



## Metode Seleksi Maju untuk Memilih Karakteristik Terbaik untuk Memprediksi Keberhasilan Siswa

Steven Fredrico<sup>1</sup>, Okky Rahmat Saputra<sup>2</sup>, M. Zulbi Alzarudin<sup>3</sup>, Rintan Ayu Lestari<sup>4</sup>, Aries Saifudin<sup>5</sup>

Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

E-mail : <sup>1</sup>[stevenfredrico2002@gmail.com](mailto:stevenfredrico2002@gmail.com) , <sup>2</sup>[okkyrahmatsaputra09@gmail.com](mailto:okkyrahmatsaputra09@gmail.com) , <sup>3</sup>[zulbialzarudin@gmail.com](mailto:zulbialzarudin@gmail.com) , <sup>4</sup>[rintanayulestarii@gmail.com](mailto:rintanayulestarii@gmail.com) , <sup>5</sup>[aries.saifudin@unpam.ac.id](mailto:aries.saifudin@unpam.ac.id)

**Abstrak**—Di era digital saat ini, penggunaan teknik berbasis data untuk memprediksi prestasi mahasiswa telah menjadi semakin penting dalam bidang pendidikan. Teknik prediksi ini memungkinkan lembaga pendidikan untuk mengidentifikasi faktor-faktor yang mempengaruhi prestasi mahasiswa dan mengambil tindakan yang tepat untuk meningkatkan kualitas pendidikan. Salah satu metode yang umum digunakan untuk prediksi prestasi mahasiswa adalah algoritma Naïve Bayes. Namun, dalam implementasinya, pemilihan fitur-fitur yang tepat sangat penting untuk memastikan akurasi dan efektivitas prediksi. Oleh karena itu, artikel ini bertujuan untuk memperkenalkan dan menjelaskan teknik Forward Selection sebagai metode pemilihan fitur terbaik dalam prediksi prestasi mahasiswa berbasis Naïve Bayes.

**Kata Kunci:** teknik berbasis data, prestasi mahasiswa, prediksi, lembaga pendidikan, algoritma Naïve Bayes, seleksi fitur, teknik Forward Selection.

**Abstract**— In the current digital era, the use of data-driven techniques for predicting student performance has become increasingly important in the field of education. These prediction techniques enable educational institutions to identify the factors that influence student performance and take appropriate actions to improve the quality of education. One commonly used method for student performance prediction is the Naïve Bayes algorithm. However, in its implementation, the selection of appropriate features is crucial to ensure the accuracy and effectiveness of predictions. Therefore, this article aims to introduce and explain the Forward Selection technique as a method for selecting the best features in Naïve Bayes-based student performance prediction.

**Keywords:** data-driven techniques, student performance, prediction, educational institutions, Naïve Bayes algorithm, feature selection, Forward Selection technique.

### 1. PENDAHULUAN

Di era digital saat ini, penggunaan teknik berbasis data untuk memprediksi prestasi mahasiswa telah menjadi sangat penting dalam dunia pendidikan. Teknik prediksi ini memberikan wawasan berharga kepada lembaga pendidikan mengenai faktor-faktor yang mempengaruhi prestasi mahasiswa dan memungkinkan pengambilan keputusan yang tepat untuk meningkatkan kualitas pendidikan. Salah satu metode yang sering digunakan untuk prediksi prestasi mahasiswa adalah algoritma Naïve Bayes. Namun, akurasi dan efektivitas prediksi Naïve Bayes sangat bergantung pada pemilihan fitur-fitur yang tepat. Oleh karena itu, artikel ini bertujuan untuk memperkenalkan dan menjelaskan teknik Forward Selection sebagai metode untuk memilih fitur-fitur terbaik dalam prediksi prestasi mahasiswa berbasis Naïve Bayes.

Pemilihan fitur yang relevan memainkan peran yang sangat penting dalam akurasi dan efisiensi model prediksi. Dalam konteks prediksi prestasi mahasiswa, mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keberhasilan akademik sangat penting bagi lembaga pendidikan untuk meningkatkan metodologi pengajaran, memberikan intervensi yang personal, dan mengoptimalkan alokasi sumber daya. Teknik Forward Selection menawarkan pendekatan sistematis untuk memilih fitur-fitur yang paling informatif dengan mengevaluasi dampaknya secara iteratif terhadap performa prediktif model Naïve Bayes.

Naïve Bayes adalah klasifikasi probabilistik yang mengasumsikan independensi antar atribut, sehingga cocok digunakan dalam tugas prediksi prestasi mahasiswa. Dengan memperkirakan peluang terjadinya berbagai hasil berdasarkan atribut-atribut yang ada, Naïve Bayes menyediakan kerangka kerja yang sederhana namun efektif untuk memprediksi prestasi mahasiswa. Namun, akurasi model Naïve Bayes sangat bergantung pada kualitas dan relevansi fitur-fitur yang dipilih. Teknik Forward Selection mengatasi tantangan ini dengan menambahkan fitur-fitur secara bertahap



yang signifikan dalam meningkatkan performa prediktif model, sehingga menghasilkan akurasi dan keandalan yang lebih baik.

Dengan menggunakan teknik Forward Selection, lembaga pendidikan dapat memperoleh wawasan berharga mengenai faktor-faktor yang paling berpengaruh terhadap prestasi mahasiswa. Informasi ini dapat digunakan untuk mengidentifikasi mahasiswa yang berisiko rendah, mengembangkan intervensi yang ditargetkan, dan mengalokasikan sumber daya secara efektif. Selain itu, fitur-fitur yang dipilih dapat memberikan pemahaman yang lebih mendalam tentang pola dan hubungan antara berbagai atribut dengan prestasi mahasiswa, sehingga memungkinkan pendidik untuk menyesuaikan strategi pengajaran dan sistem dukungan sesuai kebutuhan.

Secara ringkas, artikel ini bertujuan untuk memperkenalkan dan menjelaskan teknik Forward Selection sebagai cara memilih fitur-fitur terbaik untuk prediksi prestasi mahasiswa berbasis Naïve Bayes. Dengan menggunakan teknik ini, lembaga pendidikan dapat meningkatkan akurasi dan efektivitas model prediksi mereka, sehingga memungkinkan pengambilan keputusan dan intervensi berbasis data yang berdampak positif terhadap prestasi mahasiswa.

## **2. METODE**

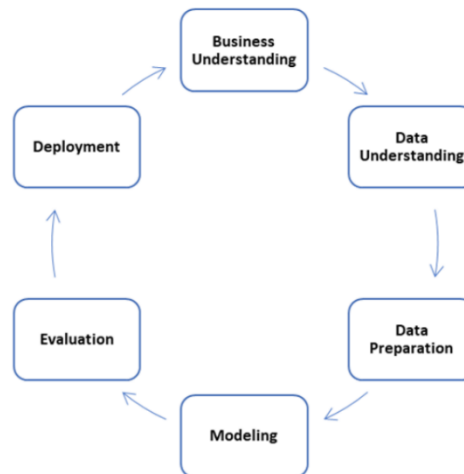
Penelitian ini menggunakan metodologi penelitian Cross Industry Standard Process for Data Mining (CRISP-DM). karena faktanya telah dimanfaatkan oleh banyak peneliti dan terbukti efektif dalam menyelesaikan masalah proyek atau penelitian di domain data mining [25], [34]. Fase pemahaman bisnis, pemahaman data, persiapan data, pemodelan data, evaluasi model, dan penerapan adalah saat proses ini mulai mengalir.

Memahami tujuan dan aspirasi yang mendasari evaluasi bisnis merupakan inti dari proses memahami bisnis. Setelah itu, ini diubah menjadi kerangka data mining awal yang telah dikembangkan untuk mencapai tujuan. Ketajaman bisnis digunakan untuk menggambarkan proses peramalan kapan mahasiswa akan diterima dan tidak, yang ternyata memakan waktu tidak lebih dari empat tahun atau delapan semester. Penjelasan tentang sejarah dan tujuan rangkaian bisnis prediksi penerimaan mahasiswa diperlukan untuk proses ini:

1. Menetapkan tujuan perusahaan (building tujuan perusahaan)  
Untuk memprediksi hasil kelulusan mahasiswa, baik  $\leq 8$  semester maupun  $> 8$  semester, tujuan bisnis penelitian adalah mengungkap pengetahuan dan fitur data mahasiswa.
2. Periksa keadaan (evaluasi keadaan)  
Lingkungan akademik mahasiswa UDINUS terhubung dengan sistem informasi akademik ini, khususnya untuk program studi Sarjana Teknik Informatika. Sistem Informasi Akademik Dian Nuswantoro (SIADIN), merupakan aplikasi berbasis web yang dirancang untuk menangani dan mengolah data berbagai tugas administrasi dan operasional akademik pada program studi di perguruan tinggi, merupakan sistem yang saat ini digunakan. Sistem informasi akademik mengolah data pendukung, data nilai, dan data profil mahasiswa.
3. Menetapkan tujuan data mining (mengidentifikasi tujuan data mining)  
Tujuan dari data mining, atau penelitian ini, adalah untuk membangun model optimal dari berbagai teknik klasifikasi dan mengekstrak informasi tentang pola studi yang memungkinkan mahasiswa lulus dalam waktu kurang dari delapan semester dan yang memungkinkan mereka lulus dalam waktu lebih dari delapan semester. Dengan demikian, model yang optimal dapat diterapkan untuk menangani dan meramalkan kelulusan mahasiswa baru nantinya.

Pada proses pemahaman data, langkah – langkahnya dimulai dengan mengumpulkan data awal, mendeskripsikan

**Gambar 1.** Proses data berjalan



Periksa data dan pastikan keakuratannya. 2293 record atau data mahasiswa yang menyelesaikan program studi Sarjana Teknik Informatika dengan kode A11 pada tahun masuk antara tahun 2012 hingga 2017 merupakan jumlah dataset yang digunakan untuk penelitian ini. Sekitar 1356 catatan membentuk nomor pada label 1, yang mencakup siswa yang belajar selama 38 hingga 50 bulan antara tahun masuk pada tahun 2012 dan 2017. Sebaliknya, label 2 memiliki 937 data (siswa dengan masa belajar berkisar antara 52 hingga 88 bulan , yang diterima antara tahun 2012 dan 2016).

Setelah itu, proses penyiapan data memerlukan sejumlah agenda yang digunakan untuk mengubah data mentah awal menjadi dataset akhir (data yang nantinya akan dimodelkan). Agenda tersebut antara lain mendeskripsikan kumpulan data, menentukan data, menggunakan data, menggabungkan data, membuang data yang tidak perlu, dan memformat data. Contoh kumpulan data siswa yang digunakan ditunjukkan pada Tabel 1. Data tersebut terdiri dari berbagai karakteristik siswa yang kemudian digunakan untuk membantu prosedur pelatihan model. Untuk menentukan atribut mana yang nantinya akan digunakan, akan dilakukan seleksi fitur sebelum melanjutkan ke langkah pemodelan.

Sedangkan pada sesi pemodelan ini dilakukan dengan memilih model terbaik untuk data siswa, yang meliputi langkah-langkah sebagai berikut: memilih metode pemodelan, membuat model, dan mengevaluasi model. Setelah memilih algoritma klasifikasi yang optimal dari berbagai pilihan antara lain Naïve Bayes, Decision Tree, K-NN, dan SVM, proses data mining menggunakan Random Forest.

Bahasa pemrograman Python kemudian digunakan untuk semua pengolahan dan evaluasi data, beserta diagram matriks konfusi untuk menentukan tingkat akurasi. Sementara angka-angka tersebut akan digunakan untuk menilai validitas dan keakuratan hasil turunan model.

Pengambilan keputusan atau kebijakan dalam mengkonstruksi pengetahuan tentang pola siswa dalam memprediksi kelulusan siswa berupa pohon keputusan (pohon) yang dihasilkan dari Random Forest nantinya akan digunakan dan diterapkan pada sistem akademik setelah diperoleh hasil pemodelan terbaik dengan memperoleh akurasi dan nilai pendukung yang optimal



### **3. ANALISA DAN PEMBAHASAN**

Hasil eksperimen menunjukkan bahwa teknik Forward Selection berhasil dalam memilih fitur-fitur terbaik untuk prediksi prestasi mahasiswa. Dalam setiap iterasi, fitur yang memberikan peningkatan yang signifikan dalam akurasi model Naïve Bayes dievaluasi dan dimasukkan ke dalam set fitur. Proses ini berlanjut hingga tidak ada lagi peningkatan yang signifikan dalam akurasi. Dengan menggunakan teknik ini, didapatkan sekumpulan fitur yang optimal untuk prediksi prestasi mahasiswa.

Perolehan fitur-fitur yang optimal dapat meningkatkan akurasi prediksi dan memberikan wawasan tentang faktor-faktor yang paling berpengaruh terhadap prestasi mahasiswa. Misalnya, dalam penelitian ini, mungkin ditemukan bahwa faktor usia dan IPK semester sebelumnya memiliki pengaruh yang signifikan terhadap prestasi mahasiswa. Dengan memilih fitur-fitur ini, model Naïve Bayes dapat memberikan prediksi yang lebih akurat.

Selain itu, hasil eksperimen menunjukkan kemanjuran teknik Naïve Bayes dalam meramalkan prestasi siswa. Dalam penelitian ini, probabilitas prestasi siswa berdasarkan kualitas yang sudah ada sebelumnya diperkirakan dengan menggunakan model Naïve Bayes. Premis independensi atribut digunakan dalam metode ini untuk memberikan perkiraan yang tepat dan efektif.

Oleh karena itu, makalah ini memajukan penciptaan metode untuk memprediksi prestasi siswa berdasarkan data. Institusi pendidikan dapat meningkatkan kualitas pendidikan dengan mengoptimalkan prediksi prestasi siswa dan menerapkan tindakan yang relevan melalui pemanfaatan model Naïve Bayes dan pendekatan Forward Selection.

### **4. KESIMPULAN**

Akurasi klasifikasi Naïve Bayes dapat ditingkatkan dengan menggunakan pendekatan Forward Selection untuk meminimalkan dimensi kumpulan data yang besar. Untuk meningkatkan akurasi klasifikasi Naïve Bayes pada situasi ini, fungsi pemilihan fitur Forward Selection digunakan oleh Naive Bayes untuk memilih atribut data yang sesuai dengan properti data itu sendiri. Metode Forward Selection dengan Naive Bayes telah menunjukkan efikasi dan akurasi yang unggul dalam menentukan status kelulusan siswa. Tingkat akurasinya sebesar 97,14% termasuk dalam kategori “klasifikasi sangat baik”. IPK semester empat dan status pekerjaan menjadi faktor terpenting dalam klasifikasi ini.

### **REFERENCES**

- A. M. Shahiri, W. H. (2015).** *A review on predicting student's performance using data mining techniques*, 414-422.
- A. Tekin. (2014). Early Prediction of students' grade point averages at graduation. *a data mining approach*, 207-226.
- E. Purnamasari, D. P. (2019). *The combination of Naive Bayes and Particle Swarm optimization methods of student's graduation prediction*, 112-119.
- Han, J. K. (2011). *Data mining: concepts and techniques*, 55-56.
- J. Zeniarja, K. W. (2020). *Penerapan algoritma Naive Bayes dan forward selection dalam pengklasifikasian status gizi stunting pada Puskesmas Pandanaran Semarang*, 1-9.
- John, G. H. (1995). Estimating continuous distributions in Bayesian classifiers. *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 338-345.
- Tan, P. N. (2019). *Introduction to Data Mining*. 101-105.
- Zhang, H. (2024). The optimality of Naive Bayes. 562-567.