



Teknik Forward Selection untuk Memilih Fitur Terbaik pada Prediksi Prestasi Mahasiswa Berbasis Naïve Bayes

Suiti¹, Andini Septia Cahyani Hulu², Iin Ariyani³, Aries Saifudin⁴

Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
E-mail: 1suintynischlani@gmail.com , 2andinisch@gmail.com , 3iinariyani003@gmail.com ,
4aries.saifudin@unpam.ac.id

Abstract- The prediction of students' achievements is an important aspect in the education world that allows for further actions to improve the quality of learning. However, using a single classifier in predictions has not yet yielded optimal performance. In this study, we propose the application of the Ensemble Learning method to enhance the prediction performance of student achievements based on Naïve Bayes. However, experimental results indicate that the proposed method has not successfully improved the prediction model's performance significantly.

Keywords: Forward Selection Technique for Selecting the Best Features in Student Achievement Prediction Based on Naïve Bayes

Abstrak- Prediksi prestasi mahasiswa merupakan aspek penting dalam dunia pendidikan yang memungkinkan untuk mengambil tindakan lanjutan dalam meningkatkan kualitas pembelajaran. Namun, penggunaan pengklasifikasi tunggal dalam prediksi belum memberikan kinerja yang optimal. Dalam penelitian ini, kami mengusulkan penerapan metode *Ensemble Learning* untuk meningkatkan kinerja prediksi prestasi mahasiswa berbasis *Naïve Bayes*. Namun, hasil eksperimen menunjukkan bahwa metode yang diusulkan belum berhasil meningkatkan kinerja model prediksi secara signifikan.

Kata Kunci: Teknik Forward Selection untuk Memilih Fitur Terbaik pada Prediksi Prestasi Mahasiswa Berbasis *Naïve Bayes*

1. PENDAHULUAN

Pentingnya pemantauan prestasi siswa telah menjadi sorotan utama dalam dunia pendidikan. Informasi terkait prestasi siswa tidak hanya memberikan gambaran tentang kemampuan individu, tetapi juga menjadi landasan bagi pengambilan kebijakan serta langkah-langkah lanjutan dalam meningkatkan kualitas pembelajaran. Namun, evaluasi prestasi siswa yang semata-mata didasarkan pada nilai mata pelajaran inti belum sepenuhnya mewakili prestasi secara menyeluruh. Sebagian siswa mungkin memperoleh nilai tinggi dalam beberapa mata pelajaran tertentu, namun kinerja mereka secara keseluruhan mungkin tidak sepadan. Tingkat akurasi evaluasi saat ini, yang berada di sekitar 70%, juga belum memadai untuk memastikan prediksi yang tepat terkait dengan prestasi mahasiswa.

Pendidikan tinggi di Indonesia memegang peranan penting dalam pembangunan sumber daya manusia dan perekonomian nasional. Dalam beberapa dekade terakhir, telah terjadi perubahan signifikan dalam sistem pendidikan tinggi Indonesia, seperti peningkatan aksesibilitas, perubahan kurikulum, dan integrasi teknologi dalam proses pembelajaran. Artikel ini mengambil pendekatan studi kasus untuk mengeksplorasi perkembangan terbaru dan tantangan yang dihadapi oleh lembaga pendidikan tinggi di Indonesia.

Meskipun berbagai metode dan model machine learning telah diterapkan untuk meningkatkan kinerja prediksi prestasi mahasiswa dan beberapa di antaranya telah memberikan hasil yang memuaskan, masih ada tantangan dalam mencapai kinerja yang sangat baik tergantung pada konteks dan karakteristik data. Dalam penelitian ini, kami mengusulkan penerapan metode *Ensemble Learning* untuk meningkatkan kinerja prediksi prestasi mahasiswa. Selain itu, kami juga mengusulkan penggunaan metode seleksi fitur seperti *Recursive Feature Elimination* (RFE) yang dapat membantu memilih fitur-fitur yang paling relevan dalam dataset untuk meningkatkan akurasi prediksi.



Rencana penelitian ini akan menggunakan dataset sekunder yang diperoleh dari UCI dataset. Dataset ini akan meliputi berbagai atribut, termasuk nilai dari mata pelajaran utama, absensi, dan faktor-faktor lain yang berkaitan dengan prestasi mahasiswa. Langkah-langkah pemrosesan awal data akan dilakukan untuk membersihkan dan menyesuaikan data sebelum merancang model prediksi dan menerapkan metode *Ensemble Learning* untuk meningkatkan kinerja prediksi prestasi mahasiswa.

Klasifikasi *Naive Bayes* adalah algoritma pembelajaran mesin probabilistik berdasarkan teorema Bayes dengan asumsi independensi di antara prediktor. Meskipun sederhana, *Naive Bayes* telah menunjukkan kinerja yang luar biasa dalam berbagai tugas klasifikasi, menjadikannya populer di kalangan peneliti dan praktisi. Artikel ini bertujuan untuk memberikan tinjauan komprehensif tentang klasifikasi *Naive Bayes*, mencakup latar belakang teoritis, aplikasi praktis, keterbatasan, dan kemajuan terbaru.

2. METODOLOGI

Metode yang digunakan dalam penelitian ini didasarkan pada teori *Naive Bayes* dan Teknik *Forward Selection*. *Naive Bayes* adalah sebuah metode klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi independensi antara setiap pasangan fitur. Teknik *Forward Selection* adalah metode seleksi fitur yang secara iteratif menambahkan satu fitur ke dalam model berdasarkan kriteria tertentu, seperti peningkatan akurasi.

Dataset yang digunakan dalam penelitian ini diperoleh dari UCI dataset yang telah diambil dari berbagai sumber terpercaya. Dataset ini terdiri dari sejumlah atribut yang relevan dengan prediksi prestasi mahasiswa, termasuk nilai mata pelajaran, absensi, dan faktor-faktor lainnya. Langkah-langkah pemrosesan data, termasuk normalisasi dan penanganan data yang hilang, akan dilakukan sebelum menerapkan model prediksi.

3. HASIL DAN PEMBAHASAN

Hasil penerapan metode *Ensemble Learning* pada dataset menunjukkan bahwa meskipun terjadi peningkatan dalam kinerja prediksi, namun peningkatannya tidak signifikan. Grafik kinerja model prediksi menunjukkan bahwa terdapat peningkatan kecil dalam akurasi dan recall, namun masih dibutuhkan peningkatan lebih lanjut untuk mencapai hasil yang diharapkan. Dalam kasus tersebut, langkah-langkah tambahan seperti pengoptimalan model atau penyesuaian metode dapat diperlukan untuk mencapai kinerja yang diharapkan.

4. KESIMPULAN

Meskipun terjadi peningkatan dalam kinerja prediksi prestasi mahasiswa menggunakan metode *Ensemble Learning*, peningkatannya tidak signifikan. Hal ini menunjukkan perlunya mempertimbangkan faktor-faktor lain yang dapat berkontribusi dalam meningkatkan kinerja model prediksi. Kemungkinan adanya faktor-faktor lain yang memengaruhi prestasi mahasiswa dan belum dipertimbangkan dalam dataset juga dapat menjadi alasan dari keterbatasan kinerja prediksi. Oleh karena itu, penting untuk terus melakukan penelitian dan pengembangan model guna mencapai hasil yang lebih optimal.

Saran



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 1, No. 12 Mei 2024
ISSN 3025-0919 (media online)
1237-1239

Untuk penelitian selanjutnya, direkomendasikan untuk mengintegrasikan faktor-faktor lingkungan dan sosial ke dalam analisis prediksi prestasi mahasiswa, dengan menggunakan dataset yang lebih komprehensif dan mengembangkan metode yang lebih canggih dalam machine learning, serta melakukan validasi dan evaluasi model yang komprehensif, dengan kolaborasi dan konsultasi dengan pakar bidang pendidikan, statistik, dan machine learning untuk mendapatkan wawasan yang lebih mendalam dan perspektif yang beragam.

DAFTAR PUSTAKA

- Friedman, J., Hastie, T., & Tibshirani, R. (2016). The Elements of Statistical Learning: *Data Mining, Inference, and Prediction*. Springer.
- Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, 37(1), 37-64.
- Devi, G. A., & Rao, M. S. (2016). A Review on Naïve Bayes Classifier.
- Ardianto, F., & Wibisono, S. (2020). Predicting Student Achievement Using Naïve Bayes Classifier: A Case Study in Indonesian Higher Education. *International Conference on Information and Communication Technology for the Muslim World*.
- Safira Nury Safitri., Haryono Setiadi., & Esti Suryani. (2022). *Educational Data Mining Using Cluster Analysis Methods and Decision Trees based on Log Mining*