

## Analisis Pengaruh Media Sosial terhadap Produktivitas Akademik Mahasiswa menggunakan Metode *Decision Tree* dan *Random Forest*

Chatarina Enny Murwaningtyas<sup>1\*</sup>, Angel Kristiamita<sup>2</sup>, Agatha Lintang Antika Ika Putri<sup>3</sup>,  
Fibelia Dwi Puspaningrum<sup>4</sup>, Carolina Dhinda Putri Mahanani<sup>5</sup>

<sup>1,2,3,4,5</sup> Pendidikan Matematika, FKIP, Universitas Sanata Dharma

\*Email: enny@usd.ac.id

### Abstrak

Penelitian ini bertujuan untuk mengevaluasi pengaruh penggunaan media sosial terhadap produktivitas akademik mahasiswa Universitas Sanata Dharma Yogyakarta, diukur melalui Indeks Prestasi Kumulatif (IPK). Metode yang digunakan melibatkan dua model pembelajaran mesin: *Decision Tree* dan *Random Forest*. Data diolah menggunakan teknik penskalaan yang tahan terhadap outlier dan penyeimbangan data melalui teknik oversampling. Hasil penelitian menunjukkan bahwa model *Random Forest* memiliki performa superior dengan akurasi, presisi, recall, dan *F1-score* masing-masing sebesar 90%. Sementara itu, model *Decision Tree* menunjukkan akurasi sebesar 80%, dengan presisi 86%, recall 80%, dan *F1-score* 82%. Analisis pentingnya fitur menunjukkan bahwa 'Fakultas' dan 'Jenis Kelamin' adalah faktor paling signifikan dalam memprediksi IPK mahasiswa. Penelitian ini menyimpulkan bahwa penggunaan *Random Forest* dengan teknik penyeimbangan data dapat meningkatkan akurasi dan keandalan prediksi, memberikan wawasan tentang pemanfaatan media sosial untuk meningkatkan produktivitas akademik mahasiswa.

**Kata kunci:** Media Sosial, Produktivitas Akademik, *Decision Tree*, *Random Forest*.

### Abstract

This study aims to evaluate the impact of social media usage on the academic productivity of Universitas Sanata Dharma Yogyakarta students, measured through their Grade Point Average (GPA). The methods employed involve two machine learning models: *Decision Tree* and *Random Forest*. The data were processed using outlier-resistant scaling techniques and data balancing through oversampling. The results show that the *Random Forest* model outperformed with an accuracy, precision, recall, and *F1-score* of 90% each. Meanwhile, the *Decision Tree* model achieved 80% accuracy, with a precision of 86%, recall of 80%, and *F1-score* of 82%. Feature importance analysis revealed that 'Faculty' and 'Gender' are the most significant factors in predicting students' GPA. This study concludes that employing *Random Forest* with data balancing techniques can enhance prediction accuracy and reliability, providing insights into the optimal use of social media to improve students' academic productivity.

**Keywords:** Social Media, Academic Productivity, *Decision Tree*, *Random Forest*.

### PENDAHULUAN

Teknologi telah menjadi kebutuhan fundamental dalam kehidupan manusia, sejajar dengan kebutuhan dasar lain seperti pangan dan pakaian (Wirany dkk., 2022). Teknologi informasi dan komunikasi, memiliki peran penting dalam memudahkan komunikasi dan mempercepat pertukaran informasi yang sangat dibutuhkan dalam arus globalisasi. Dari masa ke masa, teknologi informasi telah berkembang dari media cetak dan tulis, seperti surat kabar dan majalah, menjadi media elektronik yang lebih canggih seperti telepon dan televisi. Saat ini, era

komunikasi interaktif ditandai oleh integrasi teknologi informasi seperti telepon, radio, dan televisi yang kemudian berkembang menjadi internet, mengubah cara kita menyampaikan informasi dan berinteraksi dalam masyarakat baik di perkotaan maupun pedesaan (Siregar dan Nasution, 2020).

Jumlah pengguna internet di Indonesia tergolong cukup tinggi. Menurut Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), jumlah pengguna internet di Indonesia pada tahun 2024 mencapai 221 juta jiwa dari total populasi 278 juta jiwa pada tahun 2023. Selain

itu, secara keseluruhan, *We Are Social* mencatat terdapat 139 juta pengguna media sosial di Indonesia pada Januari 2024. Peningkatan pengguna internet yang didominasi oleh penggunaan media sosial menunjukkan bahwa masyarakat Indonesia telah meleak teknologi. Banyak masyarakat menggunakan media sosial karena dapat memfasilitasi komunikasi sehingga individu dapat terhubung dengan orang di seluruh dunia dengan cepat. Selain itu, *platform-platform* di media sosial juga menjadi sarana untuk mendapatkan dan menyebarkan informasi, termasuk informasi terkait bidang pendidikan.

Media sosial membantu memfasilitasi mahasiswa dengan komunitas akademik *online* (Berry, 2019). *Platform* yang sering digunakan oleh mahasiswa dalam aktivitas akademik meliputi YouTube, Instagram, TikTok, dan X (Twitter). Media sosial ini menyediakan foto dan video yang mempermudah mahasiswa mendapatkan informasi mengenai pembelajaran. Belajar menggunakan media sosial memungkinkan mahasiswa untuk belajar lebih fleksibel dan aktif (Mendoza dkk., 2022). Saat ini, banyak dosen yang memanfaatkan media sosial untuk membantu mahasiswa dalam memahami materi. Dengan begitu, mahasiswa dapat menemukan sendiri pemahamannya dan mencari informasi mengenai materi di media sosial, baik untuk belajar maupun mengerjakan tugas.

Saat ini, tidak sedikit *influencer* pendidikan yang membuat penjelasan mengenai materi pembelajaran menjadi lebih menyenangkan, sehingga meningkatkan motivasi belajar. Terkadang, beberapa *influencer* pendidikan mengajak *audiens* untuk bertanya terkait tugas, yang kemudian dibahas melalui video. Pada unggahan video tersebut, terdapat kolom komentar yang memberikan kesempatan kepada *audiens* untuk berkomunikasi satu sama lain mengenai tugas yang dibahas. Selain dari *influencer*, kolaborasi antar mahasiswa dalam satu kelas dapat dilakukan dengan berbagi tugas atau catatan melalui media sosial. Hal ini dapat mengundang umpan balik atau komentar dari teman atau *audiens* lain sehingga mereka dapat belajar satu sama lain (Ogbuagu, 2021).

Banyaknya manfaat yang dapat diperoleh dari media sosial dalam konteks pembelajaran akan berpengaruh terhadap produktivitas mahasiswa. Dalam bidang akademik, produktivitas mahasiswa dapat dinilai dengan melihat Indeks Prestasi Kumulatif (IPK). IPK merupakan nilai keberhasilan studi seorang mahasiswa

selama menyelesaikan tingkatan atau tugas-tugas akademik (Rizki dkk., 2023). Berdasarkan penelitian Marna dkk. (2020), IPK dapat dikelompokkan menjadi dua kategori, yaitu kategori IPK lebih besar atau sama dengan 3 dan kategori IPK kurang dari 3. Semakin tinggi nilai IPK berarti semakin tinggi pula keberhasilan studi, yang mencerminkan produktivitas akademik seorang mahasiswa.

Dalam konteks akademik, penggunaan media sosial menawarkan potensi besar untuk mendukung kegiatan belajar mengajar, kolaborasi, dan akses sumber daya pendidikan. Hal ini sejalan dengan Apriansyah dan Antoni (2018), yang mengatakan bahwa media sosial dapat digunakan sebagai alat belajar yang mendorong kolaborasi mahasiswa serta pengembangan keterampilan. Namun, di sisi lain, ada kekhawatiran bahwa penggunaan media sosial yang berlebihan dapat mengganggu konsentrasi dan mengurangi produktivitas akademik mahasiswa. IPK bergantung pada Indeks Prestasi Semester (IPS) yang dapat mengalami kenaikan atau penurunan. Oleh karena itu, untuk mendapatkan IPK yang tinggi, mahasiswa harus berkomitmen untuk mempertahankan IPS. Penting untuk memahami bagaimana penggunaan media sosial mempengaruhi produktivitas akademik mahasiswa.

Dalam penelitian ini, metode *Decision Tree* digunakan untuk memprediksi pengaruh penggunaan media sosial terhadap produktivitas akademik mahasiswa. *Decision Tree* adalah algoritma yang dapat mengidentifikasi hubungan antara variabel *independen* dan *dependen* melalui pembentukan pohon keputusan. Algoritma ini telah digunakan secara luas dalam berbagai penelitian untuk memprediksi hasil akademik, termasuk dalam penelitian yang menggunakan data demografis dan jaringan dukungan mahasiswa untuk memprediksi GPA (Frazier dkk., 2023). *Decision Tree* memungkinkan peneliti untuk memahami variabel mana yang paling signifikan dalam mempengaruhi IPK, memberikan wawasan mendalam mengenai faktor-faktor yang berkontribusi terhadap keberhasilan akademik.

Penelitian yang dilakukan oleh Diantika dkk. (2024) menggunakan *Random Forest* dalam memprediksi ketepatan waktu kelulusan mahasiswa dan memperoleh akurasi sebesar 0,90. Selain itu, penelitian oleh Saleh dkk. (2022) yang juga menggunakan *Random Forest* dalam

klasifikasi kelulusan mahasiswa, menghasilkan akurasi sebesar 0,90. Hasil penelitian ini menunjukkan bahwa *Random Forest* dapat memberikan hasil prediksi yang akurat dalam konteks akademik.

Penelitian ini juga menegaskan pentingnya membandingkan berbagai algoritma *machine learning* dalam konteks prediksi performa akademik mahasiswa. Sebagai contoh, penelitian oleh Budiyanto dkk. (2024) membandingkan kinerja algoritma Naïve Bayes, *Random Forest*, dan C4.5 dalam memprediksi kelulusan dengan predikat *cum laude*. Hasil penelitian menunjukkan bahwa Naïve Bayes memiliki kinerja terbaik, namun *Random Forest* juga menunjukkan hasil yang sangat kompetitif. Hal ini menunjukkan bahwa *Random Forest* dan *Decision Tree* dapat memberikan hasil prediksi yang akurat dan menjadi dasar yang kuat untuk penelitian lebih lanjut dan implementasi praktis dalam sistem pendidikan.

Penelitian ini bertujuan untuk mengetahui pengaruh media sosial terhadap IPK mahasiswa Universitas Sanata Dharma Yogyakarta dengan menggunakan *Decision Tree* dan *Random Forest*. Dengan memahami pengaruh ini, diharapkan dapat memberikan wawasan baru tentang bagaimana media sosial dapat dimanfaatkan secara optimal untuk meningkatkan produktivitas akademik mahasiswa. Hasil penelitian ini juga diharapkan dapat menjadi acuan bagi institusi pendidikan dalam merancang kebijakan yang mendukung penggunaan media sosial secara positif dalam proses pembelajaran.

## TINJAUAN PUSTAKA

### 2.1 Media Sosial

Media sosial merupakan salah satu perkembangan kemajuan teknologi yang digunakan sebagai wadah untuk interaksi sosial sehingga dapat mempermudah komunikasi, berpartisipasi, berbagi ilmu pengetahuan, dan membentuk komunitas yang dapat diakses secara *online*. Media sosial mendorong perguruan tinggi dalam meningkatkan pembelajaran kolaboratif, kreativitas mahasiswa, berbagi informasi, penerbitan, dan partisipasi mahasiswa (Nasiruddin & Rapa', 2022).

Media sosial memiliki beberapa kelebihan. Pengguna dapat mengeksplorasi berbagai hal, seperti pencapaian prestasi, kebutuhan akademik, dan inspirasi dari

pencapaian orang lain, serta mencari peluang dan dukungan untuk mencapai tujuan dan ambisi mereka. Namun, media sosial juga memiliki kekurangan, seperti ketergantungan pada informasi, kesulitan dalam fokus pada pembelajaran, dan dapat merusak manajemen waktu karena keasyikan bermain media sosial.

### 2.2 Produktivitas Mahasiswa

Produktivitas adalah interaksi antara individu dengan lingkungan untuk menghasilkan sesuatu yang mendukung dirinya dalam menghadapi tantangan dari luar. Pada kalangan mahasiswa, produktivitas erat kaitannya dengan keberhasilan dalam menjalani proses perkuliahan. Septiani dan Triariani (Septiani & Triariani, 2022) memaparkan bahwa tingkat produktivitas dipengaruhi oleh faktor teknis dan faktor manusia. Faktor teknis meliputi penerapan metode kerja yang efektif dan efisien, sedangkan faktor manusia mencakup upaya yang dilakukan untuk menyelesaikan tugas dan tanggung jawab.

Pada mahasiswa, faktor teknis yang mempengaruhi produktivitas adalah metode dan model pembelajaran yang digunakan dalam perkuliahan. Sementara itu, faktor manusia mencakup usaha mahasiswa dalam menyelesaikan tugasnya. Indikator yang dapat digunakan untuk menilai tanggung jawab mahasiswa dapat dilihat dari IPK mereka.

### 2.3 Decision Tree

*Decision tree* adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi dengan cara memecah dataset menjadi subset-subset berdasarkan fitur-fitur tertentu. Setiap percabangan dalam pohon keputusan mewakili keputusan berdasarkan nilai fitur, dan setiap daun akhir mewakili hasil atau label kelas. Algoritma ini sangat intuitif dan mudah diinterpretasikan karena strukturnya yang mirip dengan proses pengambilan keputusan manusia. Menurut penelitian, pohon keputusan efektif dalam menangani data yang memiliki hubungan non-linear dan interaksi antar fitur yang kompleks (Frazier dkk., 2023).

Namun, pohon keputusan memiliki kelemahan utama yaitu kecenderungan untuk *overfitting*, terutama pada dataset yang kecil atau memiliki banyak fitur. *Overfitting* terjadi ketika model terlalu cocok dengan data pelatihan, sehingga kehilangan kemampuan untuk menggeneralisasi pada data baru. Untuk

mengatasi masalah ini, metode seperti *pruning* dan penggunaan algoritma *ensemble* seperti *Random Forest* telah dikembangkan untuk meningkatkan kestabilan dan akurasi model (Budiyanto dkk., 2024).

#### 2.4 *Random Forest*

*Random Forest* adalah algoritma pembelajaran mesin yang dikenal memiliki performa terbaik dalam mengklasifikasi dataset dengan jumlah besar (Khalim dkk., 2023). Algoritma ini bekerja dengan membuat banyak pohon keputusan yang kemudian dikombinasikan untuk menghasilkan prediksi yang lebih akurat dan stabil. Setiap pohon dalam hutan dibangun menggunakan metode *bagging*, yang melibatkan pembuatan beberapa *subset* data dari *dataset* asli melalui proses pengambilan sampel acak dengan penggantian. Metode ini membantu mengurangi *overfitting* dengan memperkenalkan variasi dalam proses pelatihan, sehingga model lebih tahan terhadap fluktuasi dan pola yang tidak biasa dalam *dataset* (Schonlau & Zou, 2020).

Menurut penelitian, *Random Forest* adalah algoritma yang kuat karena kemampuannya untuk menghasilkan prediksi yang akurat dan tahan terhadap *overfitting*, terutama dalam kasus data yang kompleks. Dengan pendekatan *ensemble learning*, *Random Forest* mampu menangani berbagai jenis data, baik yang bersifat kategorikal maupun kontinu, serta dapat digunakan untuk tugas-tugas klasifikasi dan regresi. Kekuatan *Random Forest* dalam mengelola data yang besar dan kompleks menjadikannya pilihan yang populer dalam berbagai aplikasi pembelajaran mesin di Indonesia (Budiyanto dkk., 2024).

#### 2.5 Pengaruh Media Sosial terhadap Produktivitas Mahasiswa

Satria dan Hermanto (2022) meneliti pengaruh media sosial terhadap semangat belajar mahasiswa selama pembelajaran daring menggunakan algoritma Naïve Bayes. Penelitian ini menunjukkan bahwa penggunaan media sosial yang bijak dapat meningkatkan semangat belajar mahasiswa, namun penggunaan yang berlebihan berdampak negatif pada manajemen waktu mereka. Sementara itu, Hariyanti dkk. (2024) membandingkan algoritma *Decision Tree* dan Naïve Bayes dalam menganalisis pengaruh media sosial dan pola tidur terhadap prestasi akademik siswa. Hasil penelitian ini

mengungkapkan bahwa *Decision Tree* lebih unggul dalam interpretasi visual dan akurasi dibandingkan dengan Naïve Bayes dalam konteks ini.

Penelitian ini juga relevan dengan studi oleh Pribadi dan Arsyad (2021), yang menganalisis upaya mahasiswa dalam mengatasi adiksi media sosial menggunakan algoritma Naïve Bayes. Penelitian ini menunjukkan bahwa penggunaan media sosial yang tidak terkendali dapat berdampak signifikan terhadap manajemen waktu mahasiswa, mengurangi produktivitas mereka. Sementara itu, Fadli dkk. (Fadli dkk., 2021) menyoroti hubungan negatif yang signifikan antara adiksi gawai dengan prestasi belajar mahasiswa. Penelitian ini mengungkapkan bahwa semakin tinggi tingkat adiksi gawai, semakin rendah prestasi akademik mahasiswa, dengan pengaruh sebesar 16% dalam analisis regresi sederhana.

Namun, penelitian ini melangkah lebih jauh dengan mengeksplorasi bagaimana media sosial dapat dimanfaatkan secara positif untuk mendukung produktivitas akademik. Penelitian ini lebih menekankan pada membandingkan dua algoritma, *Decision Tree* dan *Random Forest*, untuk menentukan model yang terbaik dan paling sesuai dengan data yang digunakan. Dengan demikian, penelitian ini tidak hanya memberikan analisis menyeluruh tetapi juga menawarkan rekomendasi model berdasarkan performa dalam konteks produktivitas akademik mahasiswa. Dengan demikian, penelitian ini menawarkan kontribusi baru dalam memahami pengaruh media sosial melalui pendekatan pembelajaran mesin yang lebih komprehensif.

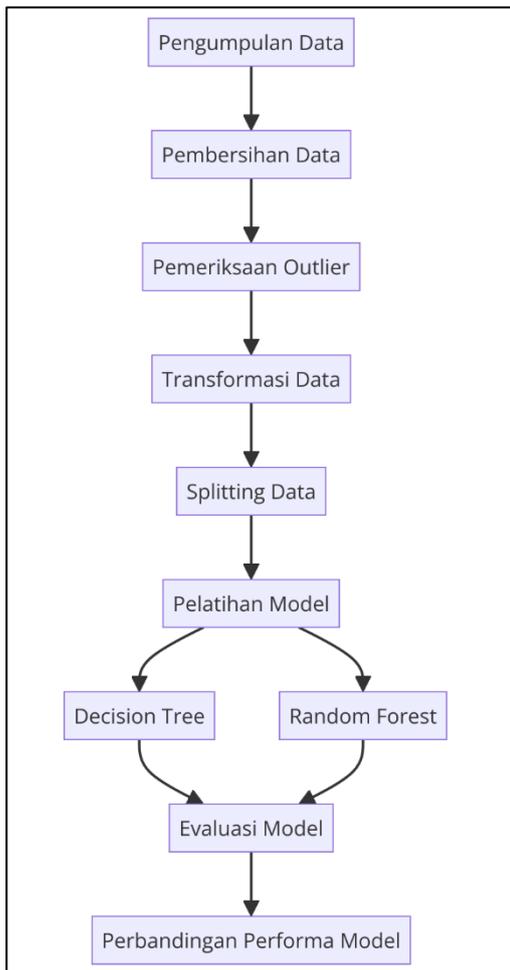
#### METODE PENELITIAN

Pada penelitian ini, metodologi yang digunakan bertujuan untuk menganalisis dampak penggunaan media sosial terhadap produktivitas mahasiswa. Secara garis besar, metode penelitian dapat dilihat pada Gambar 1.

#### 3.1 Pengumpulan Data

Pada penelitian ini, telah terkumpul data sebanyak 136. *Dataset* yang diperoleh ditunjukkan pada Tabel 1. Data dikumpulkan melalui penyebaran angket kepada mahasiswa S1 dan vokasi di Universitas Sanata Dharma. Instrumen ini mencakup aspek demografi, jenis dan durasi penggunaan media sosial, serta perspektif mahasiswa terhadap pengaruh media sosial pada produktivitas akademik. Setiap

indikator dirancang untuk mengumpulkan informasi yang relevan dengan penelitian ini.



Gambar 1. Alur diagram penelitian

Tabel 1. Instrumen Pengumpulan Data

No	Indikator	Isian Jawaban
1.	Jenis Kelamin	Laki-laki Perempuan
2.	Usia	(dalam Tahun)
3.	Fakultas	a. FKIP b. Non FKIP
4.	IPK	a. <3.00 b. ≥ 3.00
5.	Jenis media sosial yang digunakan untuk aktivitas pendidikan	a. Youtube b. Tiktok c. Instragram d. X(twitter)
6.	Durasi penggunaan media sosial dalam aktivitas akademik perhari	(dalam jam)
7.	Pernyataan tentang penggunaan media	a. Ya b. Tidak

No	Indikator	Isian Jawaban
8.	social untuk aktivitas akademik keikutsertaan komunitas media sosial dalam akademik	a. Ya b. Tidak
9.	Pernyataan terlihat diskusi akademik pada media sosial	a. Ya b. Tidak
10.	Perspektif Mahasiswa terhadap pengaruh media sosial pada aktivitas akademik	a. Sangat Setuju b. Setuju c. Tidak Setuju d. Sangat Tidak Setuju

### 3.2 Pembersihan Data

Pembersihan data merupakan langkah penting dalam analisis data untuk memastikan kualitas dan akurasi sebelum diterapkan pada model *Decision Tree* dan *Random Forest*. Langkah pertama dalam pembersihan data adalah menghapus kolom yang tidak relevan, seperti kolom nama dan usia, karena dianggap tidak berpengaruh terhadap produktivitas mahasiswa. Selanjutnya, dilakukan pengecekan *missing value* untuk memastikan bahwa data yang digunakan lengkap dan kredibel.

Pengecekan *missing value* adalah proses krusial untuk mengidentifikasi dan menangani data yang hilang. Data yang lengkap dan berkualitas akan menghasilkan model yang lebih akurat dan dapat meminimalkan kesalahan prediksi. Hasil pengecekan *missing value* ditunjukkan pada Gambar 2, yang memastikan *dataset* siap untuk tahap analisis berikutnya dengan model *Decision Tree* dan *Random Forest*.

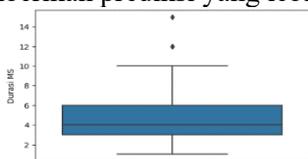
Jenis Kelamin	0
Fakultas	0
IPK	0
Jenis MS	0
Durasi MS	0
MS Belajar	0
Komunitas MS	0
Diskusi MS	0
Perspektif MS	0
dtype: int64	

Gambar 2. Data Missing Value

### 3.3 Pemeriksaan Outlier

Setelah melakukan pengecekan nilai yang hilang (*missing value*), langkah selanjutnya adalah mengecek *outlier*, seperti yang

ditunjukkan pada Gambar 3. Meskipun terdapat *outlier* dalam data, *outlier* tersebut tidak dihilangkan karena setiap mahasiswa memiliki durasi belajar yang berbeda-beda. Hal ini penting untuk mempertahankan keanekaragaman data dan memastikan model *Decision Tree* dan *Random Forest* dapat menangkap variasi dalam pola penggunaan media sosial dan durasi belajar mahasiswa. Kedua model ini mampu menangani *outlier* dengan baik. *Decision Tree*, dengan pendekatan hierarkisnya, dan *Random Forest*, dengan sifat *ensemble*-nya yang menggabungkan hasil dari banyak pohon keputusan, membuat model lebih *robust* dan dapat memberikan prediksi yang lebih akurat.



Gambar 3. *Outlier* data Durasi MS

### 3.4 Transformasi Data

Setelah itu, dilakukan proses *encoding* pada *dataset*. *Encoding* dilakukan agar semua data non-numerik diubah menjadi format numerik, sehingga lebih mudah untuk diterapkan pada model pembelajaran mesin seperti *Decision Tree* dan *Random Forest*. Proses ini penting karena algoritma pembelajaran mesin membutuhkan input data dalam bentuk numerik untuk melakukan perhitungan dan analisis. Sebagai contoh, kategori seperti jenis kelamin, jenis media sosial yang digunakan, dan perspektif mahasiswa terhadap penggunaan media sosial diubah menjadi kode numerik. Hal ini ditunjukkan pada Gambar 4.

Proses *encoding* memastikan bahwa semua variabel yang digunakan dalam model memiliki format yang sesuai dan dapat diolah oleh algoritma. Dengan mengubah data kategorikal menjadi numerik, kita menghilangkan potensi masalah yang dapat muncul dari interpretasi data yang tidak konsisten dan mempermudah proses analisis lebih lanjut. *Encoding* juga membantu dalam meningkatkan efisiensi model *Decision Tree* dan *Random Forest* dalam memproses data dan menghasilkan prediksi yang akurat. Hasil akhir dari proses *encoding* ini adalah *dataset* yang siap untuk dianalisis dan diolah lebih lanjut menggunakan model *Decision Tree* dan *Random Forest*, memastikan bahwa setiap aspek data diambil ke dalam pertimbangan dalam analisis akhir.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 136 entries, 0 to 135
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Jenis Kelamin         136 non-null    int32
1   Fakultas              136 non-null    int32
2   IPK                   136 non-null    int32
3   Durasi MS             136 non-null    int64
4   MS Belajar           136 non-null    int32
5   Komunitas MS         136 non-null    int32
6   Diskusi MS           136 non-null    int32
7   Perspektif MS        136 non-null    int64
8   Jenis MS_Instagram   136 non-null    int32
9   Jenis MS_Tiktok      136 non-null    int32
10  Jenis MS_X (twitter) 136 non-null    int32
11  Jenis MS_Youtube     136 non-null    int32
dtypes: int32(10), int64(2)
memory usage: 7.6 KB
None
```

Gambar 4. Hasil *encoding* data

### 3.5 Splitting Data

Setelah proses *encoding* selesai, tahap selanjutnya adalah membagi *dataset* menjadi dua bagian: data pelatihan (*training set*) dan data pengujian (*testing set*). Proses ini dikenal sebagai *splitting*. *Splitting* dilakukan dengan rasio 0,7 untuk data pelatihan dan 0,3 untuk data pengujian. Hal ini memastikan bahwa model dapat dilatih dengan satu *subset* data dan diuji dengan *subset* data lainnya yang tidak pernah dilihat oleh model sebelumnya. Proses ini penting untuk mengukur kinerja model secara objektif dan memastikan bahwa model tidak *overfitting*.

### 3.6 Pelatihan Model

Keberhasilan penggunaan *Decision Tree* dan *Random Forest* dapat dilihat dari kinerja model terhadap data uji. Dalam penelitian ini, model *Decision Tree* dan *Random Forest* dikembangkan menggunakan *package sklearn*. Langkah pertama adalah menentukan parameter utama, seperti jumlah pohon (*n\_estimators*) yang akan digunakan dalam *Random Forest*, dan menetapkan nilai untuk *seed* generator acak dengan *random\_state = 42*.

Proses pelatihan ini memungkinkan model untuk mempelajari informasi yang cukup sehingga dapat melakukan prediksi yang akurat. Secara keseluruhan, baik model *Decision Tree* maupun *Random Forest* mampu membuat prediksi yang akurat dan memastikan hasil yang konsisten pada data uji, sehingga memberikan gambaran yang jelas tentang kinerja model dalam situasi yang nyata.

**3.7 Evaluasi Model**

Model yang telah dilatih dievaluasi menggunakan data pengujian. Metrik kinerja seperti akurasi, *precision*, *recall*, dan *F1-score* dihitung untuk menilai seberapa baik model dalam membuat prediksi. Evaluasi ini penting untuk memastikan bahwa model dapat bekerja dengan baik pada data baru.

**3.8 Perbandingan Performa Model**

Tahap ini melibatkan perbandingan performa model dengan menggunakan berbagai metrik kinerja untuk menentukan model terbaik. Model dengan performa terbaik pada data pengujian akan dipilih sebagai model akhir yang paling efektif untuk prediksi.

**HASIL DAN PEMBAHASAN**

**4.1 Dekripsi Data**

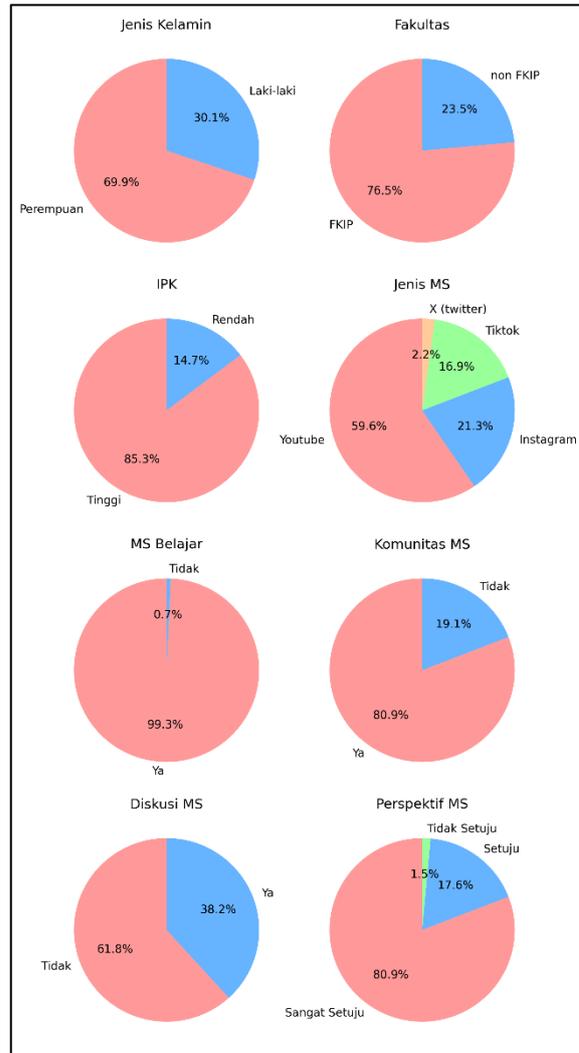
Berdasarkan data yang diperoleh dari 136 responden mahasiswa Universitas Sanata Dharma, deskripsi kualitatif menunjukkan beberapa temuan penting terkait penggunaan media sosial dalam aktivitas akademik. Mayoritas responden adalah perempuan (69.9%), dan sebagian besar berasal dari Fakultas Keguruan dan Ilmu Pendidikan (FKIP) dengan persentase 76.5%. Hal ini menunjukkan dominasi perempuan dalam populasi mahasiswa yang diteliti serta konsentrasi yang lebih tinggi dari fakultas FKIP dibandingkan fakultas lain.

Sebagian besar responden memiliki IPK tinggi ( $\geq 3.00$ ), yaitu sebanyak 85.3%, sementara hanya 14.7% responden yang memiliki IPK di bawah 3.00. Data ini menunjukkan bahwa mayoritas mahasiswa yang berpartisipasi dalam penelitian ini memiliki prestasi akademik yang baik.

Platform media sosial yang paling banyak digunakan untuk aktivitas pendidikan adalah YouTube (59.6%), diikuti oleh Instagram (21.3%), TikTok (16.9%), dan X (Twitter) (2.2%). YouTube terlihat menjadi platform utama bagi mahasiswa dalam mendukung aktivitas akademik mereka, diikuti oleh Instagram dan TikTok.

Hampir semua responden (99.3%) menggunakan media sosial untuk aktivitas belajar, dengan hanya 0.7% yang tidak menggunakannya. Ini menunjukkan tingginya adopsi media sosial sebagai alat belajar di kalangan mahasiswa. Selain itu, sebagian besar

responden (80.9%) ikut serta dalam komunitas media sosial yang berhubungan dengan akademik, sedangkan 19.1% tidak ikut serta. Hal ini mengindikasikan bahwa komunitas *online* berperan penting dalam mendukung aktivitas akademik mahasiswa.



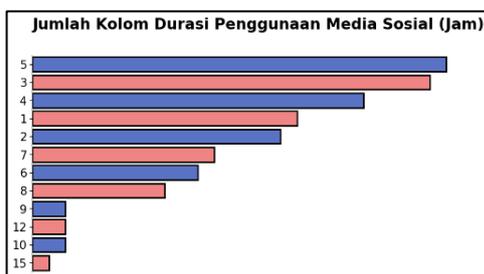
Gambar 5. Dikripsi Data

Meskipun banyak mahasiswa menggunakan media sosial untuk belajar, sebanyak 61.8% responden tidak terlibat dalam diskusi akademik di media sosial, sementara 38.2% terlibat dalam diskusi tersebut. Ini menunjukkan bahwa tidak semua mahasiswa aktif dalam diskusi akademik, meskipun mereka menggunakan media sosial sebagai alat belajar.

Mayoritas responden (80.9%) sangat setuju bahwa media sosial berpengaruh positif terhadap aktivitas akademik mereka, diikuti oleh yang setuju (17.6%), tidak setuju (1.5%), dan

tidak ada yang sangat tidak setuju. Pandangan mayoritas mahasiswa terhadap pengaruh positif media sosial menunjukkan bahwa *platform* ini efektif dalam mendukung produktivitas akademik mereka.

Dari hasil analisis deskriptif di atas, dapat disimpulkan bahwa media sosial memiliki peran signifikan dalam mendukung aktivitas akademik mahasiswa Universitas Sanata Dharma. YouTube adalah platform yang paling banyak digunakan, dan sebagian besar mahasiswa menggunakan media sosial untuk belajar dan berpartisipasi dalam komunitas akademik online. Namun, tidak semua mahasiswa terlibat aktif dalam diskusi akademik di media sosial.



Gambar 6. Durasi Penggunaan Media Sosial

Secara keseluruhan, Gambar 6 ini menunjukkan variasi yang cukup besar dalam durasi penggunaan media sosial oleh mahasiswa, dengan mayoritas menggunakan media sosial selama 3 hingga 5 jam per hari. Durasi penggunaan yang ekstrem, seperti 9 jam atau lebih, cenderung lebih jarang terjadi. Data ini penting untuk memahami kebiasaan penggunaan media sosial di kalangan mahasiswa dan bagaimana hal ini dapat mempengaruhi produktivitas akademik mereka.

#### 4.2 Pemodelan dengan *DecisionTree* dan *Random Forest*

Penelitian ini menerapkan beberapa teknik untuk meningkatkan performa model pembelajaran mesin dalam menangani data yang tidak seimbang dan menghasilkan prediksi yang lebih akurat. Teknik-teknik ini diimplementasikan dalam bentuk *pipeline* yang efisien, yang menggabungkan beberapa langkah pemrosesan data dan pelatihan model.

Langkah awal dalam *pipeline* adalah menggunakan *RobustScaler*, sebuah teknik penskalaan yang lebih tahan terhadap *outlier* dibandingkan metode penskalaan lainnya seperti *StandardScaler*. Selanjutnya, untuk menangani masalah ketidakseimbangan data, *pipeline*

menerapkan SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE berfungsi untuk menyeimbangkan jumlah sampel dari setiap kelas dengan cara membuat sampel sintetis dari kelas minoritas. Ini membantu model untuk tidak bias terhadap kelas mayoritas dan meningkatkan akurasi prediksi untuk kelas minoritas.

Model pembelajaran mesin yang digunakan dalam pipeline ini adalah *DecisionTreeClassifier* dan *RandomForestClassifier*. *Decision Tree* adalah model yang sederhana namun kuat dalam mengklasifikasikan data berdasarkan aturan-aturan keputusan yang dihasilkan dari data pelatihan. *Random Forest*, di sisi lain, adalah model yang lebih kompleks yang menggabungkan hasil dari banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi *overfitting*.

Untuk memastikan bahwa model bekerja dengan parameter terbaik, *GridSearchCV* digunakan. *GridSearchCV* mencoba berbagai kombinasi parameter *hyperparameter* dan memilih yang terbaik berdasarkan kinerja pada data validasi. Untuk *DecisionTreeClassifier*, parameter yang dicari termasuk kedalaman maksimum pohon (*max\_depth*) dan kriteria pemisahan (*criterion*). Untuk *RandomForestClassifier*, parameter yang dicari termasuk jumlah pohon dalam hutan (*n\_estimators*), jumlah fitur yang dipertimbangkan untuk pemisahan di setiap simpul (*max\_features*), kedalaman maksimum pohon (*max\_depth*), dan kriteria pemisahan (*criterion*).

Setelah proses pencarian parameter terbaik selesai, model dievaluasi menggunakan data uji. Fungsi *evaluasi\_model* digunakan untuk menghitung dan menampilkan metrik performa seperti akurasi, *presisi*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan performa model dalam mengklasifikasikan data ke dalam kelas yang tepat. Selain itu, laporan klasifikasi juga memberikan gambaran detail tentang performa model pada setiap kelas, baik untuk IPK Rendah maupun IPK Tinggi.

	precision	recall	f1-score	support
IPK Rendah	0.40	0.67	0.50	6
IPK Tinggi	0.94	0.83	0.88	35
accuracy			0.80	41
macro avg	0.67	0.75	0.69	41
weighted avg	0.86	0.80	0.82	41

Gambar 7. Akurasi dengan *Decision Tree*

Berdasarkan hasil *Decision Tree*, Gambar 7, model memiliki akurasi keseluruhan sebesar 80%. *Presisi* untuk kelas IPK Rendah adalah

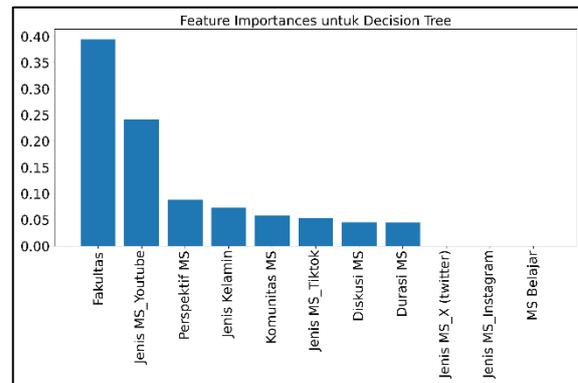
40%, dengan *recall* sebesar 67%, menghasilkan *F1-score* sebesar 50%. Hal ini menunjukkan bahwa model cukup baik dalam mendeteksi IPK Rendah, tetapi memiliki banyak *false positives*. Untuk kelas IPK Tinggi, model menunjukkan *presisi* yang sangat baik sebesar 94%, *recall* sebesar 83%, dan *F1-score* sebesar 88%, menunjukkan bahwa model ini sangat efektif dalam mengklasifikasikan sampel dengan IPK Tinggi. Nilai *macro average* dari *presisi*, *recall*, dan *F1-score* masing-masing adalah 67%, 75%, dan 69%, menunjukkan keseimbangan performa antara kedua kelas. Namun, nilai *weighted average* menunjukkan bahwa model lebih condong memberikan performa lebih tinggi pada kelas dengan jumlah sampel yang lebih banyak, yakni IPK Tinggi, dengan nilai *presisi* 86%, *recall* 80%, dan *F1-score* 82%. Hal ini menunjukkan bahwa *Decision Tree* lebih akurat dalam mengklasifikasikan kelas mayoritas (IPK Tinggi) dibandingkan dengan kelas minoritas (IPK Rendah).

	precision	recall	f1-score	support
IPK Rendah	0.67	0.67	0.67	6
IPK Tinggi	0.94	0.94	0.94	35
accuracy			0.90	41
macro avg	0.80	0.80	0.80	41
weighted avg	0.90	0.90	0.90	41

Gambar 8. Akurasi dengan *Random Forest*

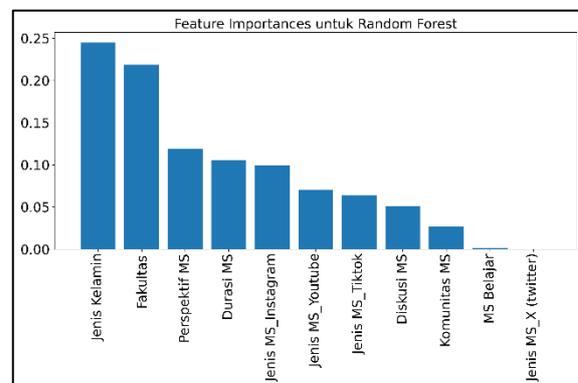
Sementara itu, berdasarkan hasil *Random Forest*, Gambar 8, model memiliki akurasi keseluruhan sebesar 90%. *Presisi* untuk kelas IPK Rendah adalah 67%, dengan *recall* sebesar 67%, menghasilkan *F1-score* sebesar 67%. Ini menunjukkan bahwa model memiliki keseimbangan yang baik dalam mendeteksi IPK Rendah dengan sedikit *false positives* dan *false negatives*. Untuk kelas IPK Tinggi, model menunjukkan performa yang sangat baik dengan *presisi* sebesar 94%, *recall* sebesar 94%, dan *F1-score* sebesar 94%, menandakan kemampuan yang sangat baik dalam mengklasifikasikan sampel dengan IPK Tinggi. Nilai *macro average* dari *presisi*, *recall*, dan *F1-score* masing-masing adalah 80%, menunjukkan bahwa model memiliki performa yang seimbang antara kedua kelas. Nilai *weighted average* menunjukkan bahwa model memberikan performa yang konsisten dan tinggi pada kelas mayoritas (IPK Tinggi) dengan *presisi*, *recall*, dan *F1-score* masing-masing sebesar 90%. Secara

keseluruhan, *Random Forest* menunjukkan performa yang unggul dan konsisten, mampu menangani ketidakseimbangan data dengan baik dan memberikan hasil yang lebih akurat dan andal dibandingkan dengan *Decision Tree*.



Gambar 9. *Feature Importance Decision Tree*

Selain evaluasi kinerja, analisis fitur penting dalam kedua model juga memberikan wawasan berharga. Berdasarkan Gambar 9, grafik *feature importance* untuk *Decision Tree*, terlihat bahwa fitur 'Fakultas' memiliki pengaruh terbesar dalam model ini, diikuti oleh 'Jenis MS\_Tiktok', 'Jenis Kelamin', dan 'Durasi MS'. Hal ini menunjukkan bahwa faktor-faktor ini sangat penting dalam menentukan klasifikasi IPK mahasiswa. Di sisi lain, berdasarkan Gambar 10, model *Random Forest* juga menunjukkan bahwa fitur 'Fakultas' dan 'Jenis Kelamin' memiliki pengaruh terbesar, diikuti oleh 'Jenis MS\_Youtube' dan 'Durasi MS'. Meskipun ada perbedaan dalam urutan kepentingan fitur, kedua model sepakat bahwa 'Fakultas' adalah faktor yang paling signifikan.



Gambar 10. *Feature Importance Random Forest*

Secara keseluruhan, implementasi teknik-teknik tersebut berhasil meningkatkan performa model dalam mengklasifikasikan data dengan akurasi dan keseimbangan yang baik antara kedua kelas. *Random Forest* terbukti lebih efektif dibandingkan *Decision Tree*, terutama dalam menangani ketidakseimbangan data dan memberikan prediksi yang lebih akurat untuk kedua kelas. Analisis *feature importance* juga memberikan pemahaman lebih dalam tentang faktor-faktor yang mempengaruhi prediksi model, yang dapat digunakan untuk pengembangan kebijakan atau intervensi yang lebih tepat sasaran di masa mendatang. Implementasi yang cermat dari teknik-teknik ini dapat digunakan sebagai pedoman untuk penelitian lebih lanjut dan aplikasi praktis lainnya dalam bidang pembelajaran mesin dan analisis data.

## SIMPULAN

Penelitian ini menunjukkan bahwa penerapan beberapa teknik dalam *pipeline* pemrosesan data dan pemodelan mampu meningkatkan performa model pembelajaran mesin dalam menangani data yang tidak seimbang. Teknik-teknik yang digunakan meliputi penskalaan data dengan *RobustScaler* dan *oversampling* kelas minoritas menggunakan SMOTE. Dua model pembelajaran mesin yang diuji, yaitu *DecisionTreeClassifier* dan *RandomForestClassifier*, dievaluasi menggunakan *Grid SearchCV* untuk menemukan parameter hyperparameter terbaik.

Hasil penelitian menunjukkan bahwa model *Random Forest* memiliki performa yang lebih baik dibandingkan dengan *Decision Tree* dalam semua metrik evaluasi, termasuk akurasi, *presisi*, *recall*, dan *F1-score*. *Random Forest* mampu menangani ketidakseimbangan data dengan lebih baik, menghasilkan prediksi yang lebih konsisten dan akurat untuk kedua kelas.

Analisis *feature importance* juga mengungkap bahwa fitur 'Fakultas' dan 'Jenis Kelamin' merupakan faktor yang paling signifikan dalam memprediksi IPK mahasiswa. Meskipun ada perbedaan dalam urutan kepentingan fitur antara *Decision Tree* dan *Random Forest*, kedua model sepakat bahwa 'Fakultas' adalah faktor yang paling signifikan.

Dengan demikian, *Random Forest* dengan teknik-teknik yang diterapkan dalam penelitian ini terbukti efektif dalam meningkatkan akurasi dan keandalan prediksi, yang dapat digunakan untuk mendukung pengambilan keputusan di

bidang pendidikan. Implementasi yang cermat dari teknik-teknik ini dapat dijadikan pedoman untuk penelitian lebih lanjut dan aplikasi praktis lainnya dalam analisis data dan pembelajaran mesin.

Untuk penelitian selanjutnya, beberapa arah penelitian dapat dilakukan. Pertama, eksplorasi fitur tambahan, seperti data perilaku mahasiswa atau aktivitas non-akademik, dapat memberikan wawasan lebih komprehensif tentang faktor yang memengaruhi produktivitas akademik. Kedua, model pembelajaran mesin lain, seperti XGBoost atau LightGBM, dapat dibandingkan untuk mengevaluasi apakah ada algoritma yang lebih unggul dalam memprediksi IPK mahasiswa. Ketiga, penelitian dapat diperluas ke populasi mahasiswa dari universitas lain atau pada tingkat pendidikan yang berbeda untuk meningkatkan generalisasi hasil.

## DAFTAR PUSTAKA

- Apriansyah, & Antoni, D. (2018). Pemanfaatan Media Sosial Sebagai Media Pembelajaran Pada Mahasiswa Perguruan Tinggi di Sumsel. *Jurnal Digital Teknologi Informasi*, 1(2), 64–79. <https://doi.org/10.32502/digital.v1i2.2371>
- Berry, S. (2019). Teaching to connect: Community-building strategies for the virtual classroom. *Online Learning Journal*, 23(1), 164–183. <https://doi.org/10.24059/olj.v23i1.1425>
- Budiyanto, F. I. K., Hermadi, I., & Hardhienata, M. K. D. (2024). Prediksi Performa Akademik Mahasiswa untuk Kelulusan Predikat Cum Laude dengan Pendekatan Machine Learning Predicting Academic Performance of Students for Graduating with Cum Laude Honors using Machine Learning Approach. *Jurnal Ilmu Komputer dan Agri-informatika*, 11(1), 39–49. <https://doi.org/10.29244/jika.11.1.39-49>
- Diantika, S., Nalatissifa, H., Maulidah, N., Supriyadi, R., & Fauzi, A. (2024). Penerapan Teknik Random Oversampling Untuk Memprediksi Ketepatan Waktu Lulus Menggunakan Algoritma Random Forest. *Computer Science*, 4(1), 11–18. <https://doi.org/10.31294/coscience.v4i1.1996>
- Fadli, Justicia, A. K., & Azanita, A. (2021). Pengaruh Adiksi Gawai Terhadap Prestasi Belajar Mahasiswa Prodi Diploma III

- Akademi Farmasi Yarsi Pontianak. *Jurnal Inkofar*, 5(1), 2581–2920. <https://doi.org/10.46846/JURNALINKOFAR.V5I1.190>
- Frazier, A., Silva, J., Meilak, R., Sahoo, I., Broda, M., & Chan, D. (2023). Decision Tree-Based Predictive Models for Academic Achievement Using College Students' Support Networks. *Journal of Data Science*, 21(3), 557–577. <https://doi.org/10.6339/21-JDS1033>
- Hariyanti, I., Al-Husaini, M., & Raharja, A. R. (2024). Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa. *Technologia : Jurnal Ilmiah*, 15(2), 332–240. <https://doi.org/10.31602/TJI.V15I2.14381>
- Khalim, K. A., Hayati, U., & Bahtiar, A. (2023). Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest Dan Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, 7(1), 498–504. <https://doi.org/10.36040/jati.v7i1.6376>
- Marna, M., Maxrizal, M., & Saftari, M. (2020). Analisis Faktor yang Mempengaruhi Indeks Prestasi dengan Metode Regresi Logistik Biner. *JMPM: Jurnal Matematika dan Pendidikan Matematika*, 5(1), 12–22. <https://doi.org/10.26594/jmpm.v5i1.1807>
- Mendoza, M. D., Hutajulu, O. Y., Lubis, A. R., Rahmadani, R., & Astono Putri, T. T. (2022). Pengaruh Penggunaan Media Sosial Dalam Pendidikan Terhadap Prestasi Akademik Mahasiswa. *Jurnal Teknologi Pendidikan (JTP)*, 15(2), 68–80. <https://doi.org/10.24114/jtp.v15i2.39120>
- Nasiruddin, F. Az. zahra, & Rapa', L. G. (2022). Dampak Media Sosial Terhadap Prestasi Belajar Mahasiswa. *Edustudent: Jurnal Ilmiah Pendidikan dan Pembelajaran*, 1(3), 188–193. <https://doi.org/10.26858/edustudent.v1i3.2890>
- Ogbuagu, B. C. U. (2021). Social Media Addiction and Academic Productivity Amongst Nigerian Students: Implications. *International Journal of Knowledge, Innovation and Entrepreneurship*, 9(3), 21–40.
- Pribadi, A. W., & Arsyad, P. J. (2021). Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *JSI (Jurnal sistem Informasi) Universitas Suryadarma*, 8(2), 203–210. <https://doi.org/10.35968/JSI.V8I2.735>
- Rizki, N. A., Afifah, N., Barung, T. G., Novri, I., & Hasan, I. K. (2023). Analisis Komparasi Perbedaan IPK Mahasiswa Program Studi Pendidikan Matematika Ditinjau Dari Gaya Belajar. *Prosiding Seminar Nasional Pendidikan Matematika, Universitas Mulawarman*, 3, 1–9.
- Saleh, J., Adrian, A., & Sanger, J. (2022). Sistem Klasifikasi Kelulusan Mahasiswa dengan Algoritma Random Forest. *Jurnal Ilmiah Realtech*, 18(1), 10–14. <https://doi.org/10.52159/realtech.v18i1.4>
- Satria, F., & Hermanto. (2022). Implementasi Algoritma Naive Bayes Untuk Memprediksi Pengaruh Media Sosial Terhadap Semangat Belajar Mahasiswa di Masa Pandemi Covid 19. *Jurnal Informasi dan Komputer*, 10(1), 50–56. <https://doi.org/10.35959/JIK.V10I1.290>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 1–27. <https://doi.org/10.1177/1536867X20909688>
- Septiani, Y., & Triariani, M. (2022). Pengaruh Burnout Terhadap Produktivitas Mahasiswa PGSD Universitas Kuningan. *Jurnal Kiprah Pendidikan*, 1(3), 161–167. <https://doi.org/10.33578/kpd.v1i3.46>
- Siregar, L. Y., & Nasution, M. I. P. (2020). Perkembangan teknologi informasi terhadap peningkatan bisnis online. *Hirarki: Jurnal Ilmiah Manajemen Dan Bisnis*, 2(1), 71–75. <https://doi.org/10.30606/hjimb>
- Wiriany, D., Natasha, S., & Kurniawan, R. (2022). Perkembangan Teknologi Informasi dan Komunikasi terhadap Perubahan Sistem Komunikasi Indonesia. *Jurnal Nomosleca*, 8(2), 242–252. <https://doi.org/10.26905/nomosleca.v8i2.8821>