

BAB I

PENDAHULUAN

1.1 Latar Belakang

E-commerce merupakan konsep bisnis digital yang memanfaatkan teknologi internet dan platform elektronik untuk mendukung aktivitas pertukaran barang, jasa, maupun informasi antara penjual dan pembeli secara daring [1]. Perkembangan teknologi digital dan semakin banyaknya orang yang menggunakan internet mendorong pertumbuhan aktivitas belanja online, serta memudahkan para pelaku usaha dan konsumen dalam melakukan transaksi secara daring [2]. Sistem pembelajaran mesin dan analisis data dapat digunakan untuk menemukan pola perilaku pelanggan dari data transaksi, sehingga membantu organisasi memahami karakteristik pengguna secara lebih akurat [3].

Metode konvensional dalam memahami cara orang membeli sering kali kesulitan mengelola banyak data dan pola perilaku yang rumit. Diperlukan pendekatan analisis yang mampu memanfaatkan data secara lebih efektif. Salah satu cara yang semakin populer adalah pembelajaran mesin, yaitu pendekatan yang memungkinkan komputer belajar dari pola data dan digunakan untuk membuat model prediksi serta membantu dalam pengambilan keputusan. Statistical learning sendiri merupakan sekumpulan metode yang digunakan untuk memahami dan memodelkan hubungan dalam data sehingga dapat menghasilkan informasi yang bermanfaat bagi proses analisis dan prediksi [4]. [5] pembelajaran mesin telah menjadi tulang punggung revolusi teknologi digital karena mampu mengolah data dan menghasilkan prediksi yang bermakna dalam berbagai bidang. machine learning adalah cara yang memungkinkan komputer memahami pola dalam data, sehingga sistem bisa membuat peramalan atau mengambil keputusan secara otomatis.

Machine learning (ML) menjadi salah satu pendekatan yang relevan dalam analisis data karena mampu mengolah data dalam jumlah besar serta mengidentifikasi pola yang sulit ditemukan secara manual. Pendekatan ini memungkinkan sistem komputer mempelajari hubungan antar variabel dalam data

untuk menghasilkan model prediksi yang lebih akurat. Dalam konteks analisis data modern, metode machine learning banyak digunakan untuk menemukan pola, melakukan klasifikasi, serta mendukung proses pengambilan keputusan berbasis data[4]. Dalam analisis data, metode klasifikasi digunakan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan karakteristik yang dimilikinya. Proses ini dilakukan dengan membangun model yang mampu mempelajari pola dari data pelatihan sehingga dapat memprediksi kategori dari data baru. Beberapa algoritma yang umum digunakan dalam metode klasifikasi antara lain Logistic Regression, Naïve Bayes, dan Random Forest, yang mampu memodelkan hubungan antar variabel serta mengidentifikasi pola dalam data untuk menghasilkan prediksi yang akurat[3]

Selain itu, perkembangan mengumpulkan data dalam jumlah besar membutuhkan teknik analisis yang mampu menarik informasi berguna dari data tersebut. Salah satu metode yang sering digunakan adalah pembelajaran statistik, yaitu sekelompok teknik statistik yang digunakan untuk mengenali pola dan hubungan antar variabel dalam data yang rumit, sehingga bisa digunakan untuk menganalisis dan memprediksi berdasarkan data [4]. Cara ini memungkinkan peneliti mendapatkan informasi yang lebih terperinci dari data yang sudah ada dan membantu proses pengambilan keputusan menjadi lebih baik. Dalam praktiknya, berbagai metode pembelajaran mesin bisa digunakan untuk menganalisis data dan mencari pola yang tidak terlihat, sehingga bisa dimanfaatkan dalam berbagai bidang analisis data dan pengambilan keputusan [5]. Penerapan teknik klasifikasi dalam penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai pola perilaku belanja konsumen.

Pemanfaatan teknik klasifikasi dalam machine learning memberikan kemampuan untuk menganalisis data secara lebih sistematis dan terstruktur. Dengan memanfaatkan data historis sebagai data pelatihan, model klasifikasi dapat mempelajari pola yang terdapat dalam data sehingga mampu mengklasifikasikan data baru secara otomatis. Proses ini memungkinkan organisasi atau perusahaan memperoleh informasi yang lebih akurat mengenai perilaku pengguna atau konsumen. Oleh karena itu, penerapan metode klasifikasi dalam analisis data

menjadi salah satu pendekatan penting dalam pengolahan data modern karena mampu membantu proses pengambilan keputusan berbasis data secara lebih efektif dan efisien [3]

Pemanfaatan teknik klasifikasi dalam machine learning memberikan kemampuan untuk menganalisis data secara lebih sistematis dan terstruktur. Dengan memanfaatkan data historis sebagai data pelatihan, model klasifikasi dapat mempelajari pola yang terdapat dalam data sehingga mampu mengklasifikasikan data baru secara otomatis. Proses ini memungkinkan organisasi atau perusahaan memperoleh informasi yang lebih akurat mengenai perilaku pengguna atau konsumen. Oleh karena itu, penerapan metode klasifikasi dalam analisis data menjadi salah satu pendekatan penting dalam pengolahan data modern karena mampu membantu proses pengambilan keputusan berbasis data secara lebih efektif dan efisien [3]

Selain itu, keberhasilan penerapan metode klasifikasi juga dipengaruhi oleh kemampuan model dalam mempelajari hubungan antara variabel yang terdapat dalam dataset. Model klasifikasi yang baik tidak hanya mampu mengidentifikasi pola dalam data, tetapi juga mampu menghasilkan prediksi yang konsisten ketika diterapkan pada data baru. Oleh karena itu, pemilihan algoritma klasifikasi yang tepat menjadi faktor penting dalam proses analisis data. Penggunaan berbagai algoritma klasifikasi seperti Logistic Regression, Naïve Bayes, dan Random Forest memungkinkan peneliti membandingkan performa masing-masing metode dalam mengidentifikasi pola data serta menghasilkan prediksi yang lebih akurat [6]

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana karakteristik data perilaku belanja konsumen setelah dilakukan proses preprocessing, termasuk One-Hot Encoding dan penanganan ketidakseimbangan data?

2. Bagaimana pengaruh fitur numerik terhadap status pelanggan (Subscriber dan Non-Subscriber) berdasarkan hasil analisis visualisasi boxplot?
3. Bagaimana perbandingan kinerja algoritma dalam mengklasifikasikan status langganan, serta algoritma mana yang memberikan performa terbaik berdasarkan metrik Accuracy, Precision, Recall, dan F1-score? e-commerce.

1.3 Batasan Masalah

Agar penelitian ini terarah dan fokus, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Penelitian berfokus pada analisis komparatif kinerja algoritma pembelajaran mesin untuk klasifikasi perilaku belanja konsumen pada platform e-commerce.
2. Dataset yang digunakan berasal dari Kaggle (*Customer Shopping Behavior*) dengan total 3.900 baris data dan terdiri dari 18 kolom (atribut).
3. Analisis dilakukan menggunakan metrik evaluasi berupa akurasi, presisi, recall, dan F1-score tanpa membahas spek prediksi penjualan atau rekomendasi produk lanjutan.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah untuk menjelaskan arah secara hasil yang ingin dicapai dari kegiatan penelitian yang dilakukan. Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengatasi masalah ketidakseimbangan kelas (class imbalance) dengan menerapkan teknik SMOTE pada data latih guna menyeimbangkan distribusi target menjadi seimbang.
2. Membangun dan mengevaluasi model klasifikasi melalui analisis komparatif performa tiga algoritma *machine learning*, yaitu *Random forest*, *Logistic Regression*, dan *Naive Bayes*.
3. Menentukan algoritma terbaik berdasarkan hasil visualisasi perbandingan metrik evaluasi untuk menemukan model yang paling akurat dalam memprediksi perilaku belanja konsumen pada platform e-commerce.

1.5 Manfaat Penelitian

1.5.1 Manfaat Teoritis

Secara teoritis, penelitian ini dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang data science dan analisis perilaku konsumen, khususnya terkait penerapan algoritma machine learning dalam proses klasifikasi perilaku belanja konsumen pada e-commerce dengan menekankan pada perbandingan performa algoritma Random forest, Logistic regression dan Naïve Bayes. Hasil penelitian ini dapat dijadikan sebagai acuan bagi penelitian selanjutnya yang berfokus pada pengembangan dan optimasi model prediksi berbasis data dalam konteks pemasaran digital dan e-commerce.

1.5.2 Manfaat Praktis

Secara praktis, manfaat penelitian ini ditujukan bagi industri e-commerce dan peneliti selanjutnya. Bagi industri e-commerce, hasil analisis komparatif antara algoritma Random Forest, Logistic Regression, dan Naive Bayes dapat dimanfaatkan sebagai sistem pendukung keputusan (Decision Support System) untuk memprediksi status langganan dan loyalitas konsumen secara lebih presisi. Hal ini berdampak langsung pada optimalisasi strategi pemasaran, di mana perusahaan dapat menentukan target audiens yang lebih akurat sehingga meningkatkan efisiensi biaya operasional kampanye promosi. Sementara itu, bagi peneliti selanjutnya, kajian ini dapat menjadi referensi metodologis dan rujukan ilmiah mengenai efektivitas berbagai algoritma klasifikasi dalam memproses data perilaku belanja, khususnya pada kasus yang memerlukan penyeimbangan distribusi data guna menghasilkan model prediksi yang lebih stabil dan objektif. Selain itu, hasil penelitian ini juga dapat dijadikan dasar untuk pengembangan model prediksi yang lebih kompleks dengan mempertimbangkan variasi fitur yang lebih luas. Penerapan teknik evaluasi yang komprehensif dalam penelitian ini diharapkan dapat membantu peneliti lain dalam menentukan metrik yang paling relevan sesuai dengan karakteristik datanya. Model yang dihasilkan juga berpotensi diintegrasikan ke dalam sistem analitik pelanggan berbasis digital secara bertahap. Dengan demikian, penelitian ini tidak hanya bersifat akademis, tetapi juga memiliki peluang

implementasi nyata dalam praktik bisnis e-commerce.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari lima bab utama, yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi studi literatur dan dasar teori yang digunakan sebagai landasan penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang metode dan langkah-langkah penelitian yang dilakukan, meliputi objek penelitian, alur penelitian (*flowchart*), serta alat dan bahan yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pengujian dari algoritma yang digunakan, disertai pembahasan secara analitis terhadap hasil yang diperoleh.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian dan saran atau masukan untuk penyempurnaan penelitian selanjutnya.