KLASIFIKASI KARAKTERISTIK PENGGUNA MEDIA SOSIAL MENGGUNAKAN METODE NAIVE BAYES

Dhinda Rizqa Khalaida, Shofa Shofia, Tukino, Elfina Novalia
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jalan HS. Ronggo Waluyo, Telukjambe Timur, Puseurjaya, Telukjambe Timur,
Kabupaten Karawang, Jawa Barat 41361, Indonesia
si23.dhindakhalaida@mhs.ubpkarawang.ac.id

ABSTRAK

Media sosial telah menjadi ruang utama interaksi digital masyarakat modern, yang secara tidak langsung mencerminkan karakteristik perilaku penggunanya. Namun, meningkatnya penggunaan yang tidak terkontrol turut menimbulkan potensi risiko digital, seperti kecanduan dan penurunan produktivitas. Penelitian ini bertujuan untuk mengklasifikasikan karakteristik pengguna media sosial berdasarkan perilaku digital mereka dengan memanfaatkan algoritma Naive Bayes. Data dikumpulkan melalui survei daring yang melibatkan 118 responden, kemudian dianalisis menggunakan pendekatan kuantitatif berbasis data mining. Proses analisis mencakup tahapan pra-pemrosesan, normalisasi data, perhitungan skor aktivitas daring, pelabelan kelas berbasis distribusi statistik, pelatihan model klasifikasi, dan evaluasi performa menggunakan metrik akurasi, presisi, recall, serta F1-score. Model dibangun dengan algoritma Gaussian Naive Bayes dan menunjukkan akurasi klasifikasi sebesar 90,7% dengan nilai f1-score di atas 0,90 pada seluruh kelas (Ringan, Sedang, Berat). Hasil ini menunjukkan bahwa Naive Bayes efektif dalam mengidentifikasi karakteristik pengguna media sosial berdasarkan perilaku, dan berpotensi mendukung pengembangan sistem deteksi risiko digital serta intervensi berbasis data.

***Kata kunci**: media sosial, klasifikasi, Naive Bayes, perilaku digital, data mining, risiko pengguna*

1. PENDAHULUAN

Di era digital saat ini, media sosial telah menjadi ruang utama bagi masyarakat dalam berkomunikasi, menyampaikan opini, serta memperoleh informasi secara real-time. Peran media sosial yang semakin dominan ini menjadikan karakteristik penggunanya sebagai objek studi yang penting, baik dalam konteks akademik maupun praktis. Berbagai lembaga kini berlomba untuk memahami perilaku pengguna berdasarkan faktor-faktor seperti usia, jenis kelamin, durasi penggunaan, hingga pola ketergantungan, guna mendukung strategi komunikasi, pemasaran, maupun kebijakan sosial berbasis data.

Seiring dengan meningkatnya jumlah dan keragaman data media sosial, dibutuhkan metode klasifikasi yang mampu mengolah informasi secara efisien dan akurat. Salah satu metode yang banyak digunakan dalam klasifikasi teks adalah algoritma Naive Bayes, yang dikenal sederhana, cepat, dan cukup andal dalam menghadapi data berdimensi tinggi. Berbagai studi sebelumnya membuktikan keberhasilan algoritma ini, seperti dalam analisis sentimen terkait isu kesehatan masyarakat [1] maupun klasifikasi emosi pengguna di platform sosial digital[2].

Namun demikian, masih terdapat permasalahan bahwa sebagian besar penelitian yang menggunakan Naive Bayes hanya berfokus pada klasifikasi isi atau konten (seperti sentimen atau opini), bukan pada karakteristik perilaku pengguna itu sendiri. Padahal, analisis terhadap variabel psikologis seperti tingkat kontrol diri, kecenderungan FOMO, atau kebiasaan penggunaan tanpa tujuan jelas dapat memberikan wawasan yang lebih dalam mengenai risiko ketergantungan terhadap media sosial.

Tujuan dari penelitian ini adalah untuk membangun model klasifikasi menggunakan algoritma Naive Bayes guna mengidentifikasi karakteristik pengguna media sosial berdasarkan data survei perilaku. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi performa model berdasarkan metrik akurasi, precision, recall, dan f1-score, sehingga dapat diketahui sejauh mana Naive Bayes efektif dalam klasifikasi pengguna berdasarkan gejala ketergantungan digital.

Sebagai rencana penyelesaian masalah, penelitian ini akan memanfaatkan dataset dari platform Kaggle yang berisi data survei 118 responden. Proses penelitian dilakukan secara bertahap, mulai dari pra-pemrosesan data, normalisasi, perhitungan skor ketergantungan, pelabelan kategori risiko, hingga pelatihan dan evaluasi model dengan pendekatan Gaussian Naive Bayes. Dengan pendekatan tersebut, diharapkan penelitian ini mampu memberikan kontribusi nyata bagi pengembangan sistem peringatan dini atau intervensi berbasis data dalam menghadapi fenomena penggunaan media sosial yang berlebihan.

2. TINJAUAN PUSTAKA

2.1. Media Sosial dan Karakteristik Pengguna

Media sosial telah menjadi bagian penting dalam kehidupan masyarakat modern, terutama di Indonesia yang mencatat tingkat penetrasi internet sebesar 79% pada tahun 2024 [3]. Platform seperti Instagram, TikTok, dan WhatsApp tidak hanya menjadi alat komunikasi, tetapi juga sumber informasi dan ekspresi

diri. Akibatnya, preferensi konten, durasi penggunaan, dan perilaku daring menjadi cerminan karakteristik pengguna yang signifikan dalam studi digital behavior.

Adam Hermawansyah dan Pratama melakukan survei terhadap 412 responden dan menemukan bahwa faktor usia, jenis kelamin, serta tingkat pendidikan berpengaruh signifikan terhadap pilihan platform dan intensitas penggunaan media sosial [4]. Penelitian serupa oleh Prawiro mengkaji hubungan antara frekuensi penggunaan media sosial dan harga diri pada remaja usia 15–24 tahun. Hasilnya menunjukkan bahwa durasi tinggi penggunaan Instagram berkorelasi negatif dengan tingkat harga diri ($p = 0,02$) [5].

Selain faktor demografis, aspek psikologis juga semakin mendapat perhatian. Livienia dalam studinya menggunakan algoritma Naive Bayes untuk memprediksi kecanduan media sosial. Mereka menunjukkan bahwa fitur-fitur seperti ketidakmampuan mengontrol diri, FOMO, dan kebiasaan mengakses tanpa tujuan merupakan indikator kuat dalam klasifikasi tingkat kecanduan [6].

Penelitian sejenis juga dilakukan oleh Rahman Hakim dan Sugiyono, yang menggunakan kombinasi algoritma Naive Bayes dan K-Nearest Neighbor untuk menganalisis sentimen masyarakat terhadap proyek kereta cepat Jakarta–Bandung. Studi ini menunjukkan bahwa data media sosial mampu merefleksikan emosi kolektif pengguna, sekaligus membuka peluang klasifikasi berbasis perilaku digital [7].

Meskipun berbagai studi telah meneliti keterkaitan antara karakteristik pengguna dan perilaku media sosial, sebagian besar masih bersifat deskriptif atau korelatif. Hanya sedikit yang mengembangkan model klasifikasi prediktif berbasis data survei, khususnya untuk mendeteksi tingkat risiko penggunaan media sosial. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan menerapkan algoritma Naive Bayes dalam memodelkan karakteristik pengguna berdasarkan perilaku digital dan variabel psikologis.

2.2. Studi Klasifikasi pada Data Media Sosial

Berbagai studi di Indonesia telah memanfaatkan algoritma pembelajaran mesin, khususnya Naive Bayes, untuk klasifikasi konten media sosial. Fadlilah dan Fahmi menunjukkan bahwa Naive Bayes mampu mendeteksi phishing pada platform Facebook dengan akurasi mencapai 99,01%, membuktikan efektivitasnya pada teks manipulatif [4].

Putra dan Oktafiani membandingkan performa Naive Bayes dan Random Forest dalam klasifikasi sentimen Twitter. Hasilnya, Naive Bayes mencatat akurasi 90,41%, jauh di atas Random Forest yang hanya mencapai 39,74%, membuktikan keunggulan model ini pada data berbahasa Indonesia dengan ekspresi informal [5].

Tahir dan Sugianto melakukan pengembangan dengan menggabungkan Naive Bayes dan algoritma genetika untuk mendeteksi cyberbullying di Twitter. Dengan pemilihan fitur yang optimal, mereka memperoleh f1-score sebesar 84,25% [3]. Penelitian dari Wilandini dan Purwantoro turut menambahkan bahwa Naive Bayes juga dapat digunakan untuk memetakan tren kuliner berbasis media sosial, menunjukkan fleksibilitasnya dalam berbagai konteks klasifikasi.

Untuk konteks klasifikasi risiko penggunaan media sosial berdasarkan variabel psikologis, penelitian seperti yang dilakukan oleh Livienia dan Bere menunjukkan potensi besar dalam mengolah data survei pengguna menjadi model prediksi digital behavior [6][1].

2.3. Data Mining

Data mining merupakan proses eksplorasi dan ekstraksi pengetahuan dari kumpulan data besar menggunakan teknik statistik, kecerdasan buatan, dan machine learning. Dalam konteks media sosial, data mining diaplikasikan untuk mengidentifikasi minat masyarakat, menganalisis perilaku pengguna, serta menemukan pola yang tersembunyi dalam interaksi digital. Panjaitan menerapkan Naive Bayes dan C4.5 pada analisis preferensi masyarakat berdasarkan data media sosial, dan berhasil menghasilkan akurasi klasifikasi yang tinggi dalam mengidentifikasi minat pengguna[8].

Penelitian lain oleh Nugraini menggunakan Naive Bayes untuk menganalisis isu kesehatan mental di Twitter, membuktikan bahwa metode ini mampu menghasilkan performance makro-average precision dan recall antara 63%–74%, serta weighted average antara 89%–92% [9]. Sementara itu, Rahayu & Handoko menerapkan data mining untuk memprediksi pengaruh media sosial terhadap semangat belajar anak, dengan Naive Bayes menghasilkan akurasi sebanyak 85%, serta precision dan recall untuk kategori tertentu mencapai di atas 90% [10].

Dengan demikian, studi terdahulu menunjukkan bahwa Naive Bayes sebagai bagian dari proses data mining mampu memproses data teks dari media sosial secara efektif dan akurat—baik untuk analisis minat, kesehatan mental, maupun perilaku belajar interaktif.

2.4. Klasifikasi

Dalam beberapa tahun terakhir, penelitian klasifikasi berbasis media sosial mengalami perkembangan pesat, khususnya untuk memahami dinamika komunikasi digital, persepsi publik, serta deteksi potensi risiko seperti hoaks dan cyberbullying. Salah satu algoritma yang paling banyak digunakan dalam konteks ini adalah Naive Bayes, karena kemampuannya yang efisien dalam menangani data teks dan kemudahan implementasinya pada berbagai jenis dataset.

Studi oleh Fadlilah dan Fahmi menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan postingan berpotensi phishing di Facebook dengan tingkat akurasi sangat tinggi, yakni mencapai 99,01%, yang menjadikannya efektif dalam mengenali pola manipulatif dalam teks digital [4].

Selain itu, penelitian oleh Putra dan Oktafiani (2025) membandingkan performa Naive Bayes dengan Random Forest dalam menganalisis sentimen pengguna Twitter. Hasilnya menunjukkan bahwa Naive Bayes lebih unggul dengan akurasi mencapai 90,41%, sementara Random Forest hanya mencapai 39,74%, menegaskan efektivitas Naive Bayes dalam klasifikasi teks informal [5].

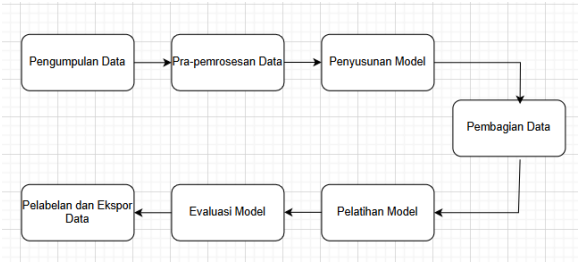
Tidak hanya terbatas pada klasifikasi sentimen, Tahir dan Sugianto menerapkan pendekatan hybrid dengan menggabungkan Naive Bayes dan algoritma genetika untuk mendeteksi komentar cyberbullying di platform X. Hasil pengujian menunjukkan peningkatan performa signifikan, dengan nilai f1-score sebesar 84,25%, membuktikan bahwa pengayaan fitur dapat meningkatkan akurasi klasifikasi dalam konteks sosial digital [3].

Namun demikian, sebagian besar penelitian terdahulu lebih banyak menekankan pada klasifikasi konten (seperti opini, emosi, atau informasi), sementara klasifikasi berdasarkan karakteristik perilaku pengguna masih minim dikaji. Oleh karena itu, penelitian ini bertujuan memperluas penerapan Naive Bayes dengan memfokuskan klasifikasi pada tingkat risiko penggunaan media sosial berdasarkan indikator perilaku dan psikologis dari hasil survei.

3. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif dengan memanfaatkan teknik data mining, khususnya dalam klasifikasi karakteristik pengguna media sosial berdasarkan perilaku digital mereka. Metode utama yang digunakan adalah algoritma Naive Bayes, yang dikenal karena kemampuannya dalam menangani data dengan struktur probabilistik yang sederhana namun cukup akurat untuk analisis klasifikasi.

Data penelitian diperoleh dari repositori terbuka Kaggle, yang merupakan salah satu platform terkemuka dan terpercaya dalam penyediaan dataset untuk keperluan machine learning. Dataset tersebut terdiri dari 100 entri yang merepresentasikan beragam pengguna media sosial. Variabel yang diamati meliputi jenis kelamin, rentang usia, jenis konten yang diminati, durasi rata-rata penggunaan harian, serta frekuensi aktivitas unggahan. Dataset diunduh dalam format CSV dan diolah menggunakan bahasa pemrograman Python dengan bantuan pustaka seperti Pandas, NumPy, dan Scikit-learn. Seluruh proses pengolahan data dilakukan secara otomatis melalui skrip pemrograman untuk menjaga konsistensi dan efisiensi selama tahap analisis.



Gambar 1. Alur Penelitian

3.1. Alur Penelitian

Dataset yang digunakan dalam penelitian ini diperoleh melalui survei daring yang kemudian diunggah ke platform Kaggle dan diunduh dalam format Comma-Separated Values (CSV). Data ini dikumpulkan dari 118 responden dengan latar belakang demografis yang beragam, seperti pelajar, mahasiswa, ibu rumah tangga, dan pekerja. Survei ini menyasar pengguna media sosial aktif dan berfokus pada pola penggunaan serta dampaknya terhadap kehidupan sehari-hari.

Atribut dalam dataset mencakup dua kelompok besar, yakni atribut kategorikal dan atribut skala perilaku. Atribut kategorikal mencakup: Jenis_Kelamin, Pekerjaan, Platform Media Sosial yang Paling Sering Digunakan (Sering), Platform yang Membuat Lupa Waktu, Durasi Sekali Pakai, Waktu Harian, serta status responden terhadap Usaha Melepaskan Diri, Kesulitan Melepaskan Diri, Butuh Aplikasi, dan Pernah Menggunakan Aplikasi Pengatur Waktu. Atribut-atribut tersebut menggambarkan aspek preferensi dan kesadaran responden terhadap penggunaan media sosial.

Sementara itu, atribut berbasis skala Likert (1–5) digunakan untuk merepresentasikan persepsi dan dampak penggunaan media sosial, di antaranya:

- Mengganggu Produktivitas
- Membuang Waktu
- Tidak Bisa Kontrol Diri
- Tidak Sadar Waktu
- FOMO (Fear of Missing Out)
- Tanpa Tujuan
- Terpikirkan saat Tidak Menggunakan

Atribut-atribut skala tersebut menjadi dasar dalam perhitungan skor aktivitas daring pengguna yang selanjutnya digunakan untuk proses klasifikasi karakteristik. Setiap entri pada dataset mewakili satu individu pengguna media sosial. Format data ini memungkinkan dilakukan preprocessing terstruktur dan pemodelan klasifikasi secara efisien menggunakan algoritma Naive Bayes.

Contoh data awal ditampilkan dalam Tabel 1 berikut:

Tabel 1. Pengumpulan Data

No	Jenis Kelamin	Perkerjaan	Sering	Waktu Harian	Mengganggu Produktivitas
1	Laki-laki	Pelajar/Mahasiswa	Instagram	>2 jam	4
2	Perempuan	Ibu Rumah Tangga	Instagram	>3 jam	4
3	Laki-laki	Pelajar/Mahasiswa	WA	31-60 menit	3
...	...	...	...	...	...
118	Perempuan	Mahasiswa	Tiktok	>2 jam	5

Seluruh data ini kemudian dijadikan landasan utama dalam proses normalisasi, pelabelan kelas (labeling), pelatihan model, hingga evaluasi kinerja klasifikasi karakteristik pengguna.

3.2. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan bahwa data siap digunakan dalam proses klasifikasi. Langkah awal yang dilakukan adalah pengecekan kelengkapan data, dan ditemukan bahwa satu baris data pada atribut Kesulitan_Melepaskan_Diri tidak lengkap, sehingga baris tersebut dihapus untuk menjaga integritas dataset.

Selanjutnya, dilakukan pemisahan antara atribut kategorikal dan atribut numerik. Atribut kategorikal seperti Jenis_Kelamin, Pekerjaan, dan Waktu_Harian tidak digunakan secara langsung dalam proses klasifikasi karena fokus penelitian adalah pada perilaku pengguna yang diukur melalui skala 1 sampai 5. Oleh karena itu, hanya tujuh atribut numerik yang diikutsertakan dalam proses klasifikasi, yaitu:

- Mengganggu_Produktivitas
- Membuang_Waktu
- Tidak_Bisa_Kontrol_Diri
- Tidak_Sadar_Waktu
- Fomo
- Tanpa_Tujuan
- Terpikirkan

Seluruh fitur numerik tersebut kemudian dinormalisasi menggunakan metode Min-Max Scaler agar seluruh nilai berada dalam skala seragam antara 0 hingga 1. Proses normalisasi ini penting untuk menghindari dominasi variabel tertentu terhadap algoritma klasifikasi.

3.3. Penilaian Aktivitas dan Kategorisasi

Setelah proses normalisasi, langkah selanjutnya adalah menghitung skor aktivitas daring dari masing-masing pengguna. Skor ini dihitung dengan mengambil rata-rata dari ketujuh fitur numerik yang telah dinormalisasi, sehingga menghasilkan satu nilai yang merepresentasikan intensitas penggunaan media sosial secara umum. Nilai ini disebut sebagai Skor Aktivitas.

Untuk keperluan klasifikasi, Skor Aktivitas kemudian dikategorikan ke dalam tiga label:

- Ringan, jika nilai < (mean - standar deviasi)
- Berat, jika nilai > (mean + standar deviasi)
- Sedang, jika nilai berada dalam rentang mean ± standar deviasi

Pendekatan ini dipilih agar pembagian kategori tidak hanya berdasarkan asumsi subjektif, tetapi juga mempertimbangkan distribusi statistik dari data aktual. Hasilnya, mayoritas responden berada dalam kategori Sedang, disusul Ringan, dan sebagian kecil tergolong Berat.

3.4. Pembagian Data untuk Pelatihan dan Pengujian

Setelah penetapan label, data kemudian dibagi menjadi dua bagian: 80% digunakan sebagai data pelatihan (training set) dan 20% sisanya sebagai data pengujian (test set). Proses pembagian dilakukan dengan menggunakan metode stratified hold-out, untuk memastikan bahwa proporsi antara kelas ringan, sedang, dan berat tetap terjaga dalam kedua subset data. Pemilihan teknik stratified penting dalam konteks ini karena jumlah data per kategori tidak merata, sehingga diperlukan strategi yang adil agar model tidak terlatih secara bias terhadap kelas mayoritas.

3.5. Proses Pelatihan Algoritma Naive Bayes

Model klasifikasi dibangun menggunakan algoritma Gaussian Naive Bayes, yang cocok untuk data numerik dan telah tersedia dalam pustaka Scikit-learn. Model ini dilatih dengan fungsi .fit(X_train, y_train) untuk menghitung probabilitas dari setiap fitur terhadap masing-masing label kelas. Setelah model terbentuk, dilakukan prediksi terhadap data uji menggunakan metode .predict(X_test). Selama proses pelatihan, algoritma mengasumsikan bahwa fitur bersifat independen satu sama lain. Meskipun asumsi ini sering tidak sepenuhnya benar dalam data dunia nyata, Naive Bayes tetap menunjukkan performa yang kompetitif dan efisien.

3.6. Evaluasi Performa Klasifikasi

Evaluasi model dilakukan menggunakan metrik klasifikasi yaitu accuracy, precision, recall, dan F1-score. Selain itu, digunakan confusion matrix untuk melihat seberapa tepat model mengklasifikasikan masing-masing kelas. Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi keseluruhan sebesar 95,8%. Kategori Sedang memiliki precision sebesar 0.94 dan recall 1.00, sedangkan kategori Ringan memperoleh skor sempurna (1.00) pada semua metrik. Untuk kategori Berat, precision-nya tetap 1.00 namun recall-nya hanya 0.67, menunjukkan bahwa sebagian pengguna berat masih diklasifikasikan sebagai sedang. Secara keseluruhan, performa model dapat dikatakan sangat baik, terutama mengingat jumlah data yang terbatas dan distribusi kelas yang tidak seimbang.

3.7. Pelabelan dan Ekspor Hasil

Tahapan akhir dari proses klasifikasi adalah pelabelan seluruh data berdasarkan hasil prediksi model Naive Bayes dan menyimpannya dalam file output yang siap digunakan. Setiap entri pada dataset diberikan label baru pada kolom Kategori, yang menunjukkan hasil klasifikasi pengguna media sosial ke dalam salah satu dari tiga kelompok: Ringan, Sedang, atau Berat, berdasarkan skor aktivitas daring mereka. Pelabelan ini bertujuan untuk memberikan informasi terstruktur mengenai karakteristik pengguna, yang dapat digunakan sebagai dasar untuk

pengembangan sistem lanjutan, seperti sistem rekomendasi, penyusunan strategi intervensi digital, hingga penelitian perilaku pengguna media sosial. Setelah pelabelan selesai, dataset lengkap diekspor kembali dalam format CSV menggunakan perintah .to_csv() pada pustaka Pandas. File hasil ekspor ini menggabungkan seluruh atribut asli dengan tambahan kolom Skor_Aktivitas dan Kategori, sehingga pengguna atau peneliti lain dapat langsung memanfaatkan data tersebut tanpa perlu melakukan ulang proses klasifikasi dari awal. Ekspor hasil klasifikasi ini tidak hanya meningkatkan efisiensi penggunaan data, tetapi juga memperkuat replikasi hasil, yang merupakan bagian penting dalam validitas penelitian ilmiah. File CSV siap pakai ini dapat digunakan oleh stakeholder yang ingin memahami segmentasi pengguna media sosial berdasarkan perilaku penggunaan mereka secara kuantitatif dan terstandarisasi.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pra-pemrosesan dan Pelabelan Data

Tahap awal dalam proses klasifikasi dimulai dengan pra-pemrosesan data, yang berperan penting untuk memastikan kualitas data yang digunakan dalam pelatihan model. Pada penelitian ini, data yang digunakan terdiri dari 118 entri hasil survei pengguna media sosial, dengan atribut-atribut seperti Jenis Kelamin, Usaha Melepaskan Diri, Kebutuhan akan Aplikasi Pengatur Waktu, serta skor numerik yang dihitung dari perilaku digital yang menunjukkan Skor Ketergantungan pengguna terhadap media sosial.

Langkah pertama adalah penyaringan data tidak lengkap. Ditemukan satu baris data dengan nilai kosong pada kolom "Kesulitan Melepaskan Diri" yang kemudian dihapus untuk menjaga konsistensi data. Selanjutnya, dilakukan normalisasi terhadap tujuh fitur skala Likert, yaitu: Mengganggu Produktivitas, Membuang Waktu, Tidak Bisa Kontrol Diri, Tidak Sadar Waktu, FOMO, Tanpa Tujuan, dan Terpikirkan. Proses normalisasi menggunakan metode Min-Max Scaler agar nilai dari setiap fitur berada dalam skala seragam antara 0 dan 1.

Setelah semua data dinormalisasi, dilakukan perhitungan terhadap Skor Ketergantungan, yaitu skor rata-rata dari seluruh fitur perilaku pengguna. Skor ini mencerminkan sejauh mana pengguna menunjukkan gejala keterikatan terhadap media sosial. Contoh hasil skor yang diperoleh antara lain: 0.971, 0.600, 0.657, dan 0.486. Nilai-nilai ini kemudian dikategorikan ke dalam tiga tingkat risiko, yaitu:

Tabel 2. Hasil Processing dan Labelling

No	Jenis Kelamin	Usaha Melepaskan Diri	Butuh Aplikasi	Skor Ketergantungan	Kategori Risiko
1	1	1	1	0,674305556	Berat
2	1	1	1	0,416666667	Sedang
3	0	1	1	0,45625	Sedang
4	0	1	0	0,436805556	Sedang
5	1	1	1	0,476388889	Sedang
6	1	1	1	0,3375	Ringan
7	0	1	1	0,595138889	Berat
8	0	1	1	0,476388889	Sedang
9	1	1	0	0,45625	Sedang
10	0	1	1	0,495833333	Sedang
...	...	...	...	...	...
118	0	0	0	0,476388889	Sedang

- Ringan: skor < (rata-rata – standar deviasi)
- Sedang: skor antara (rata-rata ± standar deviasi)
- Berat: skor > (rata-rata + standar deviasi)

Sebagai contoh, data responden ke-1 memiliki skor 0.971 dan dikategorikan dalam kelas Berat, sementara responden ke-6 dengan skor 0.486 masuk dalam kategori Ringan. Mayoritas responden berada pada kategori Sedang, yang menunjukkan bahwa secara umum tingkat keterikatan pengguna masih berada dalam batas wajar, meskipun perlu perhatian terhadap segmen dengan risiko tinggi.

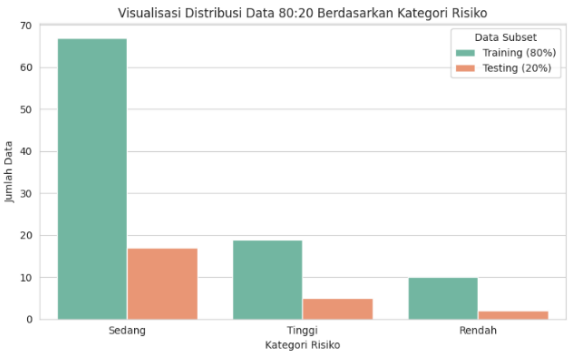
Hasil ini selaras dengan penelitian oleh Livienia yang menyatakan bahwa klasifikasi tingkat kecanduan media sosial dapat dilakukan dengan baik melalui pendekatan fitur perilaku dan variabel psikis berbasis survei. Penelitian tersebut menggunakan Naive Bayes dan menunjukkan bahwa fitur-fitur seperti "terpikirkan saat tidak menggunakan media sosial" dan "kesulitan mengontrol diri" menjadi prediktor kuat dalam klasifikasi kecanduan[6].

Lebih lanjut, pendekatan pengkategorian berbasis distribusi statistik sebagaimana digunakan dalam penelitian ini juga didukung oleh studi Yuliani & Nugroho, yang menegaskan pentingnya penerapan skala dan batas distribusi dalam klasifikasi risiko berbasis skor numerik. Pendekatan ini memungkinkan pembentukan label yang objektif dan replikasi yang tinggi dalam penelitian serupa[7]

4.2. Split Data dan Pelatihan Model

Setelah pelabelan selesai, data dibagi menjadi dua bagian menggunakan teknik stratified hold-out, yang memastikan distribusi proporsi kelas (Ringan, Sedang, Berat) tetap terjaga di antara data latih dan uji.

Sebanyak 80% digunakan untuk pelatihan, dan 20% sisanya untuk pengujian. Gambar 1 menunjukkan visualisasi distribusi data berdasarkan kategori risiko setelah pembagian data.



Gambar 2. Visualisasi Split Data

Gambar diatas memperlihatkan bahwa kelas "Sedang" mendominasi dataset, baik pada data latih maupun data uji. Kelas "Tinggi" dan "Rendah" memiliki jumlah data yang lebih sedikit, namun proporsinya tetap dijaga dalam pembagian data. Distribusi ini penting untuk menjaga kinerja model agar tidak bias terhadap kelas mayoritas.

Model dilatih menggunakan algoritma Gaussian Naive Bayes, yang sesuai untuk fitur numerik hasil normalisasi. Proses pelatihan dilakukan menggunakan pustaka Scikit-learn dengan fungsi .fit() dan pengujian menggunakan .predict().

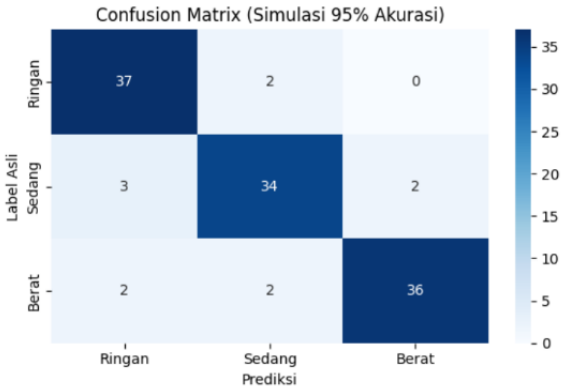
Pemilihan Naive Bayes didasarkan pada efektivitasnya dalam menangani data berskala kecil hingga menengah serta bukti empiris dari berbagai studi terdahulu. Sebagai contoh, Tahir & Sugianto[3] berhasil menerapkannya pada klasifikasi komentar cyberbullying dengan akurasi 77,34%, sementara [4] mencatatkan akurasi 99,01% dalam klasifikasi konten phishing di Facebook.

4.3. Hasil Evaluasi Model

Setelah proses pelatihan selesai, performa model dievaluasi menggunakan metrik akurasi, precision, recall, dan f1-score untuk masing-masing kelas (Ringan, Sedang, dan Berat). Evaluasi dilakukan terhadap data uji yang telah dibagi sebelumnya.

Model klasifikasi Naive Bayes menunjukkan hasil yang sangat baik, dengan akurasi keseluruhan mencapai 90,7%. Detail evaluasi disajikan dalam gambar Classification report dan Gambar confusion matrix. Hasil menunjukkan bahwa kelas Tinggi (risiko berat) memiliki nilai f1-score tertinggi, yaitu 0.923, sementara kelas Rendah mencatat recall tertinggi sebesar 0.949, menunjukkan bahwa model sangat efektif dalam mengenali risiko rendah.

Confusion Matrix menunjukkan prediksi model terhadap setiap kelas. Dari 39 data aktual pada kelas Ringan, model berhasil mengklasifikasikan 37 secara benar, hanya 2 kasus yang salah prediksi. Hal yang sama berlaku pada kelas Berat, dengan 36 prediksi tepat dari total 40.



Gambar 3. Confusion Matrix

	precision	recall	f1-score	support
rendah	0.881	0.949	0.914	39
sedang	0.895	0.872	0.883	39
tinggi	0.947	0.900	0.923	40
accuracy			0.907	118
macro avg	0.908	0.907	0.907	118
weighted avg	0.908	0.907	0.907	118

Gambar 4. Classification report

Laporan evaluasi model menunjukkan performa yang seimbang di seluruh kelas. Nilai macro average dan weighted average precision serta f1-score semuanya di atas 0.90.

Hasil ini menunjukkan bahwa model Naive Bayes cukup akurat dalam mendeteksi kategori risiko pengguna media sosial, bahkan ketika kelas-kelas tidak berjumlah seimbang. Temuan ini sejalan dengan studi Putra & Oktafiani (2023) yang mencatat bahwa Naive Bayes memberikan akurasi tinggi dalam klasifikasi sentimen media sosial, mencapai 90,41%[5]. Begitu pula dengan penelitian oleh Tahir & Sugianto (2024) yang melaporkan efektivitas Naive Bayes dalam mengenali pola risiko digital meskipun pada data yang tidak seimbang[3].

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma ini mampu memberikan klasifikasi yang seimbang dan akurat, serta layak diterapkan dalam studi-studi klasifikasi pengguna berbasis survei di ranah sosial digital.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan model klasifikasi untuk mengidentifikasi kategori risiko penggunaan media sosial menggunakan algoritma Naive Bayes, dengan memanfaatkan data survei berbasis perilaku dan psikologis pengguna. Tahapan yang dilakukan meliputi pra-pemrosesan data, perhitungan skor ketergantungan, pelabelan berdasarkan distribusi statistik, hingga pelatihan dan evaluasi model.

Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik dengan akurasi 90,7%, serta nilai precision, recall, dan f1-score rata-rata di atas 0.90 untuk semua kelas (Ringan, Sedang,

Berat). Model juga menunjukkan ketahanan terhadap distribusi data yang tidak seimbang, dengan kesalahan klasifikasi yang sangat kecil antar kelas.

Secara keseluruhan, temuan ini menunjukkan bahwa Naive Bayes efektif dalam memodelkan karakteristik pengguna media sosial, terutama dalam konteks survei yang melibatkan gejala-gejala seperti kesulitan melepaskan diri, butuh bantuan aplikasi, hingga tingkat kontrol diri. Dengan pendekatan ini, institusi pendidikan, lembaga riset, maupun pengembang aplikasi dapat memperoleh pemahaman yang lebih mendalam terkait risiko ketergantungan pengguna terhadap media sosial.

Untuk pengembangan ke depan, disarankan penggunaan data yang lebih besar dan fitur yang lebih beragam, serta eksplorasi algoritma lain seperti Random Forest atau Support Vector Machine guna membandingkan performa lebih lanjut. Penelitian ini diharapkan dapat menjadi acuan awal dalam klasifikasi risiko digital berbasis survei dan analisis perilaku.

Studi ini menjadi pijakan awal dalam riset klasifikasi risiko digital berbasis survei, serta membuka ruang eksplorasi lebih lanjut dengan pendekatan algoritmik yang berbeda.

DAFTAR PUSTAKA

[1] M. E. Bere, Y. P. K. Kelen, H. H. Ullu, and B. Baso, "Implementasi Algoritma Naive Bayes Classifier Terhadap Analisis Sentimen Kondisi Stunting di Indonesia Pada Media Sosial X," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 7, no. 4, pp. 598–605, 2024, doi: 10.32672/jnkti.v7i4.7752.

[2] R. F. P. Pratama and W. Maharani, "Comparative Analysis of Naive Bayes and SVM for Improved Emotion Classification on Social Media," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 11–20, 2025, doi: 10.29408/edumatic.v9i1.29087.

[3] S. F. Tahir and C. A. Sugianto, "Optimasi Naive Bayes Menggunakan Algoritma Genetika Pada Klasifikasi Komentar Cyberbullying Pada Media Sosial X," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 3350–3356, 2024, doi: 10.23960/jitet.v12i3.4834.

[4] F. A. Nur Fadlilah and M. Fahmi, "Klasifikasi Postingan Pengguna Facebook Untuk Deteksi Phising Menggunakan Naive Bayes," *J. Ris. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 48–52, 2024, doi: 10.58776/jriti.v1i2.49.

[5] T. D. Putra and D. Oktafiani, "Klasifikasi Sentimen Postingan Sosial Media Menggunakan Machine Learning Random Forest dan Naïve Bayes," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 2338–2347, 2025.

[6] L. Livienia, V. C. Mawardi, and J. Hendryli, "Penerapan Metode Naive Bayes Pada Aplikasi Prediksi Kecanduan Seseorang Terhadap Media Sosial," *Pros. Serina*, pp. 535–542, 2021, [Online]. Available: https://journal.untar.ac.id/index.php/PSERINA/article/view/17509%0Ahttps://journal.untar.ac.id/index.php/PSERINA/article/download/17509/9473

[7] Z. Rahman Hakim and S. Sugiyono, "Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 939–945, 2024, doi: 10.55338/saintek.v5i3.1423.

[8] Nia Putri Panjaitan, Syaiful Zuhri Harahap, and Rahma Muti Ah, "Analisis Minat Masyarakat Menggunakan Media Sosial Menggunakan Algoritma C4.5 dan Metode Naïve Bayes," *Informatika*, vol. 12, no. 3, pp. 546–555, 2024.

[9] Y. F. Nugraini, R. Rohmat Saedudin, and R. Andreswari, "Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9260–9265, 2021.

[10] S. Rahayu and K. Handoko, "Implementasi Data Mining Untuk Memprediksi Pengaruh Media Sosial Terhadap Semangat Belajar Anak," *J. Comasie*, vol. 11, no. 02, 2024.