Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 5 No. 5 (2021) 853 - 862 ISSN Media Elektronik: 2580-0760

Studi Komparasi Model Klasifikasi Berbasis Pembelajaran Mesin untuk Sistem Rekomendasi Program Studi

Rio Rizki Aryanto[1], Ahmad R Pratama[2*], Lizda Iswari[3]
[1]Program Studi Informatika – Program Magister, Fakultas Teknologi Industri, Universitas Islam Indonesia
[2,3]Jurusan Informatika, Fakultas Teknik Industri, Universitas Islam Indonesia
[1]rio.aryanto@students.uii.ac.id, [2]ahmad.rafie@uii.ac.id*, [3]lizda.iswari@uii.ac.id

Abstract

Selecting a major can be quite difficult for prospective college students. The choice may have an effect not only on their academic life, but also on their career path. Due to some restrictions as the impact of the COVID-19 pandemic, universities must find novel ways to reach prospective students and assist them in choosing their majors, one of which is a college major recommendation system. This system can assist prospective students in determining the most appropriate majors for them based on data from the current students. Unlike other existing systems that employ either a rule-based or fuzzy model, this study employs a machine learning approach using data from undergraduate students at Universitas Islam Indonesia. This paper aims to compare several clustering models (i.e., K-means, Agglomerative, Birch, and DBSCAN) for the purpose of categorizing current students, to which the results will be used for classification purposes using various approaches (i.e., single stage vs. multistage), algorithms (i.e., multinomial logistic regression, random forest, and support vector machine), and scenarios (i.e., with or without GPA-based label). Our findings indicate that the K-means model outperformed all other clustering models and that the single stage with random forest classification model performed the best across all scenarios.

Keywords: comparative study, recommendation system, major selection, machine learning classification models, single stage model, multistage model

Abstrak

Pemilihan program studi di perguruan tinggi adalah salah satu tantangan bagi calon mahasiswa baru. Pilihan yang diambil dapat berdampak pada kehidupan akademis dan jalur karier mereka. Karena beberapa pembatasan sebagai dampak dari pandemi Covid-19, perguruan tinggi harus menemukan cara inovatif untuk menjangkau calon mahasiswa baru dan membantu mereka dalam memilih program studi, salah satunya dalam bentuk sistem rekomendasi program studi. Sistem ini dapat membantu calon mahasiswa baru dalam menentukan program studi yang paling sesuai untuk mereka berdasarkan data dari mahasiswa aktif di masing-masing program studi. Tidak seperti sistem lain yang menggunakan model berbasis aturan atau fuzi, penelitian ini menggunakan pembelajaran mesin *(machine learning)* dengan data dari mahasiswa program sarjana di Universitas Islam Indonesia untuk membangun sistem rekomendasi tersebut. Penelitian ini bertujuan untuk membandingkan beberapa model klastering (K-means, Agglomerative, Birch, dan DBSCAN) untuk mengkategorikan mahasiswa aktif yang hasilnya akan digunakan untuk proses klasifikasi menggunakan berbagai metode (*single stage* vs. *multistage*), algoritme (*multinomial logistic regression*, *random forest*, dan *support vector machine*), dan skenario (dengan atau tanpa label berbasis IPK). Temuan dari penelitian ini menunjukkan bahwa model K-means mengungguli semua model klastering lainnya sementara model klasifikasi *single stage* dengan *random forest* memiliki kinerja terbaik di semua skenario.

Kata kunci: studi komparasi, sistem rekomendasi, pemilihan program studi, model klasifikasi *machine learning*, model *single stage*, model *multistage*

1. Pendahuluan

Melanjutkan studi pendidikan perguruan tinggi merupakan salah satu tujuan bagi siswa Sekolah Menengah Atas (SMA) dan sederajat. Tantangan dimulai ketika calon mahasiswa harus memilih program studi yang akan dituju. Proses pengambilan keputusan tersebut dipengaruhi oleh faktor internal seperti kondisi ekonomi, letak geografis dan prestasi akademik maupun faktor eksternal seperti kualifikasi perguruan tinggi,

Diterima Redaksi: 09-08-2021 | Selesai Revisi: 02-10-2021 | Diterbitkan Online: 25-10-2021

kualitas kompetitor dan masih banyak lagi. Keputusan pemilihan program studi sendiri dapat berimbas pada kelancaran masa studi maupun jenjang karir setamat studi. Hasil survey Indonesian Career Centre Network (ICNN) tahun 2017 yang dilansir pada [23] menunjukkan sebanyak 87% mahasiswa Indonesia merasa telah salah mengambil jurusan kuliah atau program studi. Jika dibandingkan dengan data Badan Pusat Statistik (BPS) tahun 2017, dapat diasumsikan sebanyak 8 dari 10 mahasiswa Indonesia merasakan hal tersebut. [14] mengutip survey yang sama namun memaparkan informasi lain yaitu sebanyak 71.7% pekerja di Indonesia memiliki jenis profesi yang tidak berkaitan dengan latar pendidikannya. Survey ICNN tersebut menunjukkan bahwa pemilihan program studi masih menjadi tantangan besar yang perlu mendapatkan perhatian. Oleh karenanya, penting bagi calon mahasiswa untuk dapat memilih program studi yang tepat sekaligus cocok dengan karakter atau minat mereka.

Pandemi Covid-19 membuat tantangan terkait pemilihan program studi menjadi semakin berat. Adanya pembatasan sosial berpotensi membuat calon mahasiswa kesulitan mengakses informasi terkait program studi atau universitas yang diinginkan. Bagi universitas, kondisi ini menuntut mereka untuk mengambil langkah inisiatif dan inovatif guna menjangkau calon mahasiswa. Salah satu inovasi yang dapat dilakukan adalah dengan menginisiasi sistem rekomendasi pemilihan program studi. Sistem yang dapat diakses oleh calon mahasiswa kapanpun dan dimanapun. Sistem rekomendasi dapat membantu mahasiswa untuk mengenali program studi yang cocok dengan mereka.

Di Indonesia sendiri ditemukan beberapa penelitian dengan tema sistem rekomendasi serupa. Akan tetapi kebanyakan sistem rekomendasi yang dikembangkan masih menggunakan model berbasis aturan (*rule based*) seperti pada [12], [15] dan [18] atau model fuzzy seperti pada [20]. Masih belum banyak sistem rekomendasi yang mengimplementasikan model berbasis pembelajaran mesin (*machine learning*).

Selain bertujuan untuk mengembangkan sistem rekomendasi, penelitian juga bertujuan untuk melihat bagaimana sains data khususnya terkait model *machine learning* dapat diimplementasikan pada sistem rekomendasi. Oleh karenanya, penelitian menggunakan model *machine learning* tidak tersupervisi yaitu model *clustering* sebagai model untuk mengelompokkan mahasiswa. Proses pengelompokkan dilakukan agar model klasifikasi pada sistem dapat dilatih menggunakan kelompok mahasiswa yang tepat. Implementasi model *clustering* juga merupakan pembaruan dari penelitian sebelumnya (*preliminary study*) oleh [22]. [22] mengembangkan sistem rekomendasi pemilihan program studi dengan mengimplementasikan model klasifikasi *single stage*. Alih-alih menggunakan model *clustering* sebagai alat bantu, proses pengelompokkan mahasiswa dilakukan secara sederhana dengan melihat nilai Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) dan status mahasiswa. Hal tersebut menjadi ide untuk melakukan pembaruan pada penelitian kali ini dengan mengimplementasikan model *clustering*.

Pembaruan juga dilakukan dengan melakukan studi komparasi. Penelitian mengembangkan dua jenis model klasifikasi yaitu model *single stage* dan *multistage*. Penggunaan model *multistage* dilakukan berdasarkan hasil studi literatur pada tema terkait. Penggunaan model *multistage* pada penelitian [24] terbukti memberikan performa yang lebih baik dibandingkan model *single stage*, terutama pada aspek sensitivitas model. Model *multistage* [6] bahkan mendapatkan akurasi tertinggi sebesar 98.8% setelah mengimplementasikan 8 kombinasi model berbeda. Penelitian [26] mendapatkan temuan bahwa model *multistage* mempunyai kemampuan yang lebih baik untuk mengenali kejadian (*event*) dengan jumlah kemunculan kecil dibandingkan model *single stage*. Temuan tersebut yang menjadi dasar dari penelitian ini untuk melakukan komparasi antara model klasifikasi *single stage* dengan model *multistage*.

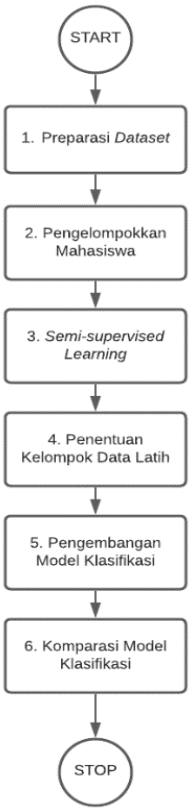
Studi komparasi juga dilakukan pada tahap awal terkait data latih yang digunakan oleh model klasifikasi. Penelitian menyiapkan 2 skenario preparasi data. Skenario pertama mempertahankan tahap preparasi yang dilakukan pada [22], sedangkan skenario kedua adalah skenario alternatif. Hal ini untuk melihat apakah pendekatan yang dilakukan pada *dataset* penelitian mampu memberikan improvisasi dibandingkan penelitian sebelumnya. Studi komparasi diakhiri dengan membandingkan tiga model klasifikasi yaitu model *single stage*, *multistage* dan model klasifikasi [22]. Sama seperti pada *preliminary study*, penelitian ini juga membandingkan 3 metode klasifikasi berbeda yaitu Multinomial Logictic Regression (MLR), Random Forest (RF) dan Support Vector Machine (SVM). Hasil komparasi dapat memberikan kesimpulan model klasifikasi terbaik dan data latih seperti apa yang dapat digunakan untuk mendukung performa model tersebut.

2. Metode Penelitian

Penelitian menggunakan tiga metode klasifikasi yaitu MLR, RF dan SVM yang diimplementasikan baik pada model *single stage* maupun *multistage*. Pemilihan tiga metode tersebut dilakukan berdasarkan temuan pada penelitian [13]. Pada penelitian tersebut diketahui bahwa metode klasifikasi berbasis *machine learning* seperti MLR, RF dan SVM cocok digunakan pada *dataset* dengan ukuran yang kecil. Metode *machine learning* tersebut bekerja lebih baik dibandingkan model dengan pendekatan *deep learning*. Penelitian ini sendiri menggunakan data mahasiswa jenjang sarjana di Universitas Islam Indonesia (UII). Total terdapat 2.908

DOI: https://doi.org/10.29207/resti.v5i5.3392

data mahasiswa yang digunakan pada penelitian. Ukuran *dataset* tersebut menurut peneliti belum membutuhkan metode *deep learning* dalam pengembangan model klasifikasi. Sistem rekomendasi pemilihan program studi nantinya akan mengimplementasikan model klasifikasi dengan metode terbaik. Penelitian dilakukan dalam beberapa tahap penelitian seperti terlihat pada Gambar 1. Penelitian diawali dengan preparasi dataset, pengelompokkan mahasiswa, pengembangan model klasifikasi dan diakhiri dengan studi komparasi.



Gambar 1. Diagram Alir Langkah-Langkah Penelitian

2.1. Preparasi *Dataset*

Data mahasiswa sarjana yang digunakan pada penelitian terbagi menjadi dua jenis. Masing-masing *dataset* tersebut akan digunakan baik secara terpisah maupun bersamaan tergantung kebutuhan penelitian. *Dataset* pertama akan disebut dengan **DB1**, adalah *dataset* dengan data terkait informasi mahasiswa semasa bangku SMA. Contoh informasi yang terdapat pada DB1 antara lain adalah jenis SMA, jurusan SMA, nilai mata pelajaran tiap semester, Nilai Ebtanas Murni (NEM) dan masih beberapa data lainnya. DB1 juga memiliki data mahasiswa semasa duduk di bangku perguruan tinggi. Namun informasi tersebut dapat dikatakan hanya informasi ringkas seperti status mahasiswa, nilai IPK, jumlah SKS dan jenis program studi yang ditempuh.

Dataset terakhir akan disebut dengan **DB2**. Berbeda dengan *dataset* sebelumnya, DB2 memiliki informasi yang lebih spesifik. Data DB2 didapatkan dari mahasiswa program studi Informatika UII. Informasi pada *dataset* tersebut menunjukkan prestasi atau nilai tiap mata kuliah yang diambil oleh tiap mahasiswa selama masa studi di program studi Informatika. Karena data poin yang dimiliki berbeda, tahap preparasi *dataset* yang dilakukan juga tidak bisa disamakan.

Pada DB1 akan disiapkan dua skenario preparasi data yaitu **Skenario-A** dan **Skenario-B**. Skenario-A akan mempertahankan apa yang telah dilakukan pada [22]. Skenario tersebut akan menyeleksi mahasiswa berdasarkan nilai IPK, jumlah SKS dan status mahasiswa. Data yang akan dipertahankan pada *dataset* adalah mahasiswa dengan nilai IPK minimal 3.00, jumlah SKS minimal 80 dan status mahasiswa aktif atau lulus. Sebaliknya, Skenario-B adalah skenario yang tidak melakukan seleksi data tersebut. Meskipun menerapkan proses seleksi yang berbeda, kedua skenario tersebut menggunakan teknik agregasi yang sama pada proses preparasinya. Agregasi data dilakukan untuk mendapatkan nilai rerata tiap mata pelajaran SMA pada masing-masing mahasiswa.

Preparasi *dataset* DB2 akan dilakukan hanya dengan menggunakan teknik seleksi data. Perlu diketahui DB2 sendiri memiliki data poin seperti nilai mata kuliah, jenis mata kuliah, jenis kurikulum dan jenis semester pada mahasiswa program studi Informatika. Seleksi pada tahap preparasi akan dilakukan berdasarkan jenis kurikulum, jenis semester dan kelengkapan data pada tiap mata kuliah. Seleksi pertama dilakukan berdasarkan jenis kurikulum. Penelitian akan menggunakan mahasiswa Informatika yang mendapatkan satu dari dua jenis kurikulum terbaru yaitu kurikulum KD-2016 atau KD-2020. Digunakannya dua kurikulum tersebut dikarenakan KD-2016 dan KD-2020 mempunyai komposisi mata kuliah yang tidak jauh berbeda. Seleksi dilanjutkan dengan menggunakan jenis semester. Hanya akan digunakan data nilai mata kuliah pada periode tahun pertama mahasiswa. Artinya akan digunakan data pada periode semester 1 atau 2. Hal ini dikarenakan pada tahun pertama semua mahasiswa mendapatkan mata kuliah seragam sehingga aspek variabilitasnya masih dapat dikendalikan. Seleksi terakhir pada preparasi DB2 dilakukan dengan melihat kelengkapan data terkait nilai mata kuliah. Peneliti menemukan fakta bahwa tidak semua mata kuliah memiliki data yang lengkap. Banyak diantaranya yang memiliki data kosong (*missing value*). Hal tersebut dapat memengaruhi performa model pada tahap pengelompokkan selanjutnya, sehingga peneliti memutuskan untuk mengambil mata kuliah dimana 80% datanya tersedia.

Setelah melalui tahap preparasi maka *dataset* DB1 dan DB2 sudah siap untuk digunakan pada tahap penelitian selanjutnya. Data DB1 akan banyak digunakan dalam pengembangan model klasifikasi sedangkan data DB2 akan digunakan pada model klastering.

DOI: https://doi.org/10.29207/resti.v5i5.3392

2.2. Pengelompokkan Mahasiswa

Agar mampu memberikan hasil rekomendasi yang baik maka model klasifikasi pada sistem rekomendasi perlu dilatih menggunakan kelompok mahasiswa yang tepat. Alih-alih menggunakan keseluruhan data, perlu dilakukan seleksi untuk menentukan kelompok mahasiswa mana yang lebih representatif. Penelitian ini akan menggunakan model *clustering* sebagai alat bantu untuk mengelompokkan mahasiswa. Model tersebut akan mempelajari karakteristik mahasiswa berdasarkan nilai mata kuliahnya untuk kemudian mengelompokkannya ke dalam beberapa kelompok atau klaster. Dari hasil tersebut dapat dilakukan interpretasi secara manual untuk memutuskan kelompok mana yang akan digunakan pada tahap selanjutnya.

Model *clustering* akan menggunakan data nilai mata kuliah sebagai variabel prediktornya. Tantangan muncul dikarenakan pada penelitian data terkait nilai mata kuliah hanya dimiliki oleh mahasiswa program studi Informatika. Data tersebut tersimpan pada *dataset* DB2. Oleh sebab itu, model *clustering* hanya dapat diimplementasikan pada mahasiswa program studi Informatika. Sebagai alternatif, pada program studi selain Informatika, proses pengelompokkan mahasiswa akan dilakukan dengan pendekatan *semi-supervised learning* yang dilakukan pada tahap selanjutnya.

Beberapa metode *clustering* seperti K-means, Agglomerative, Birch dan *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) akan digunakan pada tahap ini. Selanjutnya metode tersebut akan dikomparasi guna mendapatkan metode mana yang terbaik. Komparasi dilakukan berdasarkan nilai koefisien Silhoutte.

2.3. *Semi-supervised Learning*

Tahap ini bertujuan untuk melakukan pengelompokkan pada mahasiswa program studi non-Informatika. Tahap ini diperlukan mengingat penelitian tidak memiliki data nilai mata kuliah pada program studi selain Informatika, sehingga proses pengelompokkan tidak dapat dilakukan menggunakan bantuan model *clustering*.

Semi-supervised learning diawali dengan mengembangkan model klasifikasi *single stage* yang dilatih menggunakan data mahasiswa Informatika. Variabel prediktor yang digunakan pada model *semi-supervised learning* adalah data karakteristik semasa SMA seperti jenis kelamin, hobi, jenis SMA, jurusan SMA, nilai rerata mata pelajaran seperti matematika, bahasa Indonesia, bahasa Inggris, fisika, kimia, biologi, sejarah, ekonomi, geografi, agama, dan ketrampilan/keahlian kejuruan sedangkan variabel target akan menggunakan hasil model *clustering* sebelumnya. Artinya, hasil pada model *clustering* akan dianotasikan atau dilabelkan secara manual pada untuk kemudian diasumsikan sebagai label kelompok atau kelas mahasiswa. Melihat variabel yang digunakan pada model *semi-supervised learning*, maka diperlukan data gabungan dari *dataset* DB1 dan DB2.

Tiga metode klasifikasi yaitu MLR, RF dan SVM kembali dikomparasikan pada pengembangan model klasifikasi *semi-supervised learning*. Model klasifikasi terbaik nantinya akan digunakan untuk memprediksi label kelas mahasiswa program studi non-Informatika. Hal ini mungkin dilakukan mengingat mahasiswa program studi lain memiliki atribut yang sama seperti varibel prediktor model klasifikasi *semi-supervised learning*.

Perlu menjadi catatan bahwa model klasifikasi *semi-supervised learning* yang digunakan kemungkinan besar memiliki bias. Dikarenakan model dilatih menggunakan data mahasiswa Informatika yang secara logika lebih merepresentasikan rumpun ilmu sains atau teknik. Lebih lanjut model digunakan untuk memprediksi pada program studi rumpun ilmu medis maupun sosial. Untuk mengetahui seberapa besar efek bias dari pendekatan *semi-supervised learning* maka perlu dilakukan analisis terpisah lainnya.

2.4. Penentuan Kelompok Data Latih

Setelah melakukan pengelompokkan seluruh program studi menggunakan model *clustering* maupun model *semi-supervised learning*, tahap selanjutnya adalah menentukan kelompok mahasiswa mana yang akan digunakan sebagai data latih pada model klasifikasi sistem rekomendasi pemilihan program studi. Penentuan kelompok atau kelas mahasiswa tersebut dapat dilakukan dengan melihat sebaran data baik nilai mata kuliah maupun nilai mata pelajaran pada masing-masing kelas.

Khusus mahasiswa program studi Informatika, interpretasi kelompok dapat dilakukan dengan melihat sebaran nilai mata kuliah sedangkan untuk program studi lainnya, interpretasi dapat dilakukan berdasarkan sebaran nilai mata pelajaran SMA. Interpretasi dilakukan secara manual oleh peneliti sehingga sangat dimungkinkan terdapat unsur subjektivitas pada penilaiannya.

2.5. Pengembangan Model Klasifikasi

Penelitian menggunakan pendekatan *single stage* dan *multistage* dalam mengembangkan model klasifikasi sistem rekomendasi. Kedua model tersebut memiliki variabel prediktor dan variabel target yang sama. Sebagai variabel prediktor digunakan jenis kelamin, jenis & jurusan SMA, hobi, nilai rerata masing-masing mata pelajaran matematika, bahasa Indonesia, bahasa Inggris, fisika, biologi, kimia, geografi, sejarah, ekonomi, agama dan ketrampilan/keahlian kejuruan. Sedangkan untuk variabel target akan digunakan jenis program studi. Secara intuisi model klasifikasi tersebut mirip dengan model *semi-supervised learning* namun menggunakan variabel target yang berbeda.

DOI: https://doi.org/10.29207/resti.v5i5.3392

Pada model *single stage*, prediksi dilakukan secara bersamaan tanpa membedakan segmen data tertentu. Artinya prediksi dilakukan dalam sekali jalan. Berbeda dengan model *multistage* dimana prediksi dilakukan secara bertahap melalui beberapa fase atau *stage*. Skema pada model *multistage* dapat dilihat pada Gambar 2.

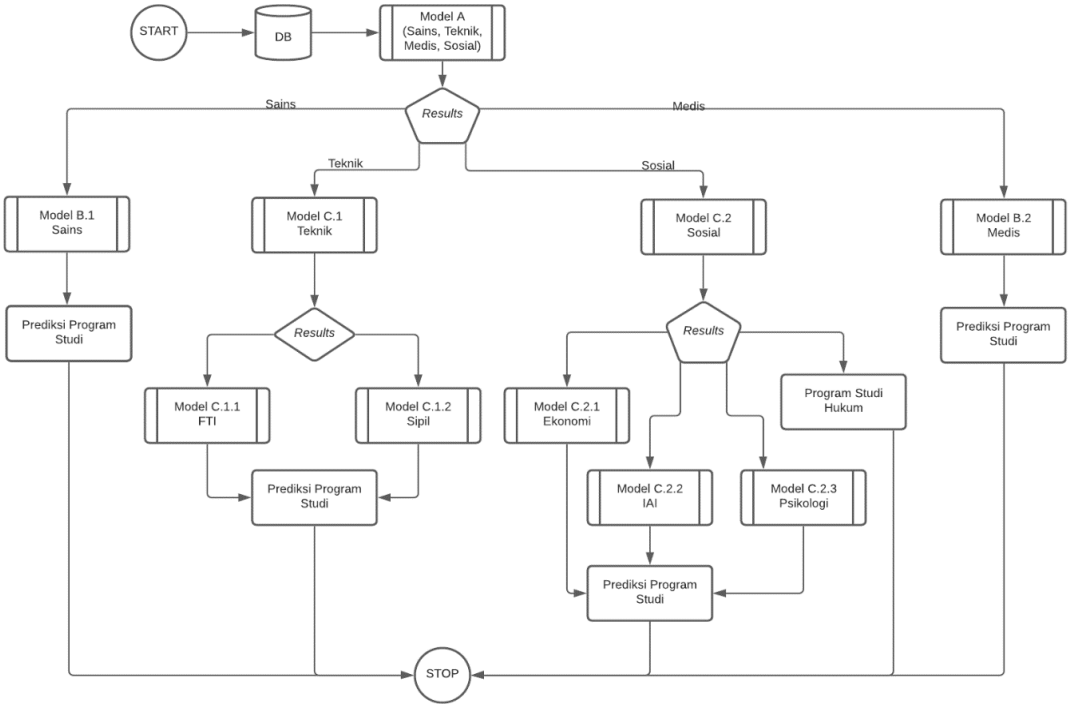
Stage I model *multistage* akan memprediksi data input dalam kategori rumpun ilmu antara sains, teknik, sosial atau medis. Jika prediksi menunjukkan rumpun ilmu sains atau medis maka prediksi program studi dapat dilakukan langsung pada *stage* II, namun jika masuk ke dalam kategori teknik atau sosial maka akan diprediksi terlebih dahulu jenis fakultasnya. Pada *stage* terakhir atau *stage* III akan diprediksi program studi pada level fakultas teknik maupun sosial.

Tahap pengembangan model klasifikasi terdiri dari preparasi, *training* & *validation* dan evaluasi. Preparasi dilakukan oleh membersihkan dan merapikan *dataset*. Preparasi dilakukan dengan teknik standardisasi yang diterapkan pada masing-masing variabel numerik pada *dataset*. Teknik standardisasi bertujuan untuk menyamakan skala numerik pada variabel numerik tersebut. Setelah preparasi, *dataset* kemudian dipecah dengan rasio 8:2 dimana rasio yang lebih besar akan digunakan sebagai data latih dan sisanya digunakan sebagai data uji.

Tahap *training* & *validation* digunakan untuk mengoptimasi performa model klasifikasi. Optimasi dilakukan dengan teknik *hyperparameter tuning* dengan implementasi metode *random search* menggunakan pembatasan parameter (*grid paramater*). Performa model klasifikasi diukur menggunakan skema *cross validation* dimana data validasi akan diambil sebanyak 20% dari data latih yang tersedia. Validasi dilakukan sebanyak 5 kali dengan menggunakan kurang lebih 100 kombinasi parameter model. Artinya, proses *hyperparameter tuning* akan melibatkan setidaknya 500 model klasifikasi berbeda untuk dicari tahu parameter seperti apa yang mampu memberikan performa terbaik. Setelah didapatkan parameter model terbaik, tahap selanjutnya adalah evaluasi model. Evaluasi dilakukan dengan mengukur performa model menggunakan data uji. Data uji adalah data yang sebelumnya sama sekali tidak digunakan pada *training* & *validation* model. Artinya data tersebut adalah data yang benar-benar baru dan belum dikenali oleh model klasifikasi.

2.6. Komparasi Model Klasifikasi

Komparasi antar model klasifikasi dilakukan berdasarkan beberapa metrik evaluator seperti *confusion matrix*, skor ROC-AUC, dan *log-loss*. Khusus pada *confusion matrix* akan lebih difokuskan menggunakan nilai akurasi dan skor F1-nya. Akurasi digunakan untuk mengukur kekuatan prediksi sedangkan skor F1 digunakan untuk mengukur keseimbangan aspek presisi dan sensitivitas model. Skor ROC-AUC menjelaskan bagaimana kekuatan model dalam membedakan antar program studi dan *log-loss* digunakan untuk mengukur kedekatan nilai probabilitas model jika dibandingkan data aslinya. Metrik evaluator tersebut biasanya digunakan pada kasus klasifikasi biner dengan jumlah



Gambar 2. Diagram Alir Model *Multistage*

DOI: https://doi.org/10.29207/resti.v5i5.3392

kelas sama dengan 2, namun penelitian yang dilakukan memiliki kasus dengan banyak kelas (*multi class*) sehingga untuk mengakomodasi kasus tersebut akan digunakan nilai rerata dari masing-masing metrik. Nilai rerata metrik didapatkan dari model klasifikasi yang menerapkan pendekatan *one-versus-one* (OVO) pada kasus *multi class* yang dimiliki.

3. Hasil dan Pembahasan

3.1. Hasil Preparasi *Dataset*

Terdapat dua skenario berbeda yang digunakan pada preparasi *dataset* DB1. Tabel 1 menunjukkan sebaran data setelah proses preparasi berdasakan program studinya. Sebaran pada program studi Informatika disajikan agar nantinya dapat dibandingkan dengan data mahasiswa Informatika yang tersedia pada *dataset* DB2.

Tabel 1. Hasil Preparasi *Dataset* DB1

Skenario	Total Mahasiswa	Mahasiswa Informatika	Mahasiswa Non-Informatika
A	1,980	76	1,904
B	2,908	116	2,792

Terlihat bahwa proporsi mahasiswa Informatika sangat kecil dibandingkan dengan ukuran *dataset* dengan proporsi kurang lebih sekitar 4%. Pada data yang menggunakan Skenario-B, didapatkan jumlah mahasiswa yang lebih banyak. Wajar mengingat skenario tersebut tidak menggunakan seleksi data seperti pada Skenario-A.

Selanjutnya akan dilihat bagaimana sebaran data pada *dataset* DB2 sebelum dan sesudah preparasi. Tidak seperti DB1 yang memiliki dua jenis skenario, pada DB2 hanya digunakan satu jenis skenario dengan menerapkan seleksi data berdasarkan jenis kurikulum, jenis semester dan kelengkapan data nilai per mata kuliah. Tabel 2 menunjukkan sebaran data pada DB2.

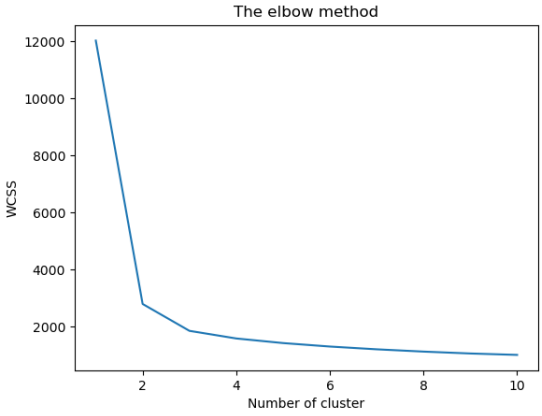
Tabel 2. Hasil Preparasi *Dataset* DB2

Sebelum Preparasi	Sesudah Preparasi
6.896 mahasiswa	1.007 mahasiswa
491 mata kuliah	6 mata kuliah

Setelah dilakukan preparasi terlihat bahwa jumlah mahasiswa Informatika yang berhasil dipertahankan berkurang cukup signifikan. Jika dibandingkan dengan data mahasiswa Informatika pada DB1, terlihat bahwa DB2 memiliki jumlah mahasiswa yang jauh lebih banyak. Namun tidak semua data tersebut dapat digunakan karena nantinya pada proses penggabungan *dataset* akan digunakan DB1 sebagai referensinya.

3.2. Hasil Pengelompokkan Mahasiswa

Tahap pengelompokkan mahasiswa Informatika dengan menggunakan model *clustering* diawali dengan proses penentuan jumlah kelompok atau klaster yang ideal yang dapat dibentuk. Penentuan jumlah kelompok ini dilakukan dengan menggunakan teknik siku (*elbow method*). Teknik tersebut memvisualisasika nilai *within cluster sum square* (WCSS) menggunakan grafik garis (*line chart*) seperti ditunjukkan pada Gambar 3. Jumlah kelompok ideal didapatkan pada saat garis WCSS pada grafik melandai.



Gambar 3. Grafik WCSS

Terlihat bahwa garis mulai melandai pada jumlah kelompok atau klaster sama dengan 2. Artinya, model *clustering* cukup mengelompokkan data ke dalam 2 jenis kelompok berbeda. Jumlah kelompok ideal tersebut kemudian digunakan pada model *clustering* sebagai salah satu parameter masukan. Komparasi beberapa model *clustering* menggunakan nilai koefisien Silhoutte dapat dilihat pada Tabel 3.

Tabel 3. Komparasi Model *Clustering*

K-means	Agglomerative	Birch	DBSCAN
0.729	0.727	0.621	0.013

Koefisien Silhoutte sendiri dapat dikatakan akan semakin bagus ketika nilainya mendekati 1. Berdasarkan tabel di atas terlihat bahwa model K-means dan Agglomerative memiliki performa yang lebih baik dibandingkan Birch maupun DBSCAN. Meskipun koefisien pada kedua model tersebut tidak berbeda signifikan, peneliti memutuskan untuk menggunakan K-means mengingat model tersebut sedikit lebih baik dibandingkan Agglomerative. Pada akhirnya, mahasiswa Informatika akan dipecah menjadi dua kelompok dengan menggunakan model K-means.

3.3. Hasil *Semi-supervised Learning*

Mengingat jumlah data mahasiswa Informatika yang jauh berbeda antara DB1 dan DB2, maka data latih yang tersedia untuk model klasifikasi *semi-supervised learning* pun tidak banyak tersedia. Setelah proses penggabungan terdapat total 34 mahasiswa pada data latih Skenario-A dan 50 mahasiswa pada data latih Skenario-B. Menggunakan data mahasiswa Informatika tersebut kemudian dikembangkan model klasifikasi *semi-supervised learning*. Tabel 4 menunjukkan

DOI: https://doi.org/10.29207/resti.v5i5.3392

performa dari masing-masing metode klasifikasi yang digunakan pada tahap ini.

Tabel 4. Performa Model Klasifikasi *Semi-supervised Learning*

	Skenario-A			Skenario-B		
	MLR	RF	SVM	MLR	RF	SVM
Avg. Accuracy	0.80	0.90	0.90	0.87	0.87	0.87
Avg. F1 Score	0.80	0.90	0.90	0.87	0.87	0.87
Avg.ROC AUC Score	0.84	0.96	1.00	0.85	0.87	0.87
Avg.Log loss	3.17	0.27	0.25	0.49	0.48	0.39

Pada model dengan data latih Skenario-A terlihat bahwa model RF dan SVM memiliki performa yang lebih baik dibandingkan model MLR, sedangkan pada Skenario-B terlihat ketiga model memiliki performa yang hampir sama. Meskipun ketiga model bekerja sama baiknya, jika dilihat berdasarkan nilai *log-loss* keunggulan terlihat dimiliki oleh model SVM. Model tersebut memiliki nilai rerata *log-loss* terkecil dibandingkan kedua model lainnya. Sehingga penelitian akan menggunakan model SVM untuk memprediksi label kelas pada program studi non-Informatika.

3.4. Hasil Penentuan Kelompok Mahasiswa

Tahap pengelompokkan mahasiswa diakhiri dengan pengambilan keputusan terkait kelompok mana yang seharusnya digunakan sebagai data latih model klasifikasi pada sistem rekomendasi pemilihan program studi. Tabel 5 menunjukkan karakteristik kelompok mahasiswa program studi Informatika.

Tabel 5. Sebaran Data Nilai Mata Kuliah Mahasiswa Informatika

	Skenario-A		Skenario-B	
	Kel. 0	Kel. 1	Kel. 0	Kel. 1
Total Mahasiswa	11	23	14	36
Rerata Jumlah Mata Kuliah Lulus	0.00	6.00	0.42	6.00
Rerata Nilai per Mata Kuliah	0.00	3.42	1.41	3.34

Dari sebaran data terlihat bahwa pada Skenario-A maupun Skenario-B, mahasiswa pada kelompok 1 memiliki hasil studi yang lebih baik. Mahasiswa pada kelompok 0 dapat dikakatan adalah kelompok mahasiswa yang tidak berhasil lulus pada 6 jenis mata kuliah. Berbeda dengan mahasiswa pada kelompok 1 yang menunjukkan rerata nilai per mata kuliahnya 3.00 atau setara dengan nilai B pada program studi Informatika.

Selanjutnya akan dilihat bagaimana sebaran data pada program studi non-Informatika. Pada program studi tersebut sebaran data tidak dapat dilihat berdasarkan nilai mata kuliah karena tidak tersedianya data. Oleh karenanya, akan dilihat sebaran data nilai mata pelajaran SMA berdasarkan kelas hasil prediksi model *semi-supervised learning*. Untuk mempermudah proses pengelompokkan dan melakukan interpretasi pada masing-masing kelas, sebaran data akan disajikan berdasarkan kelompok rumpun program studi yaitu sains, teknik, medis dan sosial. Kelompok rumpun program studi tersebut adalah kelompok yang juga digunakan pada *stage* pertama model klasifikasi *multistage*. Mengingat banyaknya jenis mata pelajaran yang tersedia maka hanya akan dilihat sebaran data pada beberapa mata pelajaran saja. Pemilihan jenis mata pelajaran disesuaikan oleh peneliti dengan rumpun program studinya. Tabel 6 menunjukkan sebaran data nilai mata pelajaran pada rumpun program studi sains.

Tabel 6. Sebaran Data Nilai Mata Pelajaran Rumpun Sains

	Skenario-A		Skenario-B	
	Kel. 0	Kel. 1	Kel. 0	Kel. 1
Matematika	55	65	55	71
Fisika	55	72	55	71
Kimia	56	71	55	72
Biologi	56	71	56	75

Terlihat mahasiswa kelompok 1 memiliki nilai rerata lebih baik dibandingkan kelompok 0. Jika perbandingkan dilakukan antara skenario, terlihat bahwa karakteristik mahasiswa pada masing-masing kelompok tidak jauh berbeda. Selanjutnya akan dilihat sebaran data nilai mata pelajaran pada rumpun program studi teknik yang disajikan pada Tabel 7.

Tabel 7. Sebaran Data Nilai Mata Pelajaran Rumpun Teknik

	Skenario-A		Skenario-B	
	Kel. 0	Kel. 1	Kel. 0	Kel. 1
Matematika	55	66	56	71
Fisika	56	64	56	70
Kimia	56	66	56	70
Biologi	56	65	56	66

Hasil serupa didapatkan pada rumpun program studi teknik. Terlihat bahwa mahasiswa kelompok 1 memiliki nilai rerata yang lebih baik dibandingkan kelompok lain. Interpretasi ketiga akan dilakukan pada rumpun program studi medis. Sebaran data nilai mata pelajaran rumpun tersebut disajikan pada Tabel 8.

Tabel 8. Sebaran Data Nilai Mata Pelajaran Rumpun Medis

	Skenario-A		Skenario-B	
	Kel. 0	Kel. 1	Kel. 0	Kel. 1
Matematika	57	76	56	76
Fisika	54	76	54	77
Kimia	57	66	57	75
Biologi	57	80	57	81

Tidak terdapat perbedaan pada rumpun program studi Medis. Mahasiswa kelompok 1 terlihat memiliki nilai rerata lebih tinggi. Peneliti menemukan sedikit perbedaan pada rumpun medis dibandingkan rumpun sebelumnya yaitu mahasiswa kelompok 1 memiliki nilai rerata mata pelajaran biologi yang lebih bagus yaitu di atas 80. Nilai rerata tersebut adalah yang tertinggi jika dibandingkan dengan mata pelajaran lain pada rumpun program studi manapun.

Interpretasi terakhir akan dilakukan pada rumpun program studi sosial. Mata pelajaran geografi, sejarah

DOI: https://doi.org/10.29207/resti.v5i5.3392

dan ekonomi akan menggantikan fisika, kimia dan biologi. Sebaran data pada rumpun sosial disajikan pada Tabel 9.

Tabel 9. Sebaran Data Nilai Mata Pelajaran Rumpun Sosial

	Skenario-A		Skenario-B	
	Kel. 0	Kel. 1	Kel. 0	Kel. 1
Matematika	51	58	50	61
Geografi	13	57	14	56
Sejarah	13	57	14	57
Ekonomi	13	58	14	58

Terlihat jelas bahwa nilai rerata mahasiswa pada rumpun sosial tidak terlalu tinggi. Akan tetapi perbandingan dilakukan antar kelompok, terlihat bahwa mahasiswa kelompok 1 memiliki nilai rerata jauh lebih tinggi. Hal yang wajar mengingat model *semi-supervised learning* dilatih menggunakan data mahasiswa Informatika dan dapat dikatakan tidak terlalu representatif untuk digunakan pada rumpun sosial. Sebagai tambahan informasi, pada tiga rumpun sebelumnya didapatkan bahwa nilai rerata mata pelajaran sosial pada masing-masing kelompok sangat rendah. Nilai rerata tertinggi hanya didapatkan pada angka sekitar 34 dan terlihat hanya pada mahasiswa kelompok 1. Artinya, meskipun tidak terlalu bekerja dengan baik pada rumpun sosial, hasil ini tetap masih dapat digunakan untuk proses penentuan kelompok mahasiswa.

Dari perbandingan menggunakan sebaran data nilai mata pelajaran pada masing-masing rumpun ilmu, dapat diambil kesimpulan bahwa mahasiswa pada kelompok 1 memiliki prestasi studi yang lebih baik dibandingkan kelompok 0. Peneliti memutuskan untuk menggunakan mahasiswa pada kelompok 1 sebagai kelompok data latih. Data mahasiswa pada kelompok tersebut akan digunakan sebagai data latih pada model klasifikasi sistem rekomendasi

3.5. Performa Model Klasifikasi *Single stage*

Pada model *single stage* dikomparasi tiga metode klasifikasi berbeda yaitu MLR, RF dan SVM. Performa dari masing-masing metode diukur menggunakan beberapa metrik evaluator. Tabel 10 menunjukkan performa dari model *single stage*.

Tabel 10. Performa Model Klasifikasi *Single stage*

	Skenario-A			Skenario-B		
	MLR	RF	SVM	MLR	RF	SVM
Avg. Accuracy	0.44	0.92	0.59	0.30	0.92	0.41
Avg. F1 Score	0.43	0.92	0.56	0.27	0.90	0.40
Avg.ROC AUC Score	0.90	0.99	0.95	0.83	0.99	0.89
Avg.Log loss	1.76	0.29	1.14	2.34	0.34	1.88

Model RF memiliki performa yang lebih baik mengungguli kedua model lainnya baik pada Skenario-A maupun Skenario-B. Model MLR dan SVM sendiri dapat dikatakan mempunyai performa yang kurang. Dari sisi data latih yang digunakan, terlihat bahwa model *single stage* menggunakan data latih Skenario-A memiliki performa yang lebih baik dibandingkan model dengan data latih Skenario-B.

3.5. Performa Model Klasifikasi *Multistage*

Model *multistage* memiliki beberapa *stage* di dalamnya, sehingga dimungkinkan memiliki model terbaik yang berbeda di setiap *stage*-nya. Sebelum melihat performa model *multistage* secara utuh akan dilihat terlebih dahulu komposisi metode klasifikasi terbaik masing-masing *stage*. Tabel 11 menunjukkan komposisi sususan metode klasifikasi pada model *multistage*.

Tabel 11. Metode Klasifikasi Terbaik tiap *Stage*

Stage	Skenario-A	Skenario-B
Stage I	RF	RF
Stage II Medic	MLR	MLR
Stage II Sains	MLR	SVM
Stage II Social	RF	RF
Stage II Teknik	SVM	RF
Stage III Social Ekonomi	RF	RF
Stage III Social IAI	MLR	RF
Stage III Social Psikologi	RF	RF
Stage III Teknik FTI	RF	RF
Stage III Teknik Sipil	SVM	RF

Metode klasifikasi RF terlihat mendominasi baik pada Skenario-A maupun Skenario-B. Pada Skenario-B, metode tersebut bahkan mendominasi dari awal. Sedangkan pada Skenario-A, kontribusi model MLR dan SVM masih banyak terlihat. Menggunakan komposisi tersebut selanjutnyan dilihat performa model *multistage*. Performa model *multistage* pada masing-masing skenario dapat dilihat pada Tabel 12.

Tabel 12. Performa Model *Multistage*

	Skenario-A	Skenario-B
Avg. Accuracy	0.82	0.85
Avg. F1 Score	0.79	0.81
Avg.ROC AUC Score	0.97	0.99
Avg.Log loss	0.26	0.17

Sedikit berbeda dengan performa model sebelumnya, pada model *multistage* performa terbaik didapatkan ketika menggunakan data latih Skenario-B. Namun dapat dikatakan perbedaan tersebut tidak terlalu signifikan. Model *multistage* pada Skenario-A dan Skenario-B memiliki performa yang sama baiknya.

3.5. Studi Komparasi

Studi komparasi akhir akan melibatkan model klasifikasi [22] yaitu model *single stage* dengan metode RF. Model tersebut akan dibandingkan dengan model *single stage* dan *multistage* terbaik yang didapatkan pada penelitian ini. Tabel 13 menunjukkan komparasi dari ketiga model.

DOI: https://doi.org/10.29207/resti.v5i5.3392

Tabel 13. Komparasi Model Klasifikasi Terbaik

	Single stage A	*Multistage B*	*Preliminary study*
Avg. Accuracy	0.92	0.85	0.86
Avg. F1 Score	0.92	0.81	0.84
Avg.ROC AUC Score	0.99	0.99	0.97
Avg.Log loss	0.29	0.17	0.66

Hasil komparasi menunjukkan bahwa model *single stage* dengan data latih Skenario-A pada penelitian memiliki performa yang lebih baik dibandingkan kedua model lainnya. Keunggulan terlihat hampir pada setiap metrik evaluator. Hanya pada nilai rerata *log-loss* model tersebut tidak lebih baik dibandingkan model *multistage*.

Model *multistage* sendiri dapat dikatakan memiliki performa yang tidak jauh berbeda dibandingkan model *preliminary study*. Model tersebut memiliki nilai akurasi yang tidak jauh berbeda, namun lebih unggul pada nilai rerata skor ROC-AUC dan nilai rerata *log-loss*.

4. Kesimpulan

Implementasi sains data khususnya model berbasis pembelajaran mesin pada pengembangan sistem rekomendasi pemilihan program studi diwujudkan dalam bentuk implementasi model klastering dan model klasifikasi. Model klastering digunakan pada tahap pengelompokkan mahasiswa untuk memastikan model klasifikasi pada sistem rekomendasi dapat dilatih menggunakan data mahasiswa yang representatif. Sedangkan model klasifikasi digunakan sebagai model final untuk memberikan hasil rekomendasi kepada pengguna sistem.

Hasil studi komparasi menunjukkan bahwa model klastering K-means mampu memberikan performa yang baik dalam tahap pengelompokkan mahasiswa. Meskipun merupakan model yang cukup sederhana, akan tetapi model tersebut menjadi alat pendukung yang baik sampai tahap akhir dalam pengembangan model klasifikasi sistem rekomendasi. Hasil komparasi juga menunjukkan bahwa dua model klasifikasi yang diinisiasi pada penelitian memiliki performa yang bagus dibandingkan dengan model sebelumnya.

Kesimpulan yang dapat diambil dari studi komparasi penelitian ini secara garis besar dapat dikategorikan menjadi dua. Pertama, penggunakan data nilai mata kuliah sebagai data poin pada proses pengelompokkan mahasiswa terbukti bekerja cukup baik. Hal yang tidak dilakukan pada penelitian sebelumnya tersebut terbukti membantu dan memberikan hasil pengelompokkan yang dapat digunakan secara utuh pada penelitia. Kedua, model *multistage* yang merupakan skema dengan pendekatan baru menunjukan performa yang cukup baik. Temuan ini dapat menjadi alternatif pada implementasi sistem rekomendasi pemilihan program studi nantinya.

Terakhir, peneliti berharap bahwa penelitian ini dapat menjadi inisiasi penelitian lain di masa depan. Terkait sistem rekomendasi pemilihan studi, improvisasi dalam pengembangan model dapat dilakukan pada tahap pengelompokkan mahasiswa. Alih-alih menggunakan teknik *semi-supervised learning*, proses pengelompokkan dapat dilakukan secara spesifik pada masing-masing program studi dengan menggunakan model klastering. Hal ini akan mengurangi bias dan harapannya mampu meningkatkan performa model klasifikasi sistem rekomendasi.

Ucapan Terimakasih

Terima kasih ditujukan kepada semua pihak yang terlibat pada penelitian ini mulai dari dosen pembimbing dan beberapa instansi internal di Universitas Islam Indonesia seperti tim Badan Sistem Informasi (BSI) dan tim Penerimaan Mahasiswa Baru (PMB) yang telah membantu menyediakan data penelitian.

Daftar Rujukan

[1] A. Andriani, "Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik ' BSI Yogyakarta ,'" *Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013)*, vol. 2013, no. SENTIKA, 2013.

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, 2013, doi: 10.1016/j.knosys.2013.03.012.

[3] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, 1995, doi: 10.1023/A:1022627411411.

[4] M. Ezz and A. Elshenawy, "Adaptive recommendation system using machine learning algorithms for predicting student's best academic program," *Education and Information Technologies*, vol. 25, no. 4, 2020, doi: 10.1007/s10639-019-10049-7.

[5] G. DS and K. K, "Developing an Intelligent Recommendation System for Course Selection by Students for Graduate Courses," *Business and Economics Journal*, vol. 7, no. 2, 2015, doi: 10.4172/2151-6219.1000209.

[6] Y. Isler, A. Narin, M. Ozer, and M. Perc, "Multi-stage classification of congestive heart failure based on short-term heart rate variability," *Chaos, Solitons and Fractals*, vol. 118, 2019, doi: 10.1016/j.chaos.2018.11.020.

[7] D. Kancherla, J. D. Bodapati, and N. Veeranjaneyulu, "Effect of different kernels on the performance of an SVM based classification," *International Journal of Recent Technology and Engineering*, vol. 7, no. 5, 2019.

[8] J. A. Konstan and J. Riedl, "Recommender systems: From algorithms to user experience," *User Modeling and User-Adapted Interaction*, vol. 22, no. 1–2. 2012. doi: 10.1007/s11257-011-9112-x.

[9] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica (Ljubljana)*, vol. 31, no. 3. 2007. doi: 10.31449/inf.v31i3.148.

[10] L. D. Kumalasari and A. Susanto, "Recommendation System of Information Technology Jobs using Collaborative Filtering Method Based on LinkedIn Skills Endorsement," *SISFORMA*, vol. 6, no. 2, 2020, doi: 10.24167/sisforma.v6i2.2240.

[11] C. Kwak and A. Clayton-Matthews, "Multinomial logistic regression," *Nursing Research*, vol. 51, no. 6, 2002, doi: 10.1097/00006199-200211000-00009.

[12] E. Marbun and S. Hansun, "Sistem Pendukung Keputusan Pemilihan Program Studi Dengan Metode Saw Dan Ahp," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, 2019, doi: 10.33096/ilkom.v11i3.432.175-183.

DOI: https://doi.org/10.29207/resti.v5i5.3392

[13] A. Mendes, J. Togelius, and L. dos Santos Coelho, "Multi-stage transfer learning with an application to selection process," in *Frontiers in Artificial Intelligence and Applications*, 2020, vol. 325. doi: 10.3233/FAIA200291.

[14] Jpnn.com, "Hasil Survei: 87 Persen Mahasiswa Pilih Jurusan Tidak Sesuai Minat," *jpnn.com*, 2019.

[15] M. R. Okaviana and R. Susanto, "Sistem Pendukung Keputusan Rekomendasi Pemilihan Program Studi Menggunakan Metode Multifactor Evaluation Process Di Sma Negeri 1 Bandung," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 3, no. 2, 2014, doi: 10.34010/komputa.v3i2.2389.

[16] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, 2005, doi: 10.1080/01431160412331269698.

[17] A. Parameswaran, P. Venetis, and H. Garcia-Molina, "Recommendation systems with complex constraints: A course recommendation perspective," *ACM Transactions on Information Systems*, vol. 29, no. 4. 2011. doi: 10.1145/2037661.2037665.

[18] S. Pare, "Sistem Pendukung Keputusan Pemilihan Program Studi Pada Perguruan Tinggi," *jurnal iImiah Mustek Anim Ha*, vol. 2, no. 9, 2013.

[19] Mrs. Leena. and Dr. Mohammed, "A Multistage Feature Selection Model for Document Classification Using Information Gain and Rough Set," *International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 11, 2014, doi: 10.14569/ijarai.2014.031103.

[20] H. S. Permatasari, A. Suyatno, and A. H. Kridalaksana, "Sistem Pendukung Keputusan Pemilihan Program Studi Di Universitas Mulawarman Menggunakan Metode Tsukamoto (Studi Kasus : Fakultas MIPA)," *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 10, no. 1, 2016, doi: 10.30872/jim.v10i1.19.

[21] S. S. Poorna and G. J. Nair, "Multistage classification scheme to enhance speech emotion recognition," *International Journal of Speech Technology*, 2019, doi: 10.1007/s10772-019-09605-w.

[22] A. M. R. Pratama, R. R. Aryanto, and A. T. Pratama, "Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. 2021.

[23] I. Safutra, "Mahasiswa Mengaku Salah Jurusan, Banyak Sarjana yang Penting Lulus," Dec. 15, 2019.

[24] C. M. Salgado, M. P. Fernandes, A. Horta, M. Xavier, J. M. C. Sousa, and S. M. Vieira, "Multistage modeling for the classification of numerical and categorical datasets," 2017. doi: 10.1109/FUZZ-IEEE.2017.8015665.

DOI: https://doi.org/10.29207/resti.v5i5.3392