

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini bertujuan untuk mengembangkan model prediksi peminatan mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara berdasarkan performa akademik mahasiswa. Penelitian dilakukan dengan menerapkan tiga algoritma klasifikasi, yaitu Decision Tree, Random Forest, dan XGBoost, serta beberapa teknik pendukung seperti SMOTE untuk menangani ketidakseimbangan data, *hyperparameter tuning* untuk optimasi model, dan *feature selection* untuk meningkatkan relevansi fitur. Proses pengembangan model dilakukan dengan mengikuti kerangka kerja CRISP-ML yang terdiri dari beberapa tahapan, yaitu *business understanding & data understanding*, *data preparation*, *modeling*, *evaluation*, *deployment*, serta *monitoring & maintenance*. Model yang dikembangkan dengan menerapkan kerangka kerja tersebut, kemudian dipilih yang mampu menghasilkan akurasi tertinggi. Pada penelitian ini, Random Forest adalah model yang paling optimal. Sehingga, model ini dipakai untuk diimplementasikan ke dalam sistem prediksi berbasis web.

Adapun ditemukan beberapa variabel akademik yang berpengaruh secara signifikan terhadap hasil prediksi peminatan. Penelitian ini mengidentifikasi bahwa fitur-fitur berupa nilai mata kuliah ‘Mobile Application Development’, ‘Data Visualization’, ‘IS Project Management’, ‘Seminars on Information System’, dan ‘Data Modeling’ merupakan 5 fitur dengan tingkat *importance* tertinggi yang mengindikasikan bahwa fitur-fitur ini memiliki kontribusi yang signifikan dalam membedakan kecenderungan peminatan mahasiswa. Hasil dari analisis *importance* dan seleksi fitur menunjukkan bahwa variabel-variabel tersebut memiliki hubungan kuat dengan output peminatan, terutama pada data mahasiswa angkatan 2021-2022 yang memiliki distribusi data lebih stabil dan kurikulum yang lebih homogen, serta jumlah data yang lebih banyak.

Berdasarkan hasil eksperimen dan evaluasi model, dapat disimpulkan bahwa model prediksi yang dikembangkan mampu memberikan rekomendasi peminatan mahasiswa berbasis performa akademik dengan cukup baik. Di antara ketiga algoritma yang digunakan, Random Forest terbukti menghasilkan performa paling optimal, khususnya ketika diaplikasikan pada data mahasiswa angkatan 2021-2022. Pada segmentasi data tersebut, setelah diterapkannya teknik SMOTE dan *hyperparameter tuning*, Random Forest berhasil mencapai tingkat akurasi tertinggi sebesar 78,08%. Hal ini menunjukkan bahwa pemodelan prediksi peminatan memiliki potensi untuk diimplementasikan secara nyata dalam mendukung pengambilan keputusan akademik.

Lebih lanjut, penelitian ini menunjukkan bahwa teknik pemrosesan data memiliki dampak yang signifikan terhadap performa model. Penerapan SMOTE terbukti dapat meningkatkan sensitivitas model terhadap kelas minoritas, meskipun terkadang menyebabkan penurunan akurasi keseluruhan akibat kemungkinan munculnya *noise* dari data sintetis. Sementara itu, *hyperparameter tuning* mampu memberikan hasil yang lebih stabil, meskipun efek peningkatannya terbatas apabila tidak didukung oleh kualitas data yang memadai. Selain itu, *feature selection* hanya memberikan hasil yang efektif pada algoritma tertentu, khususnya XGBoost, yang memiliki kemampuan adaptif terhadap fitur dalam jumlah terbatas.

Eksperimen yang dilakukan dengan memisahkan data berdasarkan angkatan juga menghasilkan temuan penting. Pemisahan data menjadi angkatan 2021-2022 dan 2023 memberikan peningkatan akurasi model secara signifikan, khususnya pada data angkatan 2021-2022 yang memiliki distribusi data lebih stabil dan kurikulum yang lebih homogen. Hasil ini mempertegas bahwa perbedaan karakteristik data, baik dari segi kurikulum, jumlah data, maupun penyebaran nilai, sangat memengaruhi performa model. Oleh karena itu, segmentasi data berdasarkan angkatan atau kelompok akademik perlu dipertimbangkan dalam pengembangan sistem prediksi berbasis *machine learning* di lingkungan pendidikan tinggi.

Walaupun nilai akurasi model dalam penelitian ini belum melampaui hasil penelitian-penelitian sebelumnya, hal tersebut dapat dipahami mengingat kompleksitas objek prediksi yang bersifat subjektif, keterbatasan fitur non-akademik yang digunakan, serta distribusi kelas yang tidak seimbang. Namun demikian, model yang dikembangkan dalam penelitian ini tetap dapat digunakan secara fungsional oleh mahasiswa Program Studi Sistem Informasi UMN sebagai alat bantu awal dalam menentukan peminatan. Dengan memberikan rekomendasi berbasis data akademik yang objektif, sistem ini dapat memberikan manfaat nyata dalam mendukung proses pengambilan keputusan akademik secara lebih terarah.

Dengan demikian, seluruh rumusan masalah telah terjawab dan tujuan penelitian telah tercapai, serta manfaat sistem dapat diimplementasikan dalam mendukung proses pengambilan keputusan akademik mahasiswa secara lebih terarah dan berbasis data.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan acuan untuk pengembangan sistem dan arah penelitian selanjutnya guna meningkatkan kualitas model serta relevansi implementasinya dalam dunia pendidikan:

- A. Saran Penelitian Selanjutnya
 - 1) Penambahan Fitur non-Akademik.

Penelitian selanjutnya disarankan untuk mempertimbangkan penggunaan fitur-fitur non-akademik seperti minat pribadi, motivasi belajar, pengalaman organisasi, serta hasil tes psikologi. Fitur-fitur ini dapat membantu menangkap faktor-faktor subjektif yang turut memengaruhi kecenderungan peminatan mahasiswa, sehingga dapat meningkatkan akurasi dan personalisasi model prediksi.

2) Perluasan *Dataset* dan Segmentasi Data.

Dataset yang digunakan dalam penelitian ini masih terbatas pada beberapa angkatan tertentu. Oleh karena itu, pengumpulan data dari lebih banyak angkatan atau integrasi lintas tahun akademik akan meningkatkan representativitas dan kemampuan generalisasi model, khususnya terhadap angkatan yang belum memiliki pola data yang stabil, seperti tahun 2023.

3) Eksplorasi Algoritma *Machine Learning* lain.

Selain model berbasis pohon keputusan yang digunakan dalam penelitian ini (Decision Tree, Random Forest, XGBoost), model alternatif seperti Support Vector Machine (SVM), LightGBM, CatBoost, serta metode deep learning seperti neural network dapat dieksplorasi untuk dibandingkan performanya. Eksplorasi ini penting untuk memperoleh model yang lebih optimal dan lebih adaptif terhadap berbagai bentuk dan kompleksitas data mahasiswa.

4) Perhatian terhadap Potensi Bias dari Teknik SMOTE

Penggunaan SMOTE memang membantu mengatasi ketidakseimbangan data dengan menambah data sintetis pada kelas minoritas. Namun, berdasarkan hasil penelitian, SMOTE juga berpotensi menimbulkan bias dan penurunan generalisasi model akibat *noise* dari data sintetis. Oleh karena itu, penelitian selanjutnya disarankan untuk:

- a. Menggabungkan SMOTE dengan teknik undersampling pada kelas mayoritas untuk menjaga keseimbangan data secara lebih stabil.
- b. Menggunakan varian SMOTE yang lebih adaptif, seperti SMOTEENN atau SMOTETomek, yang menggabungkan oversampling dan cleaning untuk mengurangi *noise*.
- c. Memastikan evaluasi model tidak hanya berdasarkan akurasi, tetapi juga mempertimbangkan precision, recall, dan F1-score khususnya pada kelas minoritas.
- d. Melakukan validasi silang (stratified k-fold) agar model tetap teruji dalam berbagai distribusi data.

B. Saran Pengembangan Sistem

1) Integrasi ke Sistem Akademik.

Sistem prediksi peminatan yang telah dikembangkan dapat diintegrasikan langsung ke dalam sistem informasi akademik Sistem Informasi Universitas Multimedia Nusantara, sehingga mahasiswa dapat mengakses rekomendasi peminatan secara langsung saat proses pemilihan peminatan berlangsung. Hal ini akan meningkatkan keterjangkauan dan efektivitas penggunaan sistem dalam pengambilan keputusan akademik oleh mahasiswa.

2) Evaluasi Efektivitas Sistem.

Untuk mengukur kebermanfaatan sistem secara nyata, perlu dilakukan evaluasi lanjutan terhadap hasil prediksi model dengan pilihan peminatan yang benar-benar diambil oleh mahasiswa setelah menggunakan sistem ini. Hal ini dapat menjadi dasar dalam menilai tingkat akurasi fungsional dan efektivitas sistem dalam konteks implementasi jangka panjang di Program Studi Sistem Informasi UMN.

Dengan mengikuti saran-saran di atas, diharapkan penelitian dan sistem yang telah dikembangkan dapat terus disempurnakan, baik dari segi performa model, ketepatan prediksi, maupun kebermanfaatan implementasi dalam institusi pendidikan tinggi.

