 (88.75%) juga lebih tinggi daripada Naive Bayes (87.25%), memperkuat kesimpulan bahwa C45 memberikan keseimbangan yang lebih baik antara precision dan recall dalam prediksi.

Algoritma C45 mampu menangani atribut kontinu dan kategorikal dengan baik, serta menghasilkan pohon keputusan yang mudah diinterpretasikan. Hal ini memberikan nilai tambah dalam membantu pengambil keputusan di Universitas Pamulang memahami faktor-faktor yang mempengaruhi registrasi ulang mahasiswa baru.

Naive Bayes unggul dalam kecepatan dan efisiensi komputasi. Algoritma ini sederhana dalam implementasinya karena mengasumsikan independensi antar atribut. Meskipun asumsi ini bisa menjadi kelemahan jika atribut tidak benar-benar independen, algoritma ini tetap memberikan akurasi yang cukup tinggi dengan efisiensi yang lebih baik.

Berdasarkan hasil penelitian, algoritma C45 lebih disarankan untuk digunakan dalam memprediksi penerimaan mahasiswa baru di Universitas Pamulang karena kinerja akurasinya yang lebih tinggi. Namun, Naive Bayes dapat menjadi alternatif yang baik jika waktu dan sumber daya komputasi menjadi kendala.

Untuk meningkatkan keakuratan prediksi di masa mendatang, disarankan untuk melakukan analisis lebih lanjut dengan menambahkan lebih banyak variabel yang mungkin mempengaruhi keputusan registrasi ulang calon mahasiswa, serta mencoba kombinasi teknik data mining lainnya seperti ensemble methods yang dapat menggabungkan kelebihan dari beberapa algoritma. Temuan penelitian ini memberikan wawasan yang berharga bagi pengambil keputusan di Universitas Pamulang dalam meningkatkan efisiensi dan efektivitas proses penerimaan mahasiswa baru.

4.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang diperoleh, terdapat beberapa saran yang dapat diberikan untuk meningkatkan keakuratan prediksi dan efisiensi proses penerimaan mahasiswa baru di Universitas Pamulang:



Untuk meningkatkan keakuratan prediksi, disarankan menambahkan lebih banyak variabel yang mungkin mempengaruhi keputusan registrasi ulang calon mahasiswa. Variabel-variabel tambahan ini bisa berupa latar belakang keluarga, motivasi memilih Universitas Pamulang, tingkat kepuasan terhadap fasilitas kampus, dan faktor-faktor eksternal lainnya.

Menggunakan teknik ensemble methods, seperti Random Forest atau Gradient Boosting, dapat menggabungkan kelebihan dari beberapa algoritma dan meningkatkan keakuratan prediksi. Teknik ini dapat memberikan hasil yang lebih andal dan robust dibandingkan dengan menggunakan satu algoritma saja.

Melakukan pra-pemrosesan data yang lebih mendalam, termasuk penanganan missing values, outliers, dan normalisasi data, dapat meningkatkan kualitas data dan, pada akhirnya, akurasi prediksi. Penggunaan teknik seperti imputation untuk missing values dan scaling untuk normalisasi dapat membantu mencapai hasil yang lebih baik.

Model prediksi perlu diperbarui secara berkala dengan data terbaru untuk memastikan bahwa model tetap relevan dan akurat. Pemantauan kinerja model secara terus-menerus dan melakukan retraining dengan data baru dapat membantu menjaga akurasi prediksi.

Memberikan pelatihan dan workshop kepada staf dan pengambil keputusan di Universitas Pamulang mengenai teknik data mining dan interpretasi hasil dapat meningkatkan pemahaman dan penggunaan hasil prediksi dalam pengambilan keputusan strategis.

Mengintegrasikan model prediksi dengan sistem informasi kampus yang sudah ada dapat mempercepat dan mempermudah akses ke hasil prediksi. Hal ini juga memungkinkan otomatisasi proses prediksi dan pemantauan hasil secara real-time.

Selain algoritma C45 dan Naive Bayes, disarankan untuk menguji kinerja metode klasifikasi lain, seperti Support Vector Machine (SVM), K-Nearest Neighbors (KNN), atau Neural Networks, untuk mencari model yang mungkin memiliki performa lebih baik.

Bekerja sama dengan perguruan tinggi lain yang memiliki situasi serupa dapat memberikan perspektif baru dan memungkinkan pertukaran data serta teknik yang dapat meningkatkan kualitas prediksi.

Dengan menerapkan saran-saran ini, diharapkan Universitas Pamulang dapat meningkatkan efisiensi dan efektivitas proses penerimaan mahasiswa baru, serta membuat keputusan yang lebih tepat berdasarkan analisis data yang akurat.

REFERENCES

- Metode, P., Bayes, N., Dan, C., Penerimaan, P., & Di, P. (n.d.). *Penerapan metode naïve bayes dan c4.5 pada penerimaan pegawai di universitas potensi utama*. 51–62.
- Susila, I. M. D., & Suradarma, I. B. (2018). Penerapan Optimasi Algoritma C45 dengan Naïve Bayes pada Pemilihan Internet Service Provider. *Eksplora Informatika*, 7(2), 16–26. <https://eksplora.stikom-bali.ac.id/index.php/eksplora/article/view/126/108>
- Yahya, N., Jananto, A., Studi, P., Informasi, S., Informasi, F. T., Stikubank, U., Mining, D., Klasifikasi, M., & Bayes, N. (2019). *UNTUK PREDIKSI KEGIATAN PENERIMAANMAHASISWA BARU (STUDI KASUS : UNIVERSITAS STIKUBANK SEMARANG)*. 2014, 978–979