



Wiyli Yustanti <wilyliyustanti@unesa.ac.id>

[JP] Submission Acknowledgement

1 message

Dr. Herry Sujaini <jurnal@untan.ac.id>
To: Wiyli Yustanti <wilyliyustanti@unesa.ac.id>

Mon, Nov 15, 2021 at 11:46 PM

Wiyli Yustanti:

Thank you for submitting the manuscript, "Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring" to JEPIN (Jurnal Edukasi dan Penelitian Informatika). With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Manuscript URL:

<https://jurnal.untan.ac.id/index.php/jepin/author/submission/50452>

Username: wiyli

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Dr. Herry Sujaini
JEPIN (Jurnal Edukasi dan Penelitian Informatika)

JEPIN<http://jurnal.untan.ac.id/index.php/jepin>



Wiyli Yustanti <wilyliyustanti@unesa.ac.id>

[JP] Editor Decision

1 message

Dr. Herry Sujaini <jurnal@untan.ac.id>
Reply-To: "Dr. Levy Olivia" <levoliv@yahoo.com>
To: Wiyli Yustanti <wilyliyustanti@unesa.ac.id>

Mon, Dec 27, 2021 at 8:30 PM

Wiyli Yustanti:

Kami telah memutuskan hasil dari manuskrip yang dikirimkan ke JEPIN (Jurnal Edukasi dan Penelitian Informatika), dengan judul "Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring".

Keputusan tim editor adalah : Diterima dengan revisi

Mohon manuskripnya diperbaiki berdasarkan komentar dan saran dari reviewer, setelah itu silahkan dikirimkan kembali lewat OJS Jepin. Manuskrip yang telah dikirimkan dikirimkan kembali selambat-lambatnya satu minggu dan manuskrip yang telah diperbaiki telah sesuai template Jepin.

Dr. Levy Olivia
Universitas Telkom
levoliv@yahoo.com

Reviewer A:

Saran perbaikan untuk Penulis :

:

2. Penjelasan abstrak kurang sistematis, sehingga tidak dapat ditarik runut hubungan antara melakukan analisa algoritma klasifikasi dengan analisa prediktif karakteristik mahasiswa pada pembelajaran daring.

2. Masih banyak kesalahan penulisan

Reviewer B:

Saran perbaikan untuk Penulis :

:

JEPIN

<http://jurnal.untan.ac.id/index.php/jepin>



Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring

Abstrak — Pelaksanaan pembelajaran secara daring di masa pandemi, khususnya proses belajar mengajar di perguruan tinggi memerlukan dukungan fasilitas seperti ketersediaan sinyal atau jaringan yang baik, ketersediaan kuota internet, serta mahasiswa juga sebaiknya memiliki perangkat penunjang seperti smartphone dan laptop. Selain faktor sarana prasarana, keberhasilan belajar akan tercapai apabila seorang mahasiswa memiliki motivasi belajar, budaya belajar, dan kondisi fisik serta psikologi yang baik. Tahap *pre-processing* menggunakan proses seleksi fitur dengan uji independen Chi-Square untuk menentukan variabel yang berpengaruh pada proses prediksi variabel respon. Hasil seleksi variabel independen menghasilkan 16 variabel yang berpengaruh dari total 22 variabel awal. Adapun jenis label pada variabel respon terdiri dari 4 kelompok yaitu mahasiswa aktif dengan fasilitas terbatas, aktif dengan fasilitas baik, pasif dengan fasilitas baik dan pasif dengan fasilitas terbatas. Pada data penelitian ini, distribusi variabel respon termasuk dalam kategori tidak seimbang (*imbalanced class*) dengan proporsi kelas terkecil adalah 2,20%. Untuk melakukan *balancing* data digunakan teknik *oversampling* sebelum tahap pemodelan. Algoritma klasifikasi yang diujicobakan terdiri dari 6 algoritma yaitu Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Regresi Logistik (LR), K-Nearest Neighbor (KNN), dan Decision Tree (DT). Hasil komparasi kinerja menunjukkan bahwa algoritma Support Vector Machine (SVM) lebih unggul dengan nilai F-1 Score 92,8% dan AUC sebesar 99,01%.

Kata kunci — Klasifikasi, F1-Score, AUC, Imbalanced, Oversampling

I. PENDAHULUAN

Pada masa pandemi, proses pembelajar di berbagai institusi pendidikan dilakukan secara daring yang menggantikan proses pembelajaran tatap muka. Pembelajaran dikelas dapat dilakukan secara virtual melalui media komunikasi yang mendukung tatap muka secara daring, seperti aplikasi media sosial atau platform e-learning. Dengan demikian, siswa tetap dapat mengikuti proses pembelajaran tanpa terhalang lokasi dan waktu. Pembelajaran secara daring sudah menjadi solusi untuk menghentikan laju persebaran virus. Hal ini dilakukan sebagai upaya pemerintah melalui kebijakannya secara bertahap untuk mewajibkan semua proses pembelajaran dilakukan secara daring sampai waktu kondusif untuk

memulai pembelajaran tatap muka. Upaya ini diharapkan bahwa melalui metode pembelajaran daring, dapat memudahkan siswa dalam mengakses setiap materi pembelajaran, berinteraksi atau melakukan diskusi antara satu dengan yang lain sehingga tercapainya hasil belajar yang optimal.

Pelaksanaan pembelajaran secara daring tentunya memerlukan dukungan ketersediaan fasilitas seperti kualitas sinyal atau jaringan yang baik, kuota internet, smartphone, dan laptop. Dukungan sarana prasarana ini, berpengaruh terhadap kualitas proses pembelajaran daring yang dapat dilihat dari hasil belajar melalui nilai yang diperoleh. Selain aspek fasilitas, faktor internal siswa juga memiliki peran dalam keberhasilan belajar. Fadila, dkk [1] menjelaskan bahwa kemandirian belajar adalah bahwa siswa dapat belajar tanpa tergantung orang lain dalam proses pemahaman materi, penerapan pengetahuan, serta penyelesaian masalah. Pada penelitian lain, Keberhasilan belajar akan tercapai bila seorang pelajar memiliki dukungan fasilitas yang baik, motivasi belajar, budaya belajar serta memiliki kondisi fisik dan psikologi yang baik. Penelitian terkait dengan faktor yang berpengaruh terhadap hasil pembelajaran daring banyak ditemukan antara lain, Anggrawan, dkk [2] dalam hasil penelitiannya menyimpulkan bahwa ada kaitan antara gaya belajar dengan performansi belajar mahasiswa pada studi Bahasa Inggris. Dimana pada pembelajaran daring, mahasiswa tipe auditori dan visual rata-rata memiliki nilai lebih unggul dibanding dengan mahasiswa tipe kinestetik. Sedangkan hasil penelitian Andrianto, dkk [3] menjelaskan secara umum, faktor yang mempengaruhi kesuksesan penerapan e-learning di Indonesia terdiri dari tiga hal yang utama yaitu infrastruktur, kualitas sumber daya manusia dan institusi pendidikan itu sendiri. Ketiga hal ini membutuhkan pemerintah untuk mempersiapkan kegiatan pembelajaran secara dari dapat berjalan lancar. Penjelasan yang senada juga ditemukan pada penelitian Bambang [4].

Aspek internal siswa juga berpengaruh terhadap kualitas proses belajar secara daring, khususnya dalam membangun motivasi dan minat belajar. Syarifah, dkk [5] menemukan bahwa ada pengaruh signifikan antara aspek penggunaan kuota internet, cara pandang, budaya belajar, kondisi ekonomi terhadap minat belajar mahasiswa. Minat belajar ini, juga dipengaruhi oleh media pembelajaran

Commented [A3]: Terlalu umum

Commented [A1]: Antar kalimat tidak nyambung, diperbaiki dan dibaca kembali

Commented [A2]: Sort by alphabeth

daring yang digunakan, hasil penelitian Nabila, dkk [6] menunjukkan bahwa lebih dari 50% mahasiswa lebih menyukai pembelajaran melalui Google Classroom (GCR) dan Whatsapp Group (WAG). Hal ini juga dibuktikan bahwa, penggunaan Google Meet pada saat pembelajaran daring memberikan nilai yang baik dan sangat baik berdasarkan hasil penelitian Sepita [7]. Penelitian Mumuh [8] dan Turmuzi [9] juga memberikan kesimpulan bahwa penggunaan GCR dan WAG menjadi pilihan terbaik untuk proses pembelajaran daring karena ramah kuota internet. Windi, dkk [10] menemukan dalam penelitiannya bahwa kemandirian belajar dipengaruhi oleh motivasi berprestasi, dimana semakin tinggi motivasi seseorang untuk belajar maka tingkat kemandiriannya juga semakin tinggi dan sebaliknya. Berbeda halnya dengan siswa berkebutuhan khusus, dimana umur, gender dan pola asuh memberikan pengaruh yang cukup signifikan [11].

Implementasi pembelajaran secara daring sudah dilaksanakan hampir dua tahun di seluruh intitusi pendidikan di Indonesia. Salah satu hasil evaluasinya dapat dilihat pada Riyanda, dkk [12] yang menyimpulkan bahwa pembelajaran secara daring baik dari sisi aspek konteks, input, proses, dan output, berhasil dengan baik dan dapat dilanjutkan. Sedangkan pada hasil evaluasi Nurlina [13], menyimpulkan bahwa dengan pembelajaran daring guru lebih aktif dalam menyiapkan bahan ajar serta kemampuan menggunakan media komunikasi digital untuk pembelajaran lebih baik. Dalam hasil penelitian lain ditemukan beberapa dampak akibat pembelajaran daring antara lain mahasiswa masih sering mengalami kebingungan, pasif, kurang produktif dan tidak kreatif, stress, penumpukan konsep/informasi yang kurang bermanfaat serta literasi bahasa yang meningkat [14]. Purniawan [15] juga memberikan informasi terkait analisis respon siswa terhadap pembelajaran daring yang menunjukkan hasil cukup baik meskipun ketersediaan kuota internet sering menjadi kendala.

Aspek faktor-faktor yang ditemukan berpengaruh dan hasil evaluasi proses pelaksanaan pembelajaran daring selama ini memberikan informasi secara kuantitatif bahwa data-data tersebut dapat dieksplorasi lebih dalam untuk membangun model prediksi karakteristik mahasiswa berdasarkan variabel-variabel yang diduga berpengaruh signifikan. Penggunaan algoritma Naïve Bayes [16] diterapkan untuk memprediksi kepuasan mahasiswa pada proses pembelajaran daring dengan menggunakan variabel yang mengukur komunikasi, suasana pembelajaran, penilaian dan penyampaian materi. Penelitian serupa juga dilakukan Natuzzuhriyyah, dkk [17] yang menggunakan tujuh variabel independen sebagai prediktor tingkat kepuasan. Algoritma klasifikasi juga digunakan pada evaluasi hasil pembelajaran daring berbasis pada analisis sentiment di media sosial Twitter [18],[19] dan [20].

Berdasarkan dengan perkembangan penelitian pembelajaran daring di masa pandemi yang cukup beragam, maka pada penelitian ini ingin dilakukan analisis faktor yang berpengaruh terhadap karakteristik mahasiswa

pada pembelajaran daring berdasarkan berbagai aspek yang ditemukan berpengaruh signifikan pada penelitian terkait dengan menggunakan pendekatan algoritma klasifikasi. Terdapat enam algoritma klasifikasi yang akan dibandingkan yaitu Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Regresi Logistik (LR), K-Nearest Neighbor (KNN), dan Decision Tree (DT). Tujuannya adalah agar proses pembelajaran dapat dimanajemen dengan baik sesuai dengan karakteristik siswa yang ada.

II. METODE

Metodologi dalam penelitian ini mengadopsi proses yang ada pada kerangka kerja *Knowledge Data Discovery* (KDD) yang diusulkan oleh Fayyad [21]. Inti utama dari kerangka KDD terdiri dari lima tahap yaitu tahap pemilihan data, pre-processing, tranformasi, pemodelan, dan evaluasi.

A. Data

Sumber data dalam penelitian iini merupakan data primer yang diperoleh dari hasil isian formulir mahasiswa peserta mata kuliah Basis Data yang diambil secara random. Didapatkan 91 responden dengan instrument isian terdiri dari 22 variabel yang meliputi aspek internal dan eksternal mahasiswa saat mengikuti perkuliahan daring. Secara rinci, variabel dependen atau data kelas/target untuk pemodelan klasifikan menggunakan nilai pada Tabel I.

TABEL I
VARIABEL DEPENDEN DATASET MAHASISWA

Simbol	Variabel	Tipe Data	Nilai Data	Keterangan
y	Kelompok Siswa	Kategorik	0	Pembelajar aktif dengan dukungan fasilitas terbatas
			1	Pembelajar aktif dengan dukungan fasilitas baik
			2	Pembelajar pasif dengan dukungan fasilitas baik
			3	Pembelajar pasif dengan dukungan fasilitas terbatas

Sedangkan untuk variabel independen secara lengkap dapat dilihat pada Tabel II. Semua data yang digunakan dalam penelitian ini bertipe kategorik dengan skala Nominal.

TABEL II
VARIABEL INDEPENDEN DATASET MAHSISWA

Simbol	Fitur	Tipe Data	Nilai Data
x_1	Jenis kelamin	Kategorik	0 = Pria, 1 = Wanita
x_2	Berat Badan	Kategorik	1 = Kurus, 2 = Normal, 3 = Gemuk, 4 = Obesitas
x_3	Kacamata	Kategorik	1 = Ya, 0 = Tidak
x_4	Pernah Sakit	Kategorik	1 = Ya, 0 = Tidak
x_5	Pernah Gangguan	Kategorik	1 = Ya, 0 = Tidak

Commented [A4]: Analisis karakteristik mahasiswa kenapa harus membandingkan algoritma klasifikasi? Tidak ada sumber relevan atau referensi terkait untuk memperkuat tujuan penelitian

Simbol	Fitur	Tipe Data	Nilai Data
	Mental		
x_6	Aktif Bertanya	Kategorik	1 = Ya, 0 = Tidak
x_7	Aktif Menjawab	Kategorik	1 = Ya, 0 = Tidak
x_8	Aktif mengerjakan Tugas	Kategorik	0 = Sebagian, 1 = Semua
x_9	Minat	Kategorik	0 = Tidak, 1 = Mungkin, 2 = Ya
x_{10}	Waktu Belajar	Kategorik	1 = < 5 jam, 2 = 5 < t < 10 jam, 3 = > 10 jam
x_{11}	Ketersediaan Sumber Belajar Lain	Kategorik	1 = Ya, 0 = Tidak
x_{12}	Aktif mencari tutorial	Kategorik	1 = Ya, 0 = Tidak
x_{13}	Mengulang Materi	Kategorik	0 = Tidak, 1 = Terkadang 2 = Selalu
x_{14}	Melakukan Latihan	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Selalu
x_{15}	Kegiatan Diskusi	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Selalu
x_{16}	Smartphone	Kategorik	1 = Ya, 0 = Tidak
x_{17}	Laptop	Kategorik	1 = Ya, 0 = Tidak
x_{18}	Internet	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Ya
x_{19}	Dukungan Lingkungan Rumah	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Ya
x_{20}	Listrik	Kategorik	1 = Ya, 0 = Tidak
x_{21}	Daerah	Kategorik	1 = Pedesaan, 2 = Perkotaan, 3 = Pesisir
x_{22}	Kemudahan akses telekomunikasi	Kategorik	1 = Buruk, 2 = Sedang, 3 = Baik

B. Pre-Processing

Pada tahap ini variabel independen (x) pada dataset mahasiswa di seleksi kembali dengan cara memilih fitur yang berpengaruh dan membuang fitur yang tidak berpengaruh sehingga diharapkan dapat meningkatkan nilai akurasi saat dilakukannya proses klasifikasi, untuk menguji variabel independen digunakan uji *Chi-Square* dengan langkah-langkah dalam pengujian sebagai berikut :

- Menentukan Hipotesis.
- Membuat Tabel Kontingensi.
- Menentukan Nilai Harapan
- Menghitung Statistik Chi-Square.
- Memutuskan Terima atau Tolak Hipotesis Null.

C. Transformasi

Bentuk transformasi dapat berupa perubahan tipe data dari kategorik menjadi numerik maupun perubahan skala data menjadi range yang standard. Pada penelitian ini dilakukan transformasi data kategorik menggunakan *Label Encoding* yaitu mengkonversi secara langsung data text menjadi nilai integer yang bermakna nominal tanpa mempertimbangkan urutan. Setelah dilakukan konversi ke numerik, selanjutnya data distandarisasi dengan rumus :

$$x_{\text{baru}} = \frac{x_{\text{lama}} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

D. Pemodelan

Proses pemodelan merupakan tahapan membangun model prediksi klasifikasi berdasarkan nilai-nilai pada variabel input terhadap variabel output (kelas). Untuk mendapatkan model prediksi yang terbaik, dilakukan eksperimen dengan beberapa algoritma klasifikasi yaitu Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Regresi Logistik (RL), Random Forest (RF) dan Support Vector Machine (SVM). Proses optimasi parameter menggunakan metode Grid Search.

E. Validasi dan Evaluasi

Metode validasi yang digunakan untuk dataset yang memiliki distribusi kelas yang tidak seimbang adalah menggunakan Teknik oversampling sehingga distribusi menjadi seimbang. Untuk memastikan bahwa semua kelas akan terwakili pada proses training dan testing, digunakan Stratified-K Cross Validation dengan nilai fold adalah 2 sampai 10. Sedangkan metrik evaluasi kinerja prediksi kelas menggunakan F1-Score. Untuk kasus distribusi kelas yang seimbang, nilai Akurasi dan F1-Score akan menghasilkan angka yang sama. Ukuran lain yang dapat digunakan adalah nilai AUC. Adapun interpretasi nilai AUC dapat mengacu pada Tabel III.

TABEL III
KATEGORI NILAI AUC

AUC Score	Interpretasi Hasil Klasifikasi
0.90 – 1.00	Sangat Baik
0.80 – 0.90	Baik
0.70 – 0.80	Sedang
0.60 – 0.70	Buruk
0.50 – 0.60	Gagal

III. HASIL DAN PEMBAHASAN

Bagian ini membahas tentang hasil tahapan pre-processing, komparasi kinerja algoritma pada tahapan pemodelan, yang menggunakan teknik validasi dan evaluasi yang sudah dijelaskan pada bagian sebelumnya.

Commented [A5]: Penjelasan tahapan di bab metodologi

A. Seleksi Variabel

Berdasarkan hasil dari seleksi fitur dengan uji independen *Chi-Square* menggunakan nilai $\alpha=10\%$ didapatkan hasil seperti pada Tabel IV.

TABEL IV
SELEKSI FITUR DATASET DENGAN UJI CHI-SQUARE

X_i	χ^2_{hitung}	Likelihood	df	χ^2_{tabel}	H_0
x_1	6,398	7,219	3	6,251	Ditolak
x_2	13,389	15,874	9	14,684	Diterima
x_3	8,146	8,737	3	6,251	Ditolak

X_i	χ^2_{hitung}	Likelihood	df	χ^2_{tabel}	H_0
x_4	17,766	18,846	3	6,251	Ditolak
x_5	19,505	19,212	3	6,251	Ditolak
x_6	11,358	12,905	3	6,251	Ditolak
x_7	5,886	5,978	3	6,251	Diterima
x_8	60,753	17,745	3	6,251	Ditolak
x_9	70,471	3,725	6	10,645	Ditolak
x_{10}	3,189	3,786	3	6,251	Diterima
x_{11}	14,402	15,091	3	6,251	Ditolak
x_{12}	3,556	5,111	3	6,251	Diterima
x_{13}	24,708	26,110	3	6,251	Ditolak
x_{14}	23,035	27,027	3	6,251	Ditolak
x_{15}	16,058	15,546	6	10,645	Ditolak
x_{16}	2,162	2,311	3	6,251	Diterima
x_{17}	27,053	15,733	3	6,251	Ditolak
x_{18}	29,044	33,973	6	10,645	Ditolak
x_{19}	32,126	34,698	6	10,645	Ditolak
x_{20}	14,783	6,237	3	6,251	Ditolak
x_{21}	7,874	9,463	6	10,645	Diterima
x_{22}	32,385	37,358	6	10,645	Ditolak

Dari hasil seleksi fitur pada Tabel III diketahui variabel independen yang tidak digunakan dalam proses selanjutnya adalah x_2 , x_7 , x_{10} , x_{12} , x_{16} , dan x_{21} , sedangkan sisanya akan digunakan sebagai variabel input pada proses pemodelan klasifikasi. Dengan demikian, pada penelitian ini, menggunakan 17 variabel yang terpilih dari 23 variabel awal.

B. Kinerja Algoritma Klasifikasi

Sebagaimana yang dijelaskan dalam bagian metode, bahwa percobaan dilakukan dengan menggunakan enam algoritma klasifikasi yang berbeda, yaitu DT, KNN, NB, LR, RF dan SVM. Untuk mengetahui apakah proses seleksi fitur mempengaruhi hasil kinerja klasifikasi, maka dilakukan pengukuran kinerja klasifikasi menggunakan nilai F1-Score sebelum dan sesudah seleksi fitur. Tabel V dan VI menunjukkan hasil dari percobaan dengan mengambil nilai K-Fold muladi dari 2 sampai 10.

TABEL V
NILAI F1-SCORE SEBELUM SELEKSI FITUR

K	DT	KNN	LR	NB	RF	SVM
2	0.681	0.772	0.840	0.780	0.836	0.889
3	0.717	0.762	0.930	0.788	0.826	0.905
4	0.691	0.800	0.879	0.779	0.879	0.851
5	0.727	0.823	0.919	0.795	0.930	0.903
6	0.721	0.815	0.909	0.785	0.912	0.879
7	0.665	0.782	0.927	0.799	0.904	0.900
8	0.699	0.798	0.920	0.778	0.886	0.881
9	0.706	0.848	0.908	0.784	0.904	0.869
10	0.667	0.838	0.908	0.764	0.925*	0.889

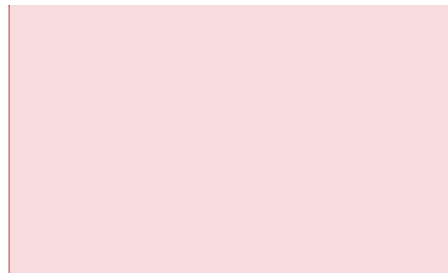
Didapatkan bahwa baik kinerja sebelum maupun sesudah seleksi fitur menunjukkan bahwa K-Fold sebanyak 10 memberikan hasil tertinggi. Untuk dataset sebelum diseleksi fitur, algoritma RF memberikan hasil kinerja prediksi tertinggi yaitu sebesar 92,5%. Sedangkan untuk dataset setelah seleksi fitur, algoritma SVM memiliki performansi yang paling tinggi yaitu sebesar 92,8% atau

sekitar 0,03% lebih unggul daripada algoritma RF pada dataset tanpa seleksi fitur.

TABEL VI
NILAI F1-SCORE SETELAH SELEKSI FITUR

	DT	KNN	LR	NB	RF	SVM
2	0.654	0.882	0.881	0.856	0.714	0.865
3	0.717	0.832	0.879	0.853	0.815	0.855
4	0.745	0.847	0.853	0.864	0.842	0.895
5	0.779	0.845	0.912	0.874	0.872	0.893
6	0.717	0.858	0.862	0.868	0.857	0.912
7	0.698	0.873	0.911	0.904	0.904	0.904
8	0.724	0.887	0.879	0.883	0.865	0.914
9	0.727	0.869	0.896	0.880	0.905	0.922
10	0.674	0.888	0.878	0.886	0.881	0.928*

Selanjutnya, analisis dilakukan dengan melihat nilai rata-rata F1-Score untuk semua algoritma pada kondisi sebelum seleksi fitur yaitu terdiri dari 22 variabel independen, dan setelah proses seleksi fitur dengan menggunakan 17 variabel. Secara umum terlihat bahwa untuk algoritma DT, KNN, NB dan SVM memiliki kinerja yang lebih tinggi dibandingkan dengan kinerja hasil model prediksi klasifikasi sebelum adanya proses seleksi fitur. Sedangkan untuk algoritma LR dan RF memiliki kinerja lebih rendah yaitu penurunan masing-masing sebesar 2,1% dan 3,9%. Secara visual, histogram nilai rata-rata F1 Score dengan jumlah fitur berbeda terlihat pada Gambar 1.



Commented [A6]: Gambar tidak muncul??

Gambar. 1 Perbandingan Nilai Rata-Rata F1-Score Algoritma Klasifikasi Sebelum dan Sesudah Seleksi Fitur

Dari keseluruhan proses eksperimen yang sudah dilakukan, kinerja tertinggi ditunjukkan oleh algoritma SVM. Selanjutnya sebagai konfirmasi dari hasil ini, dilakukan perhitungan nilai AUC. Hasil perhitungan AUC ditampilkan pada Tabel VII.

TABEL VII
NILAI AUC PADA K-FOLD 10

Jumlah Fitur	Nilai AUC untuk KFold = 10					
	DT	KNN	LR	NB	RF	SVM
22 Fitur	0.846	0.961	0.979	0.962	0.982	0.995
17 Fitur	0.715	0.865	0.884	0.874	0.851	0.899

Hasil AUC menunjukkan bahwa algoritma SVM tetap memberikan hasil prediksi yang tertinggi dibandingkan dengan algoritma yang lain. Nilai AUC dan F1-Score merupakan metrik evaluasi yang sangat kuat dan bekerja

sangat baik untuk banyak masalah klasifikasi. Salah satu perbedaan besar antara F1-Score dan AUC adalah bahwa F1-Score menekankan pada ketepatan kelas yang diprediksi dan AUC menggunakan skor yang diprediksi sebagai input. Jika dataset yang dianalisis memiliki distribusi yang sangat tidak seimbang dan hasil prediksi sangat mempertimbangkan ketepatan prediksi untuk kelas minoritas, maka yang paling tepat adalah menggunakan F1-Score. Sebaliknya, jika dataset yang dianalisis memiliki distribusi kelas yang hampir seimbang, maka dapat digunakan metrik evaluasi seperti Akurasi dan AUC.

IV. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan bahwa karakteristik mahasiswa yang mengikuti perkuliahan daring dapat diprediksi menggunakan algoritma SVM dengan sangat baik yakni memiliki nilai ketepatan prediksi kelas sebesar 92,8% dan nilai AUC 89,9% dengan menggunakan variabel input sebanyak 17 fitur, yaitu dengan mengeluarkan variabel $x_2, x_7, x_{10}, x_{12}, x_{16}$, dan x_{21} .

UCAPAN TERIMA KASIH / ACKNOWLEDGMENT

Kami mengucapkan terimakasih atas dukungan Lembaga Penelitian Universitas Negeri Surabaya atas hibah penelitian kebijakan Fakultas.

DAFTAR PUSTAKA

- [1] R. N. Fadila, T. A. Nadiroh, R. Juliana, P. Z. H. Zulfa, and I. Ibrahim, "Kemampuan Belajar Secara Daring Sebagai Prediktor Hasil Belajar Mahasiswa Pendidikan Matematika UIN Sunan Kalijaga," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 2, pp. 880–891, 2021, doi: 10.31004/cendekia.v5i2.457.
- [2] A. Anggrawan, "Analisis Deskriptif Hasil Belajar Pembelajaran Tatap Muka dan Pembelajaran Online Menurut Gaya Belajar Mahasiswa," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 18, no. 2, pp. 339–346, 2019, doi: 10.30812/matrik.v18i2.411.
- [3] R. Andrianto Pangondian, P. Insap Santosa, and E. Nugroho, "Faktor - Faktor Yang Mempengaruhi Kesuksesan Pembelajaran Daring Dalam Revolusi Industri 4.0," *Sainteks 2019*, pp. 56–60, 2019, [Online]. Available: <https://seminar-id.com/seminas-sainteks2019.html>.
- [4] B. Budhianto, "Analisis perkembangan dan faktor yang mempengaruhi keberhasilan pembelajaran daring (e-learning)," *J. AgriWidya*, vol. 1, no. 1, pp. 11–29, 2020.
- [5] S. H. Jamil and I. D. Aprilisanda, "Pengaruh Pembelajaran Daring Terhadap Minat Belajar Mahasiswa Pada Masa Pandemi Covid-19," *Behav. Account. J.*, vol. 3, no. 1, pp. 37–46, 2020, doi: 10.33005/baj.v3i1.57.
- [6] N. H. Zhafira, Y. Ertika, and Chairiyaton, "Persepsi Mahasiswa Terhadap Perkuliahan Daring Sebagai Sarana Pembelajaran Selama Masa Karantina Covid-19," *J. Bsisns dan Kaji. Strateg. Manaj.*, vol. 4, pp. 37–45, 2020.
- [7] S. F. Sepita and S. Suryanti, "Pengaruh Pembelajaran Daring Terhadap Hasil Belajar Kognitif Mahasiswa Pada Mata Kuliah Limnologi," *J. Res. Educ. Chem.*, vol. 2, no. 2, p. 102, 2020, doi: 10.25299/jrec.2020.vol2(2).5826.
- [8] M. Mulyana, B. H. Rainanto, D. Astrini, and R. Puspitasari, "Persepsi Mahasiswa Atas Penggunaan Aplikasi Perkuliahan Daring Saat Wabah Covid-19," *JAS-PT (Jurnal Anal. Sist. Pendidik. Tinggi Indones.)*, vol. 4, no. 1, p. 47, 2020, doi: 10.36339/jaspt.v4i1.301.
- [9] M. Turmuzi and N. Hikmah, "Hubungan Pembelajaran Daring Google Classroom Pada Masa COVID-19 dan Motivasi Belajar Terhadap Hasil Belajar Mahasiswa Pendidikan Matematika," *J. Cendekia J. Pendidik. Mat.*, vol. 05, no. 02, pp. 1512–1523, 2021.
- [10] W. Fitriani, H. Haryanto, and S. E. Atmojo, "Motivasi Berprestasi dan Kemandirian Belajar Mahasiswa saat Pembelajaran Daring," *J. Pendidik. Teor. Penelitian, dan Pengemb.*, vol. 5, no. 6, pp. 828–834, 2020, [Online]. Available: <http://journal.um.ac.id/index.php/jptpp/article/view/13639>.
- [11] E. M. Rakhmawati, "Analisis Faktor Pendukung Hasil Pembelajaran Daring pada Anak Berkebutuhan Khusus," *Pros. Semin. Nas. Pascasarj. Univ. Negeri Semarang*, 2020, [Online]. Available: <https://proceeding.unnes.ac.id/index.php/snpasca/article/download/569/488>.
- [12] A. R. Riyanda, K. Herlina, and B. A. Wicaksono, "Evaluasi Implementasi Sistem Pembelajaran Daring Fakultas Keguruan dan Ilmu Pendidikan Universitas Lampung," *J. IKRA-ITH Hum.*, vol. 4, no. 1, pp. 66–71, 2020, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-humaniora/article/view/669>.
- [13] N. Apriyanti, "Analisis Evaluasi Pembelajaran Daring Berorientasi pada Karakter Siswa," *J. Pendidik. Dasar*, 2020.
- [14] N. B. Argaheni, "Sistematis Review: Dampak Perkuliahan Daring Saat Pandemi COVID-19 Terhadap Mahasiswa Indonesia," *PLACENTUM J. Ilm. Kesehat. dan Apl.*, vol. 8, no. 2, p. 99, 2020, doi: 10.20961/placentum.v8i2.43008.
- [15] Purniawan and W. Sumarni, "Analisis Respon Siswa Pada Pembelajaran Daring di Masa Pandemi Covid 19," *Semin. Nas. Pascasarj. UNNES*, pp. 784–789, 2020.
- [16] A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, "Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 88–94, 2021, doi: 10.37034/jsisfotek.v3i3.137.
- [17] A. Natuzzuhriyyah and N. Nafisah, "Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 61–67, 2021, doi: 10.36805/technoexplo.v6i2.1377.
- [18] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021.
- [19] A. Tangu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.
- [20] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [21] U. Fayyad, G. P. Saphiro, and P. Smyth, "Knowledge Discovery and Data Mining : Toward a Unifying Framework," in *Knowledge Data Discovery 1996*, 1997, vol. 9, no. 6, pp. 851–860, doi: 10.3156/jfuzzy.9.6_851.

Commented [A7]: Blm ada future works/saran



JEPIN

(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. x
No. y
mm yy

Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring

Wiyli Yustanti¹, Naim Rochmawati²

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya
Kampus Unesa Keintang, Surabaya

¹wiyliyustanti@unesa.ac.id

²naimrochmawati@unesa.ac.id

Abstrak — Keberhasilan pelaksanaan pembelajaran secara daring di masa pandemi dari sisi mahasiswa dipengaruhi oleh faktor eksternal dan internal. Faktor eksternal antara lain ketersediaan sinyal atau jaringan yang baik, kuota internet, serta perangkat penunjang seperti smartphone dan laptop. Selain itu, aspek internal seperti motivasi belajar, budaya belajar, dan kondisi fisik serta psikologi yang baik juga berperan penting dalam keberhasilan proses belajar mahasiswa. Faktor-faktor ini menjadi input untuk membangun model prediksi karakteristik mahasiswa peserta daring. Pada proses pemodelan diawali dengan tahap *pre-processing* melalui seleksi fitur menggunakan uji independen Chi-Square untuk menentukan variabel yang berpengaruh pada proses prediksi variabel respon. Hasil seleksi variabel independen menghasilkan 16 variabel yang berpengaruh dari total 22 variabel awal. Adapun jenis label pada variabel respon terdiri dari 4 kelompok yaitu mahasiswa aktif dengan fasilitas terbatas, aktif dengan fasilitas baik, pasif dengan fasilitas baik dan pasif dengan fasilitas terbatas. Pada data penelitian ini, distribusi variabel respon termasuk dalam kategori tidak seimbang (*imbalanced class*) dengan proporsi kelas terkecil adalah 2,20%. Untuk melakukan *balancing* data digunakan teknik *oversampling* sebelum tahap pemodelan. Sementara algoritma klasifikasi yang diujicobakan terdiri dari 6 algoritma yaitu Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Regresi Logistik (LR), K-Nearest Neighbor (KNN), dan Decision Tree (DT). Hasil komparasi kinerja menunjukkan bahwa algoritma Support Vector Machine (SVM) lebih unggul dengan nilai F-1 Score 92,8% dan AUC sebesar 99,01%.

Kata kunci — F1-Score, AUC, Imbalanced, Klasifikasi, Oversampling

I. PENDAHULUAN

Pada masa pandemi, proses pembelajar di berbagai institusi pendidikan dilakukan secara daring dan menggantikan proses pembelajaran tatap muka. Pembelajaran dikelas dapat dilakukan secara virtual melalui media komunikasi yang mendukung tatap muka secara daring, seperti aplikasi media sosial atau platform

e-learning. Dengan demikian, siswa tetap dapat mengikuti proses pembelajaran tanpa terhalang lokasi dan waktu. Pembelajaran secara daring sudah menjadi solusi untuk menghentikan laju persebaran virus. Hal ini dilakukan sebagai upaya pemerintah melalui kebijakannya secara bertahap untuk mewajibkan semua proses pembelajaran dilakukan secara daring sampai waktu kondusif untuk memulai pembelajaran tatap muka. Upaya ini diharapkan bahwa melalui metode pembelajaran daring, dapat memudahkan siswa dalam mengakses setiap materi pembelajaran, berinteraksi atau melakukan diskusi antara satu dengan yang lain sehingga tercapainya hasil belajar yang optimal.

Pelaksanaan pembelajaran secara daring tentunya memerlukan dukungan ketersediaan fasilitas seperti kualitas sinyal atau jaringan yang baik, kuota internet, smartphone, dan laptop. Dukungan sarana prasarana ini, berpengaruh terhadap kualitas proses pembelajaran daring yang dapat dilihat dari hasil belajar melalui nilai yang diperoleh. Selain aspek fasilitas, faktor internal siswa juga memiliki peran dalam keberhasilan belajar. Fadila, dkk [1] menjelaskan bahwa kemandirian belajar adalah bahwa siswa dapat belajar tanpa tergantung orang lain dalam proses pemahaman materi, penerapan pengetahuan, serta penyelesaian masalah. Pada penelitian lain, Keberhasilan belajar akan tercapai bila seorang pelajar memiliki dukungan fasilitas yang baik, motivasi belajar, budaya belajar serta memiliki kondisi fisik dan psikologi yang baik. Penelitian terkait dengan faktor yang berpengaruh terhadap hasil pembelajaran daring banyak ditemukan antara lain, Anggrawan, dkk [2] dalam hasil penelitiannya menyimpulkan bahwa ada kaitan antara gaya belajar dengan performansi belajar mahasiswa pada studi Bahasa Inggris. Dimana pada pembelajaran daring, mahasiswa tipe auditori dan visual rata-rata memiliki nilai lebih unggul dibanding dengan mahasiswa tipe kinestetik. Sedangkan hasil penelitian Andrianto, dkk [3] menjelaskan secara umum, faktor yang mempengaruhi kesuksesan penerapan e-learning di Indonesia terdiri dari

Commented [w1]: Respon terhadap reviewer B, memperbaiki struktur kalimat agar lebih runtun untuk dibaca

Commented [w2]: Respond terhadap reviewer B, Menyusun kata kunci berdasarkan abjad

tiga hal yang utama yaitu infrastruktur, kualitas sumber daya manusia dan institusi pendidikan itu sendiri. Ketiga hal ini membutuhkan pemerintah untuk mempersiapkan kegiatan pembelajaran secara dari dapat berjalan lancar. Penjelasan yang senada juga ditemukan pada penelitian Bambang [4].

Aspek internal siswa juga berpengaruh terhadap kualitas proses belajar secara daring, khususnya dalam membangun motivasi dan minat belajar. Syarifah, dkk [5] menemukan bahwa ada pengaruh signifikan antara aspek penggunaan kuota internet, cara pandang, budaya belajar, kondisi ekonomi terhadap minat belajar mahasiswa. Minat belajar ini, juga dipengaruhi oleh media pembelajaran daring yang digunakan, hasil penelitian Nabila, dkk [6] menunjukkan bahwa lebih dari 50% mahasiswa lebih menyukai pembelajaran melalui Google Classroom (GCR) dan Whatsapp Group (WAG). Hal ini juga dibuktikan bahwa, penggunaan Google Meet pada saat pembelajaran daring memberikan nilai yang baik dan sangat baik berdasarkan hasil penelitian Sepita [7]. Penelitian Mumuh [8] dan Turmuzi [9] juga memberikan kesimpulan bahwa penggunaan GCR dan WAG menjadi pilihan terbaik untuk proses pembelajaran daring karena ramah kuota internet. Windi, dkk [10] menemukan dalam penelitiannya bahwa kemandirian belajar dipengaruhi oleh motivasi berprestasi, dimana semakin tinggi motivasi seseorang untuk belajar maka tingkat kemandiriannya juga semakin tinggi dan sebaliknya. Berbeda halnya dengan siswa berkebutuhan khusus, dimana umur, gender dan pola asuh memberikan pengaruh yang cukup signifikan [11].

Implementasi pembelajaran secara daring sudah dilaksanakan hampir dua tahun di seluruh intitusi pendidikan di Indonesia. Salah satu hasil evaluasinya dapat dilihat pada Riyanda, dkk [12] yang menyimpulkan bahwa pembelajaran secara daring baik dari sisi aspek konteks, input, proses, dan output, berhasil dengan baik dan dapat dilanjutkan. Sedangkan pada hasil evaluasi Nurlina [13], menyimpulkan bahwa dengan pembelajaran daring guru lebih aktif dalam menyiapkan bahan ajar serta kemampuan menggunakan media komunikasi digital untuk pembelajaran lebih baik. Dalam hasil penelitian lain ditemukan beberapa dampak akibat pembelajaran daring antara lain mahasiswa masih sering mengalami kebingungan, pasif, kurang produktif dan tidak kreatif, stress, penumpukan konsep/informasi yang kurang bermanfaat serta literasi bahasa yang meningkat [14]. Purniawan [15] juga memberikan informasi terkait analisis respon siswa terhadap pembelajaran daring yang menunjukkan hasil cukup baik meskipun ketersediaan kuota internet sering menjadi kendala.

Aspek faktor-faktor yang ditemukan berpengaruh dan hasil evaluasi proses pelaksanaan pembelajaran daring selama ini memberikan informasi secara kuantitatif bahwa data-data tersebut dapat diekplorasi lebih dalam untuk membangun model prediksi karakteristik mahasiswa berdasarkan variabel-variabel yang diduga berpengaruh signifikan. Penggunaan algoritma Naïve Bayes [16] diterapkan untuk memprediksi kepuasaan mahasiswa pada

proses pembelajaran daring dengan menggunakan variabel yang mengukur komunikasi, suasana pembelajaran, penilaian dan penyampaian materi. Penelitian serupa juga dilakukan Natuzzuhriyyah, dkk [17] yang menggunakan tujuh variabel independen sebagai prediktor tingkat kepuasan. Algoritma klasifikasi juga digunakan pada evaluasi hasil pembelajaran daring berbasis pada analisis sentiment di media sosial Twitter [18],[19] dan [20].

Berdasarkan dengan perkembangan penelitian pembelajaran daring di masa pandemi yang cukup beragam, maka pada penelitian ini ingin dilakukan analisis faktor yang berpengaruh terhadap karakteristik mahasiswa pada pembelajaran daring berdasarkan berbagai aspek yang ditemukan berpengaruh signifikan pada penelitian terkait dengan menggunakan pendekatan algoritma klasifikasi. Terdapat enam algoritma klasifikasi yang akan dibandingkan yaitu Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Regresi Logistik (LR), K-Nearest Neighbor (KNN), dan Decision Tree (DT). Tujuannya adalah agar proses pembelajaran dapat dimanajemen dengan baik sesuai dengan karakteristik siswa yang ada melalui pemilihan model yang menunjukkan kinerja prediksi terbaik.

II. METODE

Metodologi dalam penelitian ini mengadopsi proses yang ada pada kerangka kerja *Knowledge Data Discovery* (KDD) yang diusulkan oleh Fayyad [21]. Inti utama dari kerangka KDD terdiri dari lima tahap yaitu tahap pemilihan data, pre-processing, tranformasi, pemodelan, dan evaluasi.

A. Data

Sumber data dalam penelitian iini merupakan data primer yang diperoleh dari hasil isian formulir mahasiswa peserta mata kuliah Basis Data yang diambil secara random. Didapatkan 91 responden dengan instrument isian terdiri dari 22 variabel yang meliputi aspek internal dan eksternal mahasiswa saat mengikuti perkuliahan daring. Secara rinci, variabel dependen atau data kelas/target untuk pemodelan klasifikan menggunakan nilai pada Tabel I.

TABEL I
VARIABEL DEPENDEN DATASET MAHASISWA

Simbol	Variabel	Tipe Data	Nilai Data	Keterangan
y	Kelompok Siswa	Kategorik	0	Pembelajar aktif dengan dukungan fasilitas terbatas
			1	Pembelajar aktif dengan dukungan fasilitas baik
			2	Pembelajar pasif dengan dukungan fasilitas baik
			3	Pembelajar pasif denag dukungan fasilitas terbatas

Sedangkan untuk variabel independent secara lengkap dapat dilihat pada Tabel II. Semua data yang digunakan

Commented [w3]: Respon terhadap Reviewer B : Novelty dari penelitian ini adalah melakukan analisis faktor yang memmentukan karakteristik mahasiswa dengan pendekatan algoritma klasifikasi (supervised learning)

dalam penelitian ini bertipe kategorik dengan skala Nominal. e) Memutuskan Terima atau Tolak Hipotesis Null.

TABEL II
VARIABEL INDEPENDEN DATASET MAHSISWA

Simbol	Fitur	Tipe Data	Nilai Data
x_1	Jenis kelamin	Kategorik	0 = Pria, 1 = Wanita
x_2	Berat Badan	Kategorik	1 = Kurus, 2 = Normal, 3 = Gemuk, 4 = Obesitas
x_3	Kacamata	Kategorik	1 = Ya, 0 = Tidak
x_4	Pernah Sakit	Kategorik	1 = Ya, 0 = Tidak
x_5	Pernah Gangguan Mental	Kategorik	1 = Ya, 0 = Tidak
x_6	Aktif Bertanya	Kategorik	1 = Ya, 0 = Tidak
x_7	Aktif Menjawab	Kategorik	1 = Ya, 0 = Tidak
x_8	Aktif mengerjakan Tugas	Kategorik	0 = Sebagian, 1 = Semua
x_9	Minat	Kategorik	0 = Tidak, 1 = Mungkin, 2 = Ya
x_{10}	Waktu Belajar	Kategorik	1 = < 5 jam, 2 = 5 < t < 10 jam, 3 = > 10 jam
x_{11}	Ketersediaan Sumber Belajar Lain	Kategorik	1 = Ya, 0 = Tidak
x_{12}	Aktif mencari tutorial	Kategorik	1 = Ya, 0 = Tidak
x_{13}	Mengulang Materi	Kategorik	0 = Tidak, 1 = Terkadang 2 = Selalu
x_{14}	Melakukan Latihan	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Selalu
x_{15}	Kegiatan Diskusi	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Selalu
x_{16}	Smartphone	Kategorik	1 = Ya, 0 = Tidak
x_{17}	Laptop	Kategorik	1 = Ya, 0 = Tidak
x_{18}	Internet	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Ya
x_{19}	Dukungan Lingkungan Rumah	Kategorik	0 = Tidak, 1 = Terkadang, 2 = Ya
x_{20}	Listrik	Kategorik	1 = Ya, 0 = Tidak
x_{21}	Daerah	Kategorik	1 = Pedesaan, 2 = Perkotaan, 3 = Pesisir
x_{22}	Kemudahan akses telekomunikasi	Kategorik	1 = Buruk, 2 = Sedang, 3 = Baik

B. Pre-Processing

Pada tahap ini variabel independen (x) pada dataset mahasiswa di seleksi kembali dengan cara memilih fitur yang berpengaruh dan membuang fitur yang tidak berpengaruh sehingga diharapkan dapat meningkatkan nilai akurasi saat dilakukannya proses klasifikasi, untuk menguji variabel independen digunakan uji *Chi-Square* dengan langkah-langkah dalam pengujian sebagai berikut :

- Menentukan Hipotesis.
- Membuat Tabel Kontingensi.
- Menentukan Nilai Harapan
- Menghitung Statistik Chi-Square.

C. Transformasi

Bentuk transformasi dapat berupa perubahan tipe data dari kategorik menjadi numerik maupun perubahan skala data menjadi range yang standard. Pada penelitian ini dilakukan transformasi data kategorik menggunakan *Label Encoding* yaitu mengkonversi secara langsung data text menjadi nilai integer yang bermakna nominal tanpa mempertimbangkan urutan. Setelah dilakukan koversi ke numerik, selanjutnya data distandarisasi dengan rumus :

$$x_{\text{baru}} = \frac{x_{\text{lama}} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

D. Pemodelan

Proses pemodelan merupakan tahapan membangun model prediksi klasifikasi berdasarkan nilai-nilai pada variabel input terhadap variabel output (kelas). Untuk mendapatkan model prediksi yang terbaik, dilakuka eksperimen dengan beberapa algoritma klasifikasi yaitu Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Regresi Logistik (RL), Random Forest (RF) dan Support Vector Machine (SVM). Proses optimasi parameter menggunakan metode Grid Search.

E. Validasi dan Evaluasi

Metode validasi yang digunakan untuk dataset yang memiliki distribusi kelas yang tidak seimbang adalah menggunakan Teknik oversampling sehingga distribusi menjadi seimbang. Untuk memastikan bahwa semua kelas akan terwakili pada proses training dan testing, digunakan Stratified-K Cross Validation dengan nilai fold adalah 2 sampai 10. Sedangkan metrik evaluasi kinerja prediksi kelas menggunakan F1-Score. Untuk kasus distribusi kelas yang seimbang, nilai Akurasi dan F1-Score akan menghasilkan angka yang sama. Ukuran lain yang dapat digunakan adalah nilai AUC. Adapun interpretasi nilai AUC dapat mengacu pada Tabel III.

TABEL III
KATEGORI NILAI AUC

AUC Score	Interpretasi Hasil Klasifikasi
0.90 – 1.00	Sangat Baik
0.80 – 0.90	Baik
0.70 – 0.80	Sedang
0.60 – 0.70	Buruk
0.50 – 0.60	Gagal

III. HASIL DAN PEMBAHASAN

Bagian ini membahas tentang hasil dari setiap tahapan yang sudah dijelaskan pada bagian sebelumnya mulai dari hasil seleksi variabel yang berpengaruh terhadap variabel respon sampai dengan hasil evaluasi dan validasi untuk pemilihan model yang terbaik.

Commented [w4]: Respon terhadap Reviewer B : Perbaikan struktur kalimat

A. Seleksi Variabel

Berdasarkan hasil dari seleksi fitur dengan uji independen *Chi-Square* menggunakan nilai $\alpha=10\%$ didapatkan hasil seperti pada Tabel IV.

TABEL IV
SELEKSI FITUR DATASET DENGAN UJI CHI-SQUARE

X_i	χ^2_{hitung}	Likelihood	df	χ^2_{tabel}	H_0
x_1	6,398	7,219	3	6,251	Ditolak
x_2	13,389	15,874	9	14,684	Diterima
x_3	8,146	8,737	3	6,251	Ditolak
x_4	17,766	18,846	3	6,251	Ditolak
x_5	19,505	19,212	3	6,251	Ditolak
x_6	11,358	12,905	3	6,251	Ditolak
x_7	5,886	5,978	3	6,251	Diterima
x_8	60,753	17,745	3	6,251	Ditolak
x_9	70,471	3,725	6	10,645	Ditolak
x_{10}	3,189	3,786	3	6,251	Diterima
x_{11}	14,402	15,091	3	6,251	Ditolak
x_{12}	3,556	5,111	3	6,251	Diterima
x_{13}	24,708	26,110	3	6,251	Ditolak
x_{14}	23,035	27,027	3	6,251	Ditolak
x_{15}	16,058	15,546	6	10,645	Ditolak
x_{16}	2,162	2,311	3	6,251	Diterima
x_{17}	27,053	15,733	3	6,251	Ditolak
x_{18}	29,044	33,973	6	10,645	Ditolak
x_{19}	32,126	34,698	6	10,645	Ditolak
x_{20}	14,783	6,237	3	6,251	Ditolak
x_{21}	7,874	9,463	6	10,645	Diterima
x_{22}	32,385	37,358	6	10,645	Ditolak

Dari hasil seleksi fitur pada Tabel III diketahui variabel independen yang tidak digunakan dalam proses selanjutnya adalah $x_2, x_7, x_{10}, x_{12}, x_{16}$, dan x_{21} , sedangkan sisanya akan digunakan sebagai variabel input pada proses pemodelan klasifikasi. Dengan demikian, pada penelitian ini, menggunakan 16 variabel yang terpilih dari 22 variabel awal.

B. Kinerja Algoritma Klasifikasi

Sebagaimana yang dijelaskan dalam bagian metode, bahwa percobaan dilakukan dengan menggunakan enam algoritma klasifikasi yang berbeda, yaitu DT, KNN, NB, LR, RF dan SVM. Untuk mengetahui apakah proses seleksi fitur mempengaruhi hasil kinerja klasifikasi, maka dilakukan pengukuran kinerja klasifikasi menggunakan nilai F1-Score sebelum dan sesudah seleksi fitur. Tabel V dan VI menunjukkan hasil dari percobaan dengan mengambil nilai K-Fold muladi dari 2 sampai 10.

TABEL V
NILAI F1-SCORE SEBELUM SELEKSI FITUR

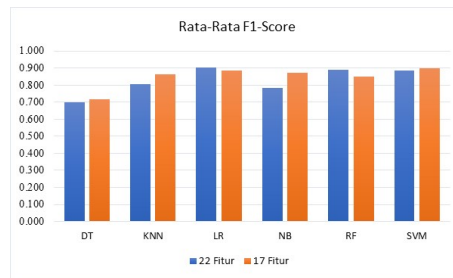
K	DT	KNN	LR	NB	RF	SVM
2	0.681	0.772	0.840	0.780	0.836	0.889
3	0.717	0.762	0.930	0.788	0.826	0.905
4	0.691	0.800	0.879	0.779	0.879	0.851
5	0.727	0.823	0.919	0.795	0.930	0.903
6	0.721	0.815	0.909	0.785	0.912	0.879
7	0.665	0.782	0.927	0.799	0.904	0.900
8	0.699	0.798	0.920	0.778	0.886	0.881
9	0.706	0.848	0.908	0.784	0.904	0.869
10	0.667	0.838	0.908	0.764	0.925*	0.889

Didapatkan bahwa baik kinerja sebelum maupun sesudah seleksi fitur menunjukkan bahwa K-Fold sebanyak 10 memberikan hasil tertinggi. Untuk dataset sebelum diseleksi fitur, algoritma RF memberikan hasil kinerja prediksi tertinggi yaitu sebesar 92,5%. Sedangkan untuk dataset setelah seleksi fitur, algoritma SVM memiliki performansi yang paling tinggi yaitu sebesar 92,8% atau sekitar 0,03% lebih unggul daripada algoritma RF pada dataset tanpa seleksi fitur.

TABEL VI
NILAI F1-SCORE SETELAH SELEKSI FITUR

	DT	KNN	LR	NB	RF	SVM
2	0.654	0.882	0.881	0.856	0.714	0.865
3	0.717	0.832	0.879	0.853	0.815	0.855
4	0.745	0.847	0.853	0.864	0.842	0.895
5	0.779	0.845	0.912	0.874	0.872	0.893
6	0.717	0.858	0.862	0.868	0.857	0.912
7	0.698	0.873	0.911	0.904	0.904	0.904
8	0.724	0.887	0.879	0.883	0.865	0.914
9	0.727	0.869	0.896	0.880	0.905	0.922
10	0.674	0.888	0.878	0.886	0.881	0.928*

Selanjutnya, analisis dilakukan dengan melihat nilai rata-rata F1-Score untuk semua algoritma pada kondisi sebelum seleksi fitur yaitu terdiri dari 22 variabel independen, dan setelah proses seleksi fitur dengan menggunakan 17 variabel. Secara umum terlihat bahwa untuk algoritma DT, KNN, NB dan SVM memiliki kinerja yang lebih tinggi dibandingkan dengan kinerja hasil model prediksi klasifikasi sebelum adanya proses seleksi fitur. Sedangkan untuk algoritma LR dan RF memiliki kinerja lebih rendah yaitu penurunan masing-masing sebesar 2,1% dan 3,9%. Secara visual, histogram nilai rata-rata F1 Score dengan jumlah fitur berbeda terlihat pada Gambar 1.



Gambar. 1 Perbandingan Nilai Rata-Rata F1-Score Algoritma Klasifikasi Sebelum dan Sesudah Seleksi Fitur

Commented [w5]: Respon terhadap Reviewer B : Gambar sudah ada

Dari keseluruhan proses eksperimen yang sudah dilakukan, kinerja tertinggi ditunjukkan oleh algoritma SVM. Selanjutnya sebagai konfirmasi dari hasil ini, dilakukan perhitungan nilai AUC. Hasil perhitungan AUC ditampilkan pada Tabel VII.

TABEL VII
NILAI AUC PADA K-FOLD 10

Jumlah Fitur	Nilai AUC untuk KFold = 10					
	DT	KNN	LR	NB	RF	SVM
22 Fitur	0.846	0.961	0.979	0.962	0.982	0.995
17 Fitur	0.715	0.865	0.884	0.874	0.851	0.899

Hasil AUC menunjukkan bahwa algoritma SVM tetap memberikan hasil prediksi yang tertinggi dibandingkan dengan algoritma yang lain. Nilai AUC dan F1-Score merupakan metrik evaluasi yang sangat kuat dan bekerja sangat baik untuk banyak masalah klasifikasi. Salah satu perbedaan besar antara F1-Score dan AUC adalah bahwa F1-Score menekankan pada ketepatan kelas yang diprediksi dan AUC menggunakan skor yang diprediksi sebagai input. Jika dataset yang dianalisis memiliki distribusi yang sangat tidak seimbang dan hasil prediksi sangat mempertimbangkan ketepatan prediksi untuk kelas minoritas, maka yang paling tepat adalah menggunakan F1-Score. Sebaliknya, jika dataset yang dianalisis memiliki distribusi kelas yang hampir seimbang, maka dapat digunakan metrik evaluasi seperti Akurasi dan AUC.

IV. KESIMPULAN DAN SARAN

Dari hasil penelitian ini dapat disimpulkan bahwa karakteristik mahasiswa yang mengikuti perkuliahan daring dapat diprediksi menggunakan algoritma SVM dengan sangat baik yakni memiliki nilai ketepatan prediksi kelas sebesar 92,8% dan nilai AUC 89,9% dengan menggunakan variabel input sebanyak 17 fitur, yaitu dengan mengeluarkan variabel $x_2, x_7, x_{10}, x_{12}, x_{16}$, dan x_{21} .

Untuk penelitian berikutnya dapat dilanjutkan dengan menambahkan variabel aspek pengajar dan institusi Pendidikan, sehingga model klasifikasi yang dibangun tidak hanya dilihat dari sudut pandang mahasiswa.

UCAPAN TERIMA KASIH / ACKNOWLEDGMENT

Kami mengucapkan terimakasih atas dukungan Lembaga Penelitian Universitas Negeri Surabaya atas hibah penelitian kebijakan Fakultas.

DAFTAR PUSTAKA

[1] R. N. Fadila, T. A. Nadiroh, R. Juliana, P. Z. H. Zulfa, and I. Ibrahim, "Kemandirian Belajar Secara Daring Sebagai Prediktor Hasil Belajar Mahasiswa Pendidikan Matematika UIN Sunan Kalijaga," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 2, pp. 880–891, 2021, doi: 10.31004/cendekia.v5i2.457.

[2] A. Anggrawan, "Analisis Deskriptif Hasil Belajar Pembelajaran Tatap Muka dan Pembelajaran Online Menurut Gaya Belajar Mahasiswa," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 18, no. 2, pp. 339–346, 2019, doi: 10.30812/matrik.v18i2.411.

[3] R. Andrianto Pangondian, P. Insap Santosa, and E. Nugroho, "Faktor - Faktor Yang Mempengaruhi Kesuksesan Pembelajaran Daring Dalam Revolusi Industri 4.0," *Sainteks 2019*, pp. 56–60, 2019, [Online]. Available: <https://seminar-id.com/seminas-sainteks2019.html>.

[4] B. Budhianto, "Analisis perkembangan dan faktor yang mempengaruhi keberhasilan pembelajaran daring (e-learning)," *J. AgriWidya*, vol. 1, no. 1, pp. 11–29, 2020.

[5] S. H. Jamil and I. D. Aprilisanda, "Pengaruh Pembelajaran Daring Terhadap Minat Belajar Mahasiswa Pada Masa Pandemi Covid-19," *Behav. Account. J.*, vol. 3, no. 1, pp. 37–46, 2020, doi: 10.33005/baj.v3i1.57.

[6] N. H. Zhafira, Y. Ertika, and Chairiyaton, "Persepsi Mahasiswa Terhadap Perkuliahan Daring Sebagai Sarana Pembelajaran Selama Masa Karantina Covid-19," *J. Bisnis dan Kaji. Strateg. Manaj.*, vol. 4, pp. 37–45, 2020.

[7] S. F. Sepita and S. Suryanti, "Pengaruh Pembelajaran Daring Terhadap Hasil Belajar Kognitif Mahasiswa Pada Mata Kuliah Limnologi," *J. Res. Educ. Chem.*, vol. 2, no. 2, p. 102, 2020, doi: 10.25299/jrec.2020.vol2(2).5826.

[8] M. Mulyana, B. H. Rainanto, D. Astrini, and R. Puspitasari, "Persepsi Mahasiswa Atas Penggunaan Aplikasi Perkuliahan Daring Saat Wabah Covid-19," *JAS-PT (Jurnal Anal. Sist. Pendidik. Tinggi Indones.)*, vol. 4, no. 1, p. 47, 2020, doi: 10.36339/jaspt.v4i1.301.

[9] M. Turmuzi and N. Hikmah, "Hubungan Pembelajaran Daring Google Classroom Pada Masa COVID-19 dan Motivasi Belajar Terhadap Hasil Belajar Mahasiswa Pendidikan Matematika," *J. Cendekia J. Pendidik. Mat.*, vol. 05, no. 02, pp. 1512–1523, 2021.

[10] W. Fitriani, H. Haryanto, and S. E. Atmojo, "Motivasi Berprestasi dan Kemandirian Belajar Mahasiswa saat Pembelajaran Daring," *J. Pendidik. Teor. Penelitian, dan Pengemb.*, vol. 5, no. 6, pp. 828–834, 2020, [Online]. Available: <http://journal.um.ac.id/index.php/jptpp/article/view/13639>.

[11] E. M. Rakhmawati, "Analisis Faktor Pendukung Hasil Pembelajaran Daring pada Anak Berkebutuhan Khusus," *Pros. Semin. Nas. Pascasarj. Univ. Negeri Semarang*, 2020, [Online]. Available: <https://proceeding.unnes.ac.id/index.php/snpasca/article/download/569/488>.

[12] A. R. Riyanda, K. Herlina, and B. A. Wicaksono, "Evaluasi Implementasi Sistem Pembelajaran Daring Fakultas Keguruan dan Ilmu Pendidikan Universitas Lampung," *J. IKRA-ITH Hum.*, vol. 4, no. 1, pp. 66–71, 2020, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-humaniora/article/view/669>.

[13] N. Apriyanti, "Analisis Evaluasi Pembelajaran Daring Berorientasi pada Karakter Siswa," *J. Pendidik. Dasar*, 2020.

[14] N. B. Argaheni, "Sistematis Review: Dampak Perkuliahan Daring Saat Pandemi COVID-19 Terhadap Mahasiswa Indonesia," *PLACENTUM J. Ilm. Kesehatan dan Apl.*, vol. 8, no. 2, p. 99, 2020, doi: 10.20961/placentum.v8i2.43008.

[15] Purniawan and W. Sumarni, "Analisis Respon Siswa Pada Pembelajaran Daring di Masa Pandemi Covid 19," *Semin. Nas. Pascasarj. UNNES*, pp. 784–789, 2020.

[16] A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, "Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naive Bayes," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 88–94, 2021, doi: 10.37034/jsisfotek.v3i3.137.

[17] A. Natuzzuhriyyah and N. Nafisah, "Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naive Bayes," *Techno Xplora J. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 61–67, 2021, doi: 10.36805/technoxplora.v6i2.1377.

[18] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021.

[19] A. Tangu Mara, E. Sediyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.

[20] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.

Commented [w6]: Respon terhadap reviewer B : sudah ditambahkan saran

Commented [w7]: Respon terhadap reviewer B : Saran untuk penelitian berikutnya

- [21] U. Fayyad, G. P. Saphiro, and P. Smyth, "Knowledge Discovery and Data Mining : Toward a Unifying Framework," in *Knowledge Data Discovery 1996*, 1997, vol. 9, no. 6, pp. 851–860, doi: 10.3156/jfuzzy.9.6_851.



Wiyli Yustanti <wilylyustanti@unesa.ac.id>

JEPIN - KELENGKAPAN ADMINISTRASI & BIAYA PENERBITAN (50452)

4 messages

Jurnal Edukasi dan Penelitian Informatika <jepin@untan.ac.id>
To: wilylyustanti@unesa.ac.id

Fri, Apr 8, 2022 at 4:21 PM

Yth. Bapak/Ibu **Wiyli Yustanti**

Bersama ini kami sampaikan bahwa paper dengan judul : "**Analisis Algoritma Klasifikasi untuk Memprediksi Karakteristik Mahasiswa pada Pembelajaran Daring**" telah **DITERIMA**.

Untuk dapat diterbitkan pada buku jurnal JEPIN edisi **Vol. 8 No. 1 April 2022**, mohon author dapat melengkapi hal-hal berikut selambat-lambatnya tanggal **Minggu, 10 April 2022**.

1. Mengisi formulir yang dapat diunduh di link berikut dan mengirimkannya kembali kepada kami melalui email ini.
https://drive.google.com/file/d/0BxTpSnf_8osMQIVoUTFhUy05TWS/view?usp=sharing
2. Mengirimkan biaya administrasi termasuk ongkos kirim 1 eks jurnal sebesar **Rp 1.000.000,-** ke No. Rekening : **BNI - 0696563721 a.n. Enda Esyudha Pratama**
3. Untuk **tambahan buku** jurnal dikenakan biaya sebesar **Rp 100.000,- / eksemplar**.

Formulir dan tanda bukti transfer dapat dikirim lewat email jepin@untan.ac.id.
Atas perhatiannya kami ucapkan terima kasih

Salam,
Dewan Redaksi JEPIN

=====
e-ISSN: 2548-9364

printed ISSN: 2460-0741<http://jurnal.untan.ac.id/index.php/jepin/index>

Wiyli Yustanti <wilylyustanti@unesa.ac.id>
To: Jurnal Edukasi dan Penelitian Informatika <jepin@untan.ac.id>

Sat, Apr 9, 2022 at 2:42 PM

Terimakasih atas Informasinya, acces untuk unduh formulir masih belum bisa, berikut saya sampaikan bukti transfer dan data untuk pengiriman jurnal hardcopy

Nama : Wiyli Yustanti
Alamat : Jalan Bogangin Baru Blok F Utara/22, Kedurus, Karangpilang, Surabaya, Jawa Timur
HP : 087854407576

Regard,

Wiyli Yustanti, S.Si, M.Kom

Phone: +6287854407576
Information Technology Dept
Faculty of Engineering
State University of Surabaya

[Quoted text hidden]



Bukti Bayar Wiyli Yustanti.jpg
52K

Jurnal Edukasi dan Penelitian Informatika <jepin@untan.ac.id>
To: Wiyli Yustanti <wilyliyustanti@unesa.ac.id>

Sat, Apr 9, 2022 at 5:18 PM

Salam,
Dewan Redaksi JEPIN

=====

e-ISSN: 2548-9364

printed ISSN: 2460-0741

<http://jurnal.untan.ac.id/index.php/jepin/index>

[Quoted text hidden]



Wiyli Yustanti <wilyliyustanti@unesa.ac.id>
To: Jurnal Edukasi dan Penelitian Informatika <jepin@untan.ac.id>

Sun, Apr 10, 2022 at 8:46 AM

Berikut kami kirimkan Formulir Hak Cipta

Regard,

Wiyli Yustanti, S.Si, M.Kom

Phone: +6287854407576
Information Technology Dept
Faculty of Engineering
State University of Surabaya

[Quoted text hidden]

