

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	17
Tabel 4. 1 Dataset	40

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan media sosial saat ini, khususnya *Instagram*, telah menjadi salah satu sarana utama dalam membangun *brand awareness* bagi institusi pendidikan. Dengan konten visual yang menarik dan konsisten, *Instagram* mampu meningkatkan kesadaran merek serta memperkuat citra institusi, termasuk perguruan tinggi, di mata calon mahasiswa dan masyarakat umum. Konten visual yang dirancang secara kreatif tidak hanya berfungsi sebagai media informasi, tetapi juga sebagai alat untuk membentuk persepsi positif terhadap identitas kampus. (Budhi Setiawati & Ismail, 2025)

Strategi promosi melalui media sosial seperti *Instagram* telah menjadi tren yang dominan di era digital saat ini. Banyak perusahaan dan pemasar memanfaatkan platform ini sebagai media promosi karena kemampuannya menjangkau audiens secara luas dan interaktif. Salah satu pendekatan kreatif dalam memasarkan produk atau jasa agar menarik perhatian konsumen sekaligus membentuk *brand awareness* adalah dengan membuat dan membagikan konten yang relevan serta menarik di media sosial, khususnya melalui fitur-fitur unggulan *Instagram* seperti *Feed*, *Story*, *Reels*, dan *Live*. (Faishal Azka Kalkautsar & Endri Listiani, 2022)

Namun, salah satu tantangan yang dihadapi adalah bagaimana mengoptimalkan elemen visual agar efektif dalam meningkatkan *brand awareness*. Tidak semua konten yang menarik secara estetika mampu membentuk persepsi yang kuat di benak audiens, terutama dalam konteks institusi pendidikan. Meski media sosial telah umum digunakan sebagai sarana promosi, kajian ilmiah tentang efektivitas elemen visual terhadap *brand awareness* masih terbatas, khususnya di perguruan tinggi. Karena itu, dibutuhkan pendekatan berbasis data untuk memahami sejauh mana konten visual mendukung peningkatan kesadaran merek dan citra institusi. (Noviyanti et al., 2025)

Seiring dengan kebutuhan untuk memahami dampak konten visual secara lebih akurat, diperlukan metode analisis berbasis data yang mampu memetakan dan mengklasifikasikan persepsi audiens terhadap konten yang disajikan. Dalam penelitian ini, algoritma *Naïve Bayes* dipilih karena kemampuannya dalam membangun model klasifikasi yang sederhana namun efektif. Algoritma ini digunakan untuk mengolah persepsi audiens terhadap elemen visual yang dipublikasikan melalui akun *Instagram*, sehingga dapat memprediksi tingkat *brand awareness* secara sistematis.

Melalui penelitian ini, diharapkan dapat dikembangkan sebuah model analisis berbasis *Naïve Bayes* yang memberikan gambaran mengenai pengaruh konten visual *Instagram* terhadap *brand awareness* Kampus IIB Darmajaya. Hasil penelitian ini tidak hanya diharapkan memberikan kontribusi dalam pengembangan strategi media sosial kampus, tetapi juga memperkaya kajian akademik di bidang penerapan teknologi klasifikasi dalam pemasaran digital berbasis media sosial.

1.2 Rumusan Masalah

Adapun masalah yang diangkat dalam penelitian ini adalah:

1. Bagaimana pengaruh elemen-elemen visual pada konten *Instagram* Kampus IIB Darmajaya terhadap persepsi audiens?
2. Bagaimana tingkat *brand awareness* audiens terhadap Kampus IIB Darmajaya berdasarkan persepsi konten visual di *Instagram*?
3. Bagaimana membangun model klasifikasi tingkat *brand awareness* berdasarkan persepsi audiens terhadap konten visual menggunakan metode *Naïve Bayes*?

1.3 Batasan Masalah

Penelitian ini hanya membahas konten visual yang diunggah oleh akun *Instagram* resmi Kampus IIB Darmajaya @darmajayathebest dan tidak mencakup analisis terhadap elemen non-visual seperti *caption*, *hashtag*, komentar, atau iklan berbayar. Selain itu, metode klasifikasi yang digunakan dibatasi hanya pada algoritma *Naïve Bayes*, tanpa membandingkannya dengan algoritma lain. Waktu pengambilan data juga dibatasi pada periode tertentu yang telah ditentukan.

Penelitian ini tidak menilai efektivitas kampanye secara keseluruhan, tetapi hanya pada aspek visual yang mempengaruhi persepsi audiens terhadap *brand awareness*.

1.4 Ruang Lingkup

Penelitian ini berfokus pada analisis persepsi audiens terhadap konten visual yang mencakup *Feed*, *Story*, dan *Reels* dari akun *Instagram* resmi Kampus IIB Darmajaya. Cakupan visual yang dianalisis meliputi elemen warna, desain, dan konsistensi tema. Data persepsi diperoleh melalui kuesioner kepada follower aktif, yang terdiri dari mahasiswa, calon mahasiswa, serta masyarakat umum. Penelitian ini menggunakan algoritma *Naïve Bayes* dalam proses klasifikasi untuk menentukan pengaruh konten visual terhadap tingkat *brand awareness*.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui persepsi audiens terhadap elemen-elemen visual pada konten *Instagram* Kampus IIB Darmajaya.
2. Mengukur tingkat *brand awareness* audiens terhadap Kampus IIB Darmajaya berdasarkan konten visual yang ditampilkan di *Instagram*.
3. Membangun model klasifikasi tingkat *brand awareness* berdasarkan persepsi terhadap konten visual menggunakan metode *Naïve Bayes*.

1.6 Manfaat Penelitian

1.6.1 Manfaat Akademis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian akademik di bidang Teknik Informatika, khususnya dalam penerapan algoritma klasifikasi *Naïve Bayes* pada analisis data media sosial. Selain itu, hasil penelitian ini dapat menjadi referensi bagi penelitian selanjutnya yang mengkaji hubungan antara elemen visual dalam konten media sosial dan tingkat *brand awareness*.

1.6.2 Manfaat Praktis

Penelitian ini dapat memberikan manfaat praktis bagi pihak Kampus IIB Darmajaya dalam mengevaluasi efektivitas strategi komunikasi dan pemasaran melalui *Instagram*. Model klasifikasi yang dibangun melalui penelitian ini dapat menjadi dasar dalam merancang konten visual yang lebih terarah dan tepat sasaran untuk meningkatkan *brand awareness*.

1.6.3 Manfaat Teknologis

Penelitian ini menunjukkan penerapan teknologi *machine learning*, khususnya algoritma *Naïve Bayes*, dalam mengolah dan menganalisis data dari media sosial. Model yang dihasilkan dapat dijadikan acuan dalam pengembangan sistem pendukung keputusan (*decision support system*) untuk mendukung strategi pemasaran digital di lingkungan institusi pendidikan.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun secara sistematis untuk memudahkan pembaca dalam memahami isi dan alur penelitian. Berikut uraian singkat isi dari masing-masing bab:

BAB I PENDAHULUAN

Berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Memuat teori-teori yang relevan dan menjadi dasar dalam pelaksanaan penelitian ini.

BAB III METODOLOGI PENELITIAN

Menjelaskan metode penelitian yang digunakan, termasuk jenis penelitian, teknik pengumpulan data, populasi dan sampel, serta metode analisis data.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Berisi penyajian hasil penelitian dan pembahasan yang dikaitkan dengan teori yang relevan, serta analisis pengaruh konten visual

Instagram terhadap tingkat *brand awareness* Kampus IIB Darmajaya.

BAB V KESIMPULAN DAN SARAN

Memuat kesimpulan dari hasil penelitian serta saran-saran yang dapat diberikan untuk pengembangan penelitian selanjutnya maupun aplikasi praktis di lapangan.

BAB II

LANDASAN TEORI

2.1 Media Sosial

Media sosial merupakan *platform* berbasis teknologi yang memungkinkan individu atau komunitas untuk membuat, berbagi, dan bertukar informasi dalam komunitas virtual. Menurut (Kaplan & Haenlein, 2010), media sosial adalah “sekelompok aplikasi berbasis internet yang dibangun di atas fondasi ideologis dan teknologi Web 2.0, yang memungkinkan penciptaan dan pertukaran konten buatan pengguna.”

Dalam lingkungan pemasaran digital yang semakin bersaing, media sosial telah berkembang menjadi salah satu alat strategis yang paling mendasar. Platform ini mendukung terjalannya dan pemeliharaan hubungan yang lebih personal dan autentik antara merek dan konsumen. Tidak seperti media konvensional yang umumnya bersifat satu arah, media sosial memungkinkan komunikasi dua arah yang aktif, di mana merek bisa berinteraksi langsung dengan publik, menjawab pertanyaan, merespons umpan balik, dan juga menyelesaikan keluhan dengan cepat. Kemampuan untuk mendapatkan umpan balik langsung ini sangat penting, memungkinkan pemasar untuk menilai sentimen masyarakat, mendeteksi tren yang muncul, dan menyesuaikan strategi mereka dengan cepat dan fleksibel. Selain itu, interaksi yang terus-menerus ini secara substansial meningkatkan keterlibatan pengguna, mengubah konsumen yang pasif menjadi pendukung merek yang aktif melalui suka, komentar, pembagian, dan bentuk partisipasi lainnya.

Selain itu, kemampuan media sosial dalam menyebarkan konten secara luas dan cepat tidak adaandingannya. Sebuah pesan atau kampanye bisa menjadi viral dalam waktu singkat, menjangkau jutaan penonton di seluruh dunia dengan biaya yang lebih efisien dibandingkan dengan metode pemasaran konvensional. Efisiensi ini, bersama dengan kemampuan penargetan audiens yang sangat akurat, menjadikan media sosial cara yang sangat efektif untuk meningkatkan kesadaran merek. Dengan kehadiran yang kontinu dan konten yang menarik serta relevan, merek dapat meningkatkan eksposur mereka, memperkuat ingatan merek dalam

pikiran konsumen, dan secara bertahap menempatkan diri sebagai pilihan utama dalam kategori produk atau layanan yang mereka tawarkan. Analitik data yang melimpah dari platform media sosial memberikan wawasan berharga tentang demografi pemirsa, pilihan konten, dan efektivitas kampanye, memungkinkan pengambilan keputusan berbasis data yang lebih cerdas dan pengoptimalan strategi pemasaran yang berkelanjutan.

2.2 Instagram sebagai Platform Promosi Visual

Instagram sebuah cara baru yang hebat untuk berbagi foto, pertama kali muncul pada Oktober 2010. Tidak lama kemudian, tempat ini berhasil menarik perhatian banyak orang di seluruh dunia, menjadikannya salah satu media sosial terbesar dengan pengguna yang sangat banyak, lebih dari 2 miliar orang pada tahun 2023 (Datareportal, 2023). Orang-orang di seluruh dunia suka *Instagram* karena bisa menyampaikan pesan dan cerita lewat gambar dan video yang menarik. Tempat ini punya banyak fitur yang mudah dipakai, seperti Feed untuk memilih konten yang bagus dan tahan lama, Stories untuk berbagi hal-hal yang asli dan sementara dengan tambahan fitur menarik seperti survei dan kuis, Reels yang menampilkan video pendek yang seru dan menghibur dengan memakai musik yang sedang populer dan efek visual, dan *Live* yang memungkinkan berbicara langsung dengan penonton saat itu juga. Semua fitur ini bisa dipakai dengan baik untuk membuat rencana promosi visual yang baru dan menarik, membuat merek atau organisasi bisa menunjukkan diri dengan cara yang beda dan menarik perhatian.

Instagram sangat penting dalam membuat merek dikenal, karena punya tiga hal utama yaitu gambar yang kuat, kemampuan menjangkau banyak orang yang berbeda-beda, dan fitur menarik yang bisa membuat pengguna lebih aktif. Gambar yang bagus memberi kesempatan bagi merek untuk membuat tampilan yang sama dan menarik, menyampaikan perasaan, cerita, dan nilai-nilai perusahaan dengan lebih baik daripada hanya memakai tulisan. Sebuah gambar atau video yang bagus bisa menyampaikan banyak hal dan tidak terhalang oleh bahasa. Dengan cara kerja yang pintar dan fitur seperti tagar dan halaman Jelajahi, *Instagram* memberi kesempatan besar untuk menjangkau banyak orang yang sesuai, termasuk target

pasar tertentu melalui iklan yang tepat. Fitur menarik seperti komentar, suka, berbagi, pesan langsung, dan stiker interaktif tidak hanya mempermudah berbicara dua arah tetapi juga membuat komunitas yang setia dan aktif di sekitar merek. Keterlibatan ini sangat berguna karena bisa mengubah pengikut menjadi orang yang menyebarkan merek dengan sukarela.

Dalam dunia pendidikan, seperti yang dikatakan oleh (Appel et al., 2020), *Instagram* sangat penting sebagai alat untuk mencapai berbagai tujuan sekolah atau universitas.

1. Pertama, tempat ini sangat baik dalam membangun dan memperkuat citra kampus dengan menunjukkan kehidupan kampus yang menarik, fasilitas yang bagus, berbagai kegiatan mahasiswa, dan suasana belajar yang memotivasi. Tampilan ini menciptakan kesan yang baik dan membuat calon mahasiswa ingin masuk.
2. Kedua, *Instagram* adalah cara penting untuk menarik perhatian calon mahasiswa baru. Dengan menunjukkan cerita sukses alumni, pendapat dari mahasiswa saat ini, hal baru dalam penelitian, prestasi akademik dan non-akademik, serta budaya kampus yang terbuka, sekolah atau universitas bisa menunjukkan kelebihan mereka, membantu calon mahasiswa membayangkan masa depan mereka di sana.
3. Ketiga, *Instagram* juga penting dalam membentuk dan menjaga pandangan positif masyarakat terhadap sekolah atau universitas. Dengan unggahan rutin tentang kegiatan pelayanan masyarakat, prestasi guru, kerja sama yang baik, dan kontribusi terhadap perkembangan ilmu pengetahuan, sekolah atau universitas bisa membangun kepercayaan di mata masyarakat. Dengan kata lain, *Instagram* seperti jendela yang terbuka, memberikan kesempatan bagi sekolah atau universitas untuk menunjukkan identitas, nilai, dan keunggulan mereka kepada banyak orang.

2.3 Konten Visual

Konten visual itu adalah semua informasi yang kita sampaikan lewat mata kita. Lebih dari sekadar teks, ia merujuk pada segala bentuk data yang disajikan melalui elemen-elemen yang bisa dilihat langsung, mulai dari gambar statis seperti foto atau grafik, desain yang menarik, hingga video yang bergerak dan penuh cerita. Tujuannya sangat jelas: yaitu untuk secara langsung dan efektif memberikan pesan atau informasi kepada siapa pun yang melihatnya, seringkali tanpa perlu banyak kata. (Agustina et al., 2019)

Di zaman sekarang, di mana arus informasi mengalir begitu deras dan semua serba cepat, perhatian kita seringkali terpecah-pecah. Dalam konteks komunikasi modern yang padat ini, konten visual punya peran yang sangat penting, bahkan bisa dibilang krusial. Mengapa demikian? Karena dia punya kemampuan istimewa untuk langsung 'menangkap' dan menahan perhatian kita, jauh lebih cepat dan kuat dibandingkan dengan tumpukan tulisan biasa. Otak kita memang dirancang untuk memproses informasi visual ribuan kali lebih cepat daripada teks. Sebuah gambar atau video yang tepat bisa langsung membangkitkan emosi, menyampaikan suasana hati, atau bahkan menceritakan kisah yang kompleks dalam hitungan detik, meninggalkan kesan yang lebih mendalam dan mudah diingat.

Maka, tak heran kalau kita melihat bagaimana konten visual ini begitu mendominasi di berbagai platform media sosial. Hampir setiap unggahan, mulai dari *Instagram* yang memang berfokus pada visual, TikTok dengan video pendeknya yang adiktif, YouTube sebagai raksasa video, bahkan platform yang dulunya berbasis teks seperti Facebook atau X (Twitter), kini semuanya memanfaatkan kekuatan gambar, video, dan grafik. Ini bukan tanpa alasan; visual memang sangat ampuh untuk beberapa tujuan utama: pertama, untuk menyampaikan pesan dengan cara yang lebih jernih dan menarik, bahkan untuk informasi yang rumit sekalipun. Kedua, untuk memengaruhi pandangan atau perasaan audiens, membangun koneksi emosional, dan membentuk citra merek yang kuat. Dan yang paling penting, visual adalah kunci untuk membangun interaksi yang lebih hidup dengan para penggunanya. Sebuah foto yang inspiratif

atau video yang lucu bisa memicu banyak like, komentar, atau bahkan dibagikan berkali-kali, menciptakan percakapan dan komunitas yang aktif di sekitar konten tersebut (Brand et al., 2022). Dengan demikian, konten visual bukan lagi sekadar pelengkap, melainkan inti dari strategi komunikasi digital yang efektif.

2.4 Persepsi Audiens terhadap Visual

Persepsi audiens bisa kita ibaratkan seperti filter atau lensa pribadi yang dipakai setiap orang. Ini adalah cara unik bagaimana seseorang menerima informasi, lalu menafsirkannya, memahaminya, dan akhirnya memberikan makna terhadap apa yang ia lihat atau dengar. Jadi, informasi yang sama bisa saja diartikan berbeda oleh dua orang yang berbeda, karena persepsi bukanlah sekadar menerima data mentah, melainkan sebuah proses aktif di dalam pikiran kita. Menurut (Schacter, 2011), seorang ahli teori persepsi, proses ini sangatlah kompleks dan dipengaruhi oleh beberapa faktor utama yang membentuk 'lensa' pribadi kita.

Faktor-faktor yang memengaruhi persepsi itu antara lain adalah pengalaman masa lalu kita, apa yang kita harapkan akan terjadi, dan juga konteks di mana informasi itu muncul. Misalnya, seseorang yang punya pengalaman kurang baik dengan suatu merek mungkin akan cenderung memandang negatif informasi baru dari merek tersebut (pengalaman). Atau, jika kita sangat menantikan pengumuman penting, kita akan lebih fokus dan mungkin menafsirkan setiap petunjuk kecil sebagai bagian dari pengumuman itu (harapan). Demikian pula, informasi yang sama bisa memiliki makna berbeda jika disampaikan dalam suasana formal dibandingkan suasana santai (konteks). Selain itu, kepercayaan pribadi, nilai-nilai yang kita anut, bahkan mood kita saat itu, turut berperan dalam membentuk bagaimana kita memandang dan merasakan sesuatu.

Ketika kita berbicara dalam ranah media sosial yang begitu dinamis dan serba visual, persepsi audiens menjadi jauh lebih krusial. Di platform-platform ini, di mana gambar dan video adalah raja, persepsi audiens akan sangat dipengaruhi oleh tiga elemen kunci: kualitas visual konten itu sendiri, estetika atau keindahan desainnya, serta seberapa konsisten identitas visual yang ditampilkan. Konten dengan kualitas visual yang buruk—misalnya buram, gelap, atau tidak jelas—bisa langsung menciptakan kesan negatif, seolah-olah merek atau institusi tersebut tidak

serius atau kurang profesional. Sebaliknya, konten dengan resolusi tinggi, pencahayaan baik, dan framing yang cerdas akan otomatis dianggap lebih kredibel dan menarik.

Inilah mengapa elemen-elemen visual seperti penggunaan warna khas institusi yang konsisten, logo yang selalu muncul dengan tepat, atau gaya desain konten yang seragam di setiap postingan, menjadi sangat penting. Konsistensi semacam ini bukan sekadar masalah estetika; ini adalah fondasi untuk memperkuat citra merek di benak audiens. Bayangkan saja, ketika kita melihat warna tertentu atau bentuk logo khas, kita langsung teringat pada merek atau institusi tertentu. Hal ini membangun familiaritas, kepercayaan, dan secara tidak langsung menegaskan kesan profesionalisme serta standar yang tinggi. Audiens akan melihat institusi sebagai entitas yang terorganisir, dapat diandalkan, dan memiliki identitas yang kuat.

Pada akhirnya, sebuah persepsi positif terhadap konten visual yang disajikan, yang berarti audiens menganggapnya menarik, relevan, terpercaya, dan berkualitas tinggi akan membawa dampak yang sangat menguntungkan. Hal ini secara langsung berkontribusi pada peningkatan *brand awareness*, artinya institusi akan semakin dikenal luas dan mudah diingat oleh banyak orang. Lebih jauh lagi, persepsi positif ini juga akan membentuk citra institusi yang kuat dan baik di mata masyarakat, membantu mereka melihat institusi tersebut sebagai tempat yang unggul, inovatif, atau peduli. Dengan demikian, mengelola persepsi melalui visual yang cermat bukanlah pilihan, melainkan keharusan strategis di era digital ini.

2.5 Brand awareness

Brand awareness, atau yang sering kita sebut sebagai kesadaran merek, adalah sebuah kemampuan dasar yang dimiliki konsumen untuk mengenali, atau bahkan secara spontan mengingat, suatu merek tertentu di antara berbagai pilihan dalam kategori produk atau jasa yang sama. Ini bukan sekadar tahu nama mereknya, tapi lebih kepada seberapa familiar dan melekatnya merek tersebut di benak konsumen saat mereka memikirkan kebutuhan atau kategori produk terkait. Menurut (Keller, 1993), seorang pakar pemasaran yang banyak meneliti tentang merek, *brand awareness* memiliki dua dimensi utama yang saling melengkapi:

1. **Brand Recognition:** Kemampuan pengenalan merek. Ini terjadi ketika konsumen mampu mengidentifikasi sebuah merek saat mereka melihat atau mendengar sesuatu yang berhubungan dengannya. Bayangkan saja, ketika Anda melihat logo tertentu di jalan, mendengar *jingle* yang familiar di radio, atau melihat kemasan produk dengan desain khas di supermarket, dan Anda langsung tahu itu merek apa, itulah *brand recognition*. Konsumen tidak perlu mengingat mereknya dari nol, cukup dengan 'bantuan' visual atau audio, mereka bisa mengenali keberadaan merek tersebut. Ini sangat penting terutama di titik-titik penjualan di mana konsumen dihadapkan pada banyak pilihan dan harus membuat keputusan cepat.
2. **Brand Recall:** Kemampuan mengingat merek. Ini adalah level kesadaran yang lebih tinggi, di mana konsumen mampu mengingat sebuah merek tanpa bantuan atau petunjuk sama sekali, hanya dengan memikirkan kategori produknya. Misalnya, ketika Anda merasa haus dan berpikir "minuman ringan," merek apa yang pertama kali muncul di pikiran Anda? Atau ketika Anda ingin membeli kopi, merek apa yang langsung Anda sebutkan tanpa melihat daftar? Kemampuan *brand recall* menunjukkan seberapa kuat merek tersebut tertanam dalam memori jangka panjang konsumen dan seberapa dominannya ia dalam kategori produk tertentu. Merek dengan *brand recall* yang tinggi seringkali menjadi pilihan utama saat konsumen secara aktif mencari solusi atau produk.

Secara strategis, *brand awareness* ini merupakan langkah paling awal dan fondasi yang tak tergantikan dalam proses membangun loyalitas pelanggan yang langgeng. Bagaimana mungkin konsumen bisa menjadi loyal jika mereka bahkan tidak tahu atau tidak mengingat merek Anda? Kesadaran merek membangun jembatan pertama antara merek dan konsumen, menumbuhkan rasa familiaritas dan kepercayaan awal. Lebih dari itu, *brand awareness* juga menjadi faktor kunci yang sangat berpengaruh dalam setiap keputusan pembelian yang dibuat konsumen. Merek yang sudah dikenal luas dan sering muncul di benak konsumen cenderung dianggap lebih terpercaya, mengurangi risiko yang dirasakan konsumen, dan

seringkali menjadi pilihan default. Ini berarti, semakin tinggi kesadaran merek, semakin besar peluang sebuah merek untuk dipilih, bahkan di tengah persaingan yang ketat, sekaligus menjadi dasar bagi seluruh ekuitas dan nilai merek di masa depan.

2.6 Data Mining dan Klasifikasi

Data mining adalah sebuah proses fundamental dalam dunia teknologi informasi, yang berfokus pada upaya mengekstrak informasi atau pola berharga yang tersembunyi di dalam kumpulan data yang sangat besar. Mengingat volume data yang terus bertumbuh secara eksponensial di era digital ini, analisis manual menjadi tidak mungkin. Oleh karena itu, data mining hadir sebagai disiplin ilmu yang memungkinkan kita untuk menggali wawasan mendalam, menemukan hubungan tersembunyi, dan mengidentifikasi tren yang tidak terlihat secara kasat mata, demi membantu kita membuat keputusan yang lebih cerdas dan efektif.

Dalam kerangka kerja data mining yang luas ini, salah satu teknik yang memiliki peran sentral dan paling sering diaplikasikan adalah klasifikasi. Ini adalah sebuah metode sistematis untuk mengkategorikan setiap 'item' data ke dalam 'kelas-kelas' atau kelompok yang sudah ditentukan sebelumnya, berdasarkan serangkaian atribut atau karakteristik spesifik yang dimiliki oleh data tersebut. Misalnya, klasifikasi dapat digunakan untuk secara otomatis memilah email ke dalam kategori 'penting' atau 'sampah', atau untuk mengidentifikasi seorang nasabah sebagai 'berisiko tinggi' atau 'berisiko rendah' berdasarkan profil finansialnya. Kemampuan pengelompokan ini sangatlah berguna karena menyederhanakan data kompleks menjadi segmen-segmen yang mudah dipahami dan dapat ditindaklanjuti.

Menurut (Han et al., 2006), klasifikasi dikategorikan sebagai teknik pembelajaran terawasi (*supervised learning*). Ini berarti algoritma belajar dari sebuah 'dataset berlabel'—yaitu kumpulan data di mana setiap entri sudah memiliki 'jawaban' atau kategori yang benar yang telah ditetapkan sebelumnya oleh manusia atau sistem. Melalui proses ini, algoritma membangun sebuah 'fungsi' atau model yang mampu memahami hubungan antara atribut data dengan label kategorinya. Setelah model terbentuk dan terlatih dengan baik menggunakan data historis ini, ia

kemudian dapat digunakan untuk memprediksi kelas atau kategori dari 'data baru' yang belum pernah ia lihat sebelumnya dengan tingkat akurasi tertentu.

Kemampuan klasifikasi untuk mengelompokkan data secara cerdas menjadikannya alat yang sangat berharga dalam berbagai aspek. Ia sangat berguna dalam pengambilan keputusan yang strategis, misalnya dengan membantu perusahaan mengidentifikasi calon pelanggan yang paling potensial untuk penawaran produk tertentu atau membantu profesional medis dalam mendiagnosis penyakit berdasarkan gejala pasien. Selain itu, klasifikasi juga sangat efektif dalam memprediksi perilaku, seperti memperkirakan apakah seorang pelanggan akan beralih ke pesaing (*churn*) atau memprediksi pola pembelian di masa depan. Lebih lanjut, teknik ini memainkan peran penting dalam analisis sentimen atau pengelompokan opini, di mana ribuan ulasan pelanggan atau komentar di media sosial dapat dikategorikan sebagai positif, negatif, atau netral. Dengan demikian, klasifikasi menjadi jembatan yang mentransformasi data mentah yang kompleks menjadi wawasan yang *actionable*, mendorong efisiensi operasional, dan memberikan keunggulan kompetitif di berbagai sektor.

2.7 Algoritma *Naïve Bayes*

Naïve Bayes adalah kumpulan algoritma klasifikasi yang didasarkan pada Teorema Bayes, yang mengasumsikan bahwa fitur (atau variabel) yang digunakan dalam prediksi adalah independent satu sama lain. Meskipun asumsi ini mungkin tidak selalu benar dalam praktik, *Naïve Bayes* sering kali memberikan hasil yang baik, terutama dalam kasus di mana ukuran dataset relatif kecil atau saat fitur memiliki hubungan yang kompleks. Salah satu algoritma klasifikasi yang berbasis pada penerapan *Teorema Bayes*, (Mulyoto, 2024) dengan asumsi sederhana bahwa setiap fitur bersifat independen terhadap fitur lainnya. Algoritma ini termasuk ke dalam kategori *supervised learning*, dimana model dibangun berdasarkan data latih yang sudah berlabel. *Teorema Bayes* menyatakan bahwa probabilitas suatu hipotesis H terhadap data D dapat dihitung menggunakan rumus:

$$P(H | D) = \frac{P(D | H) \times P(H)}{P(D)}$$

dimana:

$P(H|D)$ adalah probabilitas hipotesis setelah melihat data (*posterior probability*),

$P(D|H)$ adalah probabilitas data diberikan hipotesis benar (*likelihood*),

$P(H)$ adalah probabilitas awal hipotesis (*prior probability*),

$P(D)$ adalah probabilitas data secara keseluruhan (*evidence*).

Dalam konteks klasifikasi, *Naïve Bayes* berfungsi dengan menghitung probabilitas dari masing-masing kelas terhadap data masukan, lalu memilih kelas dengan probabilitas tertinggi. Asumsi independensi antar fitur menyederhanakan perhitungan probabilitas bersama menjadi perkalian probabilitas masing-masing fitur terhadap kelas tertentu. Dengan kata lain, *Naïve Bayes* menghitung:

$$P(C | X) \propto P(C) \times P(x_1 | C) \times P(x_2 | C) \times \dots \times P(x_n | C)$$

dimana:

C adalah kelas,

X adalah vektor fitur (x_1, x_2, \dots, x_n) ,

$P(C)$ adalah probabilitas awal kelas,

$P(x_i|C)$ adalah probabilitas fitur ke- i diberikan kelas C .

Dalam praktiknya, *Naïve Bayes* memiliki beberapa varian, antara lain *Gaussian Naïve Bayes* (untuk data yang berdistribusi normal), *Multinomial Naïve Bayes* (cocok untuk data diskrit seperti frekuensi kata), dan *Bernoulli Naïve Bayes* (untuk data biner). Meskipun asumsi independensi antar fitur jarang terpenuhi di dunia nyata, algoritma ini sering kali memberikan hasil yang cukup baik, terutama dalam masalah klasifikasi teks seperti *filtering* spam dan analisis sentimen.

Kelebihan dari *Naïve Bayes* adalah kemampuannya untuk melakukan pelatihan dan prediksi dengan sangat cepat, performa yang cukup baik pada dataset besar, serta ketahanannya terhadap fitur yang tidak relevan. Namun, kelemahannya terletak pada ketergantungannya terhadap asumsi independensi, serta potensi masalah jika suatu kombinasi fitur dan kelas tidak pernah muncul dalam data latih (menghasilkan probabilitas nol). Untuk mengatasi masalah tersebut, biasanya digunakan teknik *smoothing* seperti *Laplace smoothing*.

Dengan pendekatan yang sederhana namun efektif, *Naïve Bayes* tetap menjadi salah satu algoritma yang banyak digunakan dalam berbagai aplikasi klasifikasi. (Webb, 2016).

2.8 Google Colab

Google Colab, singkatan dari *Google Colaboratory*, adalah *platform cloud* yang disediakan oleh Google untuk menulis dan menjalankan kode *Python* langsung melalui browser tanpa memerlukan konfigurasi tambahan. *Platform* ini memanfaatkan infrastruktur *cloud Google*, memberikan lingkungan pengembangan yang kuat, dan menyediakan akses gratis ke GPU dan TPU (unit pemrosesan tensor).

Dengan fitur kolaborasi yang memungkinkan pengguna untuk berbagi notebook dan bekerja secara bersama-sama secara real-time, *Google Colab* menjadi pilihan populer bagi para pengembang, peneliti, dan pelajar untuk menjalankan proyek-proyek *Python* mereka. *Google Colab* dapat diakses melalui peramban web tanpa memerlukan instalasi perangkat lunak tambahan, serta menyediakan kemudahan dalam kolaborasi, memungkinkan pengguna untuk berbagi notebook secara langsung dengan orang lain.

Platform ini sangat berguna bagi para pengembang, peneliti, dan pemula dalam bidang pemrograman untuk menjalankan proyek-proyek *Python* tanpa harus khawatir tentang konfigurasi perangkat keras. (Ridwan Nazar, 2024).

2.9 Python

Python adalah bahasa pemrograman tingkat tinggi yang terkenal karena sintaksisnya yang sederhana, mudah dipahami, dan fleksibel. Dikembangkan oleh Guido van Rossum, *Python* memiliki berbagai kegunaan, termasuk pengembangan perangkat lunak, analisis data, kecerdasan buatan, pengembangan web, dan banyak lagi. Bahasa ini menjadi pilihan utama bagi banyak pengembang karena kombinasi antara kejelasan sintaksis dan kemampuannya dalam menangani berbagai tugas pemrograman. Dalam konteks ini, praktik pemrograman *Python* akan dilakukan menggunakan *platform Google Colaboratory*, yang memungkinkan eksekusi kode

secara langsung melalui browser tanpa memerlukan instalasi perangkat lunak tambahan. (Febby Wilyani et al., 2024).

2.10 Penelitian Terdahulu

Beberapa penelitian terdahulu relevan dengan topik ini dan memberikan dasar bagi penelitian ini, terlampir pada tabel beriku.

Tabel 2. 1 Tabel Penelitian Terdahulu

NO	Peliti	Judul	Metode	Hasil
1	(Puspitasari, 2024)	Analisis Desain Konten Visual terhadap Performa <i>Instagram</i> Universitas Utpadaka Swastika @utpas.ac.id	Kuantitatif, observasi konten dan analisis engagement	Desain konten visual yang konsisten dan informatif meningkatkan engagement serta jangkauan akun <i>Instagram</i> kampus. Relevan sebagai studi institusi pendidikan seperti IIB Darmajaya.
2	(KHAIRIL IKHSAN, 2022)	Analisis Social CRM pada Media Sosial <i>Instagram</i> Menggunakan <i>Naïve Bayes Classifier</i> untuk Mengukur Retensi Pelanggan	Naïve Bayes	Desain konten visual yang menarik secara signifikan meningkatkan <i>brand awareness</i> melalui peningkatan engagement di akun <i>Instagram</i> @Richofsecondshoes. Engagement berperan sebagai mediator yang memperkuat