

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Perguruan tinggi merupakan suatu lembaga pendidikan yang tertinggi dalam system pendidikan nasional di semua negara. Perguruan tinggi harus dapat mengembangkan potensi peserta didik agar tumbuh dan berkembang sehingga menghasilkan lulusan yang baik dan berkualitas. Baik dalam mengelola ilmu yang didapat dan mampu menerapkan dalam dunia kerja, sehingga menjadikan sebagai manusia yang jauh lebih baik dari yang sebelumnya. Mahasiswa memiliki prestasi belajar dengan kemampuan, minat, dan bakatnya, dalam perguruan tinggi memiliki beberapa program studi yang ditawarkan pada mahasiswa (Paul Suparno, M, and Press 2015). Berdasarkan Undang – Undang Republik Indonesia nomor 12 tahun 2012 tentang perguruan tinggi mengatakan bahwa pendidikan tinggi adalah jenjang pendidikan setelah pendidikan menengah yang mencakup program diploma, program sarjana, program magister, program doktor dan program profesi, seta program spesialis, yang diselenggarakan oleh perguruan tinggi berdasarkan kebudayaan bangsa indonesia (Indonesia 2012).

Program studi pada suatu perguruan tinggi merupakan suatu penawaran bagi mahasiswa dalam pemilihan sesuai dengan kompetensinya. Fenomena dilapangan menunjukkan bahwa sebagian mahasiswa memilih program studi karena faktor dari diri sendiri yaitu minat dan bakat yang dimiliki. Sebagian lainnya, memilih jurusan berdasarkan dukungan orang tua, ajakan teman, selain itu, terdapat juga fenomena pada saat memilih suatu jurusan tanpa memberikan suatu alasan terhadap

pilihannya. Fenomena tersebut dapat menimbulkan masalah dan penyesalan dikemudian hari. Masalah tersebut diantaranya, mahasiswa tidak dapat mengikuti perkuliahan dengan baik karena tidak tertarik dibidang pilihannya, kemudian menurunkan kualitas perguruan tinggi yang tidak sesuai dengan harapan, serta kemungkinan terburuknya mahasiswa tidak dapat menyelesaikan perkuliahan dengan baik yang berujung pada *drop out (DO)* (Bawantara and Pustaka 2017). Berdasarkan data di prodi Sistem Informasi ITB STIKOM Bali pada tahun 2019 jumlah mahasiswa yang lulus hanya 60% dari jumlah mahasiswa yang masuk di tahun tersebut.

Permasalahan lainnya yang menjadi hambatan kelulusan bagi sebagian besar mahasiswa setelah mereka masuk perguruan tinggi adalah pemilihan konsentrasi yang ditawarkan pada masing-masing program studi. Konsentrasi pada program studi adalah pengkhususan studi yang diambil dalam sebuah program studi yang mempelajari suatu bidang keilmuan secara lebih spesifik dan terarah. Seperti misalnya pada perguruan tinggi di ITB STIKOM Bali di program studi Sistem Informasi memiliki pilihan konsentrasi seperti (a) *Game and Multimedia*, (b) *Intelligent System*, (c) *Enterprise System*, (d) *E-Tourism*. Pemilihan konsentrasi program studi biasanya dilakukan pada perkuliahan semester V, namun tergantung dari kebijakan program studi terkait. Mahasiswa yang akan memasuki Semester V harus memikirkan matang-matang konsentrasi apa yang akan dipilih karena keputusannya nanti akan berpengaruh terhadap proses belajar dan masa depannya. Tidak adanya persiapan secara matang ketika akan memilih konsentrasi yang akan ditempuh menyebabkan beberapa mahasiswa tidak dapat mengikuti pelajaran

dengan baik dan merasa salah dalam memilih program studi yang akhirnya akan berdampak negatif bagi mahasiswa tersebut. Maka dari itu, berdasarkan permasalahan yang ada, peneliti tertarik untuk melakukan penelitian mengenai pengembangan model rekomendasi dalam memilih konsentrasi program studi dengan menggunakan *Multiclassifier Voting* dan teknik *SMOTE (Synthetic Minority Over-Sampling Technique)*. Semua informasi yang terkait pada pengembangan model dapat dimanfaatkan untuk mengetahui pola baru dari sebuah data. Pola baru ini dapat digunakan untuk tujuan yang bermanfaat misalnya pengklasifikasian. Pembentukan pola baru tersebut dapat dilakukan dengan pengolahan atau penggalian data menggunakan metode yang ada pada *Machine Learning*. *Machine Learning* merupakan bagian dari bidang kecerdasan buatan yang mempunyai fokus pada penerapan algoritma maupun metode tertentu untuk melakukan prediksi, pengenalan pola, dan klasifikasi. Pada umumnya tahapan pada *Machine Learning* adalah koleksi data, *preprocessing* data, pembentukan model, tahap pelatihan, serta tahap uji.

Pada bidang akademik, penelitian mengenai klasifikasi berbasis *machine learning* sudah pernah dilakukan seperti penelitian berikut, "*Classification of Students Based on Quality of Life and Academic Performance by using Support Vector Machine (SVM)*" pada penelitian ini bertujuan untuk mengetahui pengaruh kualitas hidup dengan performa akademik siswa dengan menggunakan metode *SVM*. Hasil penelitian menunjukkan kualitas hidup yang mempengaruhi akademik adalah akurasi 80% untuk kesehatan fisik, 86,6% untuk psikologi, 80% for hubungan sosial, 93,33% untuk lingkungan, 80% untuk kualitas hidup (Raihana and

Farah Nabilah 2018). Berikutnya penelitian mengenai “*Predicting Academic Performance of Students using a Hybrid Data Mining Approach*” yang bertujuan untuk melakukan evaluasi performa akademik siswa dengan metode klasifikasi dan *clustering*. Hasil dalam penelitian ini dengan kombinasi metode *classification* dengan *clustering* menunjukkan akurasi hasil prediksi performa lebih meningkat (Francis and Babu 2019). Selanjutnya, penelitian dengan topik “*Normalized Multiple Features Fusion Based on PCA and Multiple Classifiers Voting in CT Liver Tumor Recognition*” membahas mengenai sistem *Computer-Aided Diagnosis (CAD)* untuk mendeteksi penyakit tumor secara otomatis, cepat dan akurat. Metode yang digunakan yaitu *Multi Classifier voting* dengan hasil performa akurasi mencapai 100% (Anter and Hassenian 2018). Selain itu, penelitian dengan judul “*Comparison of Bagging and Voting Ensemble Machine Learning Algorithm as a Classifier*” dalam penelitian ini menjelaskan mengenai perbandingan metode machine learning *Ensemble Bagging dan Ensemble Voting*. Dimana kedua metode ini merupakan jenis metode yang mengkombinasikan beberapa jenis metode klasifikasi untuk mendapatkan hasil akurasi yang lebih baik. *Dataset* yang digunakan dalam penelitian ini adalah *Iris dataset* dengan 150 *instance* dan 5 *feature*. Dari hasil uji coba yang dilakukan dalam penelitian ini menunjukkan rata-rata akurasi yang didapatkan adalah 95,56% (Kabari and Onwuka 2019). Berdasarkan informasi yang ada maka, peneliti melakukan penelitian lanjutan pada bidang akademik seperti permasalahan yang telah peneliti jelaskan sebelumnya. Adapun penelitian lanjutan ini mengenai pengembangan model rekomendasi

pemilihan konsentrasi pada program studi dengan kombinasi metode *Multi Classifier Voting* dengan teknik *SMOTE*.

Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu teknik *over-sampling* yang mampu meningkatkan akurasi dari pengklasifikasi untuk kelas *minor*. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* dengan cara *sampling* ulang sampel kelas minoritas. Pada penelitian yang akan dilakukan, peneliti menerapkan metode ini untuk menyeimbangkan *dataset* dari nilai mahasiswa karena data yang ada tidak rata (*imbalanced*) disetiap kelasnya. Sedangkan proses klasifikasi pada data dengan *dataset* yang tidak seimbang merupakan masalah utama pada bidang *machine learning*. Jika bekerja pada data tidak seimbang, hampir semua algoritma klasifikasi akan menghasilkan akurasi yang jauh lebih tinggi untuk kelas mayoritas dari pada kelas minoritas.

Selain itu, peneliti juga menggunakan *Ensemble Learning* yang merupakan salah satu metode pada *machine learning* dimana cara kerjanya dengan mengkombinasikan beberapa algoritma *learning* untuk menghasilkan prediksi yang dapat meningkatkan akurasi. Ada beberapa jenis *Ensemble Learning*, yaitu *Bagging*, *Boosting*, *Stacking*, dan *Voting*. Pada penelitian ini menggunakan *Ensemble Learning* jenis *voting* dengan *hard voting* yang memiliki cara kerja dengan memilih label yang dominan dari setiap prediksi label yang didapatkan dari masing-masing algoritma atau proses perhitungan nilai akhir dari *Majority vote* dilakukan dengan cara mengitung nilai bobot tertinggi dari hasil masing-masing metode klasifikasi. Metode ini dipilih karena cara kerjanya yang sederhana tetapi

menghasilkan hasil yang maksimal. Dimana algoritma yang akan dikombinasikan untuk proses Ensemble Learning yaitu metode klasifikasi (*SVM (Support Vector Machine)*), *DT (Decision Tree)*, dan *RF (Random Forest)*.

Metode klasifikasi Random Forest bertujuan menangani dataset yang beragam seperti fitur pada nilai matakuliah. Lalu pada metode Support Vector Machine digunakan mendukung klasifikasi untuk pengolahan data berdasarkan nilai matakuliah umum dengan matakuliah khusus persemester yang menyangkut dengan pemilihan konsentrasi. Pada metode Decision Tree menangani masalah dataset distribusi yang tidak teratur.

Ada beberapa kriteria penting yang dijadikan peneliti untuk mempertimbangkan saat memilih analisis data atau metode pemodelan klasifikasi dalam penelitian ini. Salah satu kriteria terpenting ketika memilih metode adalah akurasi. Keakuratan menunjukkan sejauh mana model mampu melakukan prediksi yang benar. Berdasarkan bentuk datanya, metode yang dipilih dalam penelitian ini harus dapat memberikan skor akurasi yang lebih tinggi atau kesalahan prediksi yang lebih rendah, yang umumnya merupakan pilihan yang lebih baik. Selain itu, interpretabilitas model juga menjadi pertimbangan penting, terutama ketika diperlukan pemahaman tentang hubungan antara karakteristik dan output. Kecepatan pemrosesan, stabilitas model, dan skalabilitas juga merupakan faktor penting saat memilih metode terbaik dalam penelitian ini.

Kemudian data yang digunakan dalam penelitian ini merupakan nilai matakuliah dari mahasiswa angkatan 2017, 2018, dan 2019 pada Prodi Sistem Informasi di ITB STIKOM Bali yang telah lulus. Data yang ada ini akan melalui

proses *preprocessing*, berikutnya masuk ke proses klasifikasi. Setelah didapatkan label final, proses selanjutnya yaitu melakukan pengujian menggunakan *dataset testing*. Hasil akhir yang didapatkan yaitu nilai akurasi yang berasal dari *confusion matrix* dengan perhitungan berdasarkan nilai *False Positive (FP)*, *True Positive (TP)*, *False Negative (FN)*, dan *True Negative (TN)*.

Secara khusus, dalam penelitian ini membahas mengenai pengembangan model rekomendasi pemilihan konsentrasi pada program studi dengan kombinasi metode *Multi Classifier Voting* dengan teknik *SMOTE*. Penggunaan teknik *SMOTE* dalam penelitian ini untuk proses *oversampling* karena faktanya data yang ada tidak seimbang (*Imbalanced*) pada setiap kelasnya. Oleh karena itu, diperlukan proses *oversampling* untuk menyeimbangkan data sehingga akan dapat meningkatkan hasil akurasi prediksi dari proses klasifikasi. Kemudian dalam proses *learning* model menggunakan metode *Ensemble Learning* dengan mengkombinasikan beberapa algoritma klasifikasi yang memiliki karakteristik pembelajaran yang berbeda-beda atau *Multi Classifier* untuk menghasilkan prediksi yang dapat meningkatkan akurasi. Dalam penghitungan akurasi, *Multi Classifier Ensemble Learning* menggunakan 3 jenis algoritma klasifikasi, dimana label prediksi dari ketiga algoritma tersebut akan digabungkan menggunakan *majority vote* untuk mendapatkan hasil prediksi yang akan diuji menggunakan *dataset testing*.

Diharapkan kombinasi metode *Multi Classifier* dengan teknik *oversampling* ini mampu memberikan hasil performa terbaik. Terakhir dalam proses pengujiannya akan dibedakan hasil akurasi yang didapatkan berdasarkan pada *dataset* yang telah

diproses *oversampling* dengan *SMOTE* dengan *dataset* yang tidak diproses *oversampling (dataset original)*.

1.2 Identifikasi Masalah

Sesuai penjelasan yang telah penulis uraikan pada latar belakang maka, penulis mengidentifikasi masalah dalam penelitian ini sebagai berikut :

1. Tidak adanya persiapan secara matang dari mahasiswa ketika akan memilih konsentrasi pada program studi yang menyebabkan beberapa mahasiswa tidak dapat mengikuti pelajaran dengan baik dan merasa salah dalam memilih program studi yang akhirnya akan berdampak negatif bagi mahasiswa tersebut atau mahasiswa tidak dapat menyelesaikan perkuliahan dengan baik yang berujung pada *drop out (DO)*.
2. Dengan adanya mahasiswa yang tidak dapat menyelesaikan perkuliahan sesuai dengan batas waktu untuk penyelesaian tingkat sarjana (4 tahun) maka, akan berpengaruh pada kualitas perguruan tinggi.
3. Adanya rangkain dampak yang buruk untuk mahasiswa hingga perguruan tinggi yang disebabkan dari masalah kesalahan pemilihan konsentrasi ini, maka dibutuhkan solusi.

1.3 Pembatasan Masalah

Untuk memudahkan dalam penelitian, penulis membuat batasan masalah yang akan dibahas sebagai berikut.

1. Penelitian ini menggunakan data nilai matakuliah dari mahasiswa Angkatan 2017, 2018 dan 2019 yang telah lulus pada Perguruan Tinggi.
2. Batasan fitur dari datasets untuk penelitian pembuatan model rekomendasi pemilihan konsentrasi program studi hanya nama matakuliah dari semester I sampai IV.
3. Menggunakan teknik *SMOTE* untuk mengatasi *imbalanced* data atau data tidak seimbang.
4. Metode yang digunakan dalam pemilihan Konsentrasi Program Studi yaitu metode *Ensemble Multiclassifier Voting*.
5. Proses pengolahan data didukung menggunakan bahasa pemrograman *Python* dan *library Scikit-Learn (Sk-Learn)*.
6. Luaran penelitian ini berupa model untuk rekomendasi pemilihan konsentrasi pada program studi.

1.4 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, penulis merumuskan masalah penting dalam penelitian ini sebagai berikut.

1. Merancang model rekomendasi yang efektif untuk pemilihan konsentrasi pada program studi yang sesuai dengan nilai matakuliah mahasiswa.

2. Mengatasi jumlah *dataset* yang tidak rata atau tidak seimbang (*imbalanced*) disetiap label dari nilai mahasiswa untuk meningkatkan hasil akurasi prediksi.
3. Menentukan metode *learning* yang sesuai untuk proses klasifikasi data dalam rekomendasi pemilihan konsentrasi pada program studi.
4. Mengevaluasi model rekomendasi pemilihan konsentrasi program studi untuk menentukan tingkat performa hasil pengujian dan efektivitasnya dalam membantu mahasiswa membuat keputusan yang tepat.

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Mengembangkan model rekomendasi yang dapat memprediksi konsentrasi program studi yang sesuai dengan nilai matakuliah dan potensi calon mahasiswa.
2. Menerapkan teknik *oversampling* untuk mengatasi ketidakseimbangan data yang mungkin ada dalam dataset nilai matakuliah yang digunakan dalam pembuatan model.
3. Menggunakan metode *Ensemble Learning* untuk meningkatkan hasil prediksi dan kehandalan dari model rekomendasi pemilihan konsentrasi.
4. Menggunakan *Confusion Matrix* untuk melakukan pengukuran tingkat performa akurasi, presisi, recall, dan F1 score dari model tersebut sebagai alat evaluasi yang dapat membantu memahami sejauh mana model rekomendasi dapat memprediksi dan merekomendasikan konsentrasi program studi dengan tepat kepada mahasiswa.

1.6 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun manfaat praktis, sebagai berikut.

1.6.1 Manfaat teoritis

Manfaat teoritis yaitu hasil penelitian ini diharapkan dapat memberikan manfaat untuk pembaca atau sebagai acuan peneliti selanjutnya sebagai media referensi untuk mengembangkan ilmu pengetahuan yang berkaitan dengan *machine learning*.

1.6.2 Manfaat praktis

Untuk manfaat praktis, hasil penelitian ini diharapkan dapat memberikan manfaat bagi pihak-pihak sebagai berikut.

1. Bagi penulis

- Dapat mengembangkan pengetahuan penulis tentang proses *machine learning*.

2. Bagi institusi pendidikan

- Memberikan informasi mengenai pengklasifikasian data menggunakan metode *Multiclassifier* dan teknik *SMOTE* dalam pemilihan konsentrasi program studi.
- Dapat memberikan gambaran hasil analisa yang tepat bagi mahasiswa dalam pemilihan Konsentrasi Program Studi.

3. Bagi pihak terkait secara umum

- Dapat memberikan pengetahuan tentang penerapan *machine learning* dalam pemilihan Konsentrasi Program Studi yang tepat.

4. Bagi peneliti lainnya

- Dapat digunakan sebagai rujukan atau referensi dalam mengembangkan penelitian sejenis pada kesempatan lainnya.
- Menambah keilmuan Statistika dibidang klasifikasi *data mining machine learning*.

