BAB I PENDAHULUAN

I.1 Latar Belakang

Industri e-commerce terus mengalami pertumbuhan pesat seiring dengan digitalisasi perilaku konsumen dan ketersediaan data transaksi yang semakin besar. Data pelanggan kini menjadi sumber daya penting dalam pengambilan keputusan bisnis, khususnya untuk memahami pola perilaku dan melakukan segmentasi pasar yang lebih akurat (Chen, Xu, & Wang, 2024). Namun, kompleksitas dan tingginya dimensi data pelanggan seringkali menjadi tantangan utama dalam pengolahan informasi, terutama dalam penerapan algoritma klasifikasi (classification) berbasis machine learning (Shalev-Shwartz & Ben-David, 2014).

Permasalahan yang kerap muncul adalah banyaknya atribut atau fitur dalam dataset yang tidak semuanya relevan, yang berpotensi menurunkan performa model dan meningkatkan biaya komputasi (Zhang, Li, & Xie, 2021). Dalam konteks ini, proses penyaringan fitur atau feature selection menjadi tahapan penting untuk memastikan efisiensi dan efektivitas proses klasifikasi (Wang & Zhang, 2022). Salah satu pendekatan yang dinilai tepat untuk menangani permasalahan ini adalah penggunaan algoritma Reduct dari teori Rough Set, yang dapat menyeleksi atribut yang benar-benar signifikan tanpa menghilangkan informasi utama (Duda, Hart, & Stork, 2012).

Sebagai sumber data, penelitian ini menggunakan Feed Grains Database dari United States Department of Agriculture (USDA, 2024), yang berisi atribut numerik seperti hasil panen, harga jual, stok akhir, dan suplai komoditas pertanian. Meskipun tidak secara langsung berasal dari data pelanggan e-commerce, struktur dataset ini cocok untuk disimulasikan dalam konteks klasifikasi berbasis fitur numerik yang merepresentasikan karakteristik konsumen.

Untuk membangun sistem klasifikasi, algoritma Support Vector Machine (SVM) digunakan karena dikenal efektif dalam menangani data berdimensi tinggi dan bersifat non-linear (Vapnik, 1995). SVM juga terbukti unggul dalam berbagai studi kasus klasifikasi (Platt, 1999; Pedregosa et al., 2011). Namun, untuk memastikan bahwa model bekerja secara efisien, dibutuhkan proses seleksi fitur agar hanya atribut paling relevan yang digunakan sebagai input bagi algoritma klasifikasi (Haykin, 2009).

Dengan menerapkan algoritma Reduct, penelitian ini bertujuan menyederhanakan struktur data dan meningkatkan efisiensi proses klasifikasi. Pendekatan ini juga sejalan dengan prinsip interpretability dalam machine learning, yakni agar model tidak hanya akurat tetapi juga mudah dipahami dan dijelaskan (Chen et al., 2024). Secara keseluruhan, sistem klasifikasi yang dikembangkan dalam penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap praktik pengelolaan data pelanggan, serta memiliki potensi untuk diterapkan pada sistem informasi skala enterprise dalam pengambilan keputusan strategis.

I.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan permasalahan untuk penelitian ini adalah:

1. Bagaimana kondisi eksisting data pelanggan *e-commerce* yang kompleks dan berdimensi tinggi memengaruhi efektivitas model klasifikasi dalam *machine learning*?
2. Bagaimana algoritma *Reduct* dapat menjadi solusi untuk menyaring atribut yang tidak relevan dalam data pelanggan guna meningkatkan efisiensi dan akurasi klasifikasi di masa mendatang??

I.3 Tujuan Tugas Akhir

Penelitian ini bertujuan untuk:

1. Menjelaskan kondisi eksisting dari data pelanggan *e-commerce* yang kompleks dan berdimensi tinggi serta dampaknya terhadap performa model klasifikasi dalam *machine learning*.

2. Mengimplementasikan algoritma *Reduct* sebagai solusi penyaringan atribut yang tidak relevan guna meningkatkan efisiensi dan akurasi klasifikasi pelanggan *e-commerce* di masa mendatang.

I.4 Manfaat Tugas Akhir

Manfaat penelitian ini:

1. Secara akademis, penelitian ini dapat menjadi referensi dalam pengembangan metode penyaringan data untuk klasifikasi.
2. Secara praktis, hasil penelitian ini dapat membantu perusahaan *e-commerce* dalam mengelola data pelanggan secara lebih efisien dan efektif.

I.5 Batasan dan Asumsi Tugas Akhir

Penelitian ini dilakukan dalam ruang lingkup tertentu guna menjaga ketepatan fokus, kesesuaian konteks, serta keterbatasan sumber daya yang tersedia. Batasan-batasan yang diterapkan antara lain;

1. Data yang digunakan dalam penelitian ini adalah *Feed Grains Database* yang diperoleh dari situs resmi pemerintah Amerika Serikat (https://catalog.data.gov), dan difokuskan pada data yang dapat dimanfaatkan untuk simulasi klasifikasi perilaku pelanggan berbasis tren permintaan atau konsumsi komoditas.
2. Penelitian ini tidak menggunakan data pelanggan dari sistem *e-commerce* aktual, melainkan menggunakan *proxy data* dari sektor pertanian sebagai pendekatan terhadap analisis segmentasi dan klasifikasi berbasis pola data.
3. Model klasifikasi dibatasi pada penggunaan algoritma *Support Vector Machine* (SVM), yang diintegrasikan dengan algoritma *Reduct* untuk proses *feature selection*.
4. Lingkup eksperimen bersifat simulatif dan terbatas pada atribut yang tersedia dalam *dataset*, tanpa melakukan penggabungan dengan sumber data eksternal lainnya.

5. Waktu pengerjaan dan analisis data dibatasi dalam rentang waktu semester akhir studi, sehingga eksplorasi hanya dilakukan pada *subset data* tertentu yang telah melalui proses pembersihan..

Penelitian ini juga menggunakan beberapa asumsi untuk menyederhanakan kompleksitas dan memungkinkan proses analisis dilakukan secara sistematis, di antaranya:

1. Dataset *Feed Grains* dianggap memiliki struktur dan variasi atribut yang dapat disesuaikan untuk menggambarkan fenomena segmentasi pelanggan secara konseptual.
2. Atribut yang dipertahankan setelah penyaringan oleh algoritma *Reduct* diasumsikan memiliki pengaruh signifikan terhadap proses klasifikasi dan cukup representatif terhadap tujuan segmentasi.
3. Proses klasifikasi pelanggan dilakukan dengan asumsi bahwa data bersifat statis selama periode analisis dan tidak mengalami perubahan mendasar dalam waktu dekat.
4. Evaluasi performa model berbasis metrik akurasi, *precision*, *recall*, dan *F1-score* diasumsikan cukup untuk mengukur efektivitas algoritma dalam konteks simulatif.

I.6 Sistematika Laporan

Laporan Tugas Akhir ini disusun secara sistematis ke dalam enam bab utama, dengan susunan sebagai berikut:

1. BAB I – PENDAHULUAN

   Bab ini membahas latar belakang masalah yang menjadi dasar penelitian, perumusan masalah, tujuan dan manfaat penelitian, serta batasan penelitian. Disertakan pula sistematika penulisan untuk memberi gambaran umum isi laporan.
2. BAB II – TINJAUAN PUSTAKA

   Bab ini menyajikan kajian literatur dan landasan teori yang relevan dengan topik penelitian. Termasuk di dalamnya teori-teori tentang *machine learning*, *feature selection*, algoritma *Support Vector Machine*,

serta algoritma *Reduct*. Disertakan pula analisis penelitian terdahulu dan alasan pemilihan metode.

3. BAB III – METODOLOGI PENELITIAN

   Bab ini menguraikan kerangka pemecahan masalah dan sistematika penyelesaian masalah secara terstruktur. Termasuk di dalamnya tahapan pengumpulan data, pengolahan data, pengembangan sistem, serta metode evaluasi yang digunakan. Diagram alir dan *flowchart* turut digunakan untuk memperjelas alur proses.

4. BAB IV – ANALISIS DAN PERANCANGAN

   Bab ini menjelaskan proses pengumpulan dan analisis data numerik, serta tahap perancangan sistem klasifikasi. Termasuk dalam bab ini adalah pemilihan fitur, proses seleksi dengan algoritma *Reduct*, perancangan struktur proyek, dan dokumentasi visual seperti heatmap korelasi dan diagram proses klasifikasi.

5. BAB V – VALIDASI, ANALISIS HASIL, DAN IMPLIKASI

   Bab ini berisi hasil pengujian sistem baik dari sisi verifikasi maupun validasi, serta evaluasi terhadap kinerja model. Analisis dilakukan terhadap hasil klasifikasi sebelum dan sesudah reduksi fitur. Bab ini juga menguraikan rencana implementasi, dampak nyata dari solusi yang dikembangkan, serta potensi pengembangan ke depan.

6. BAB VI – KESIMPULAN DAN SARAN

   Bab ini memuat kesimpulan dari seluruh penelitian, menjawab tujuan penelitian yang telah ditetapkan. Selain itu, disampaikan pula saran untuk perbaikan sistem dan arah pengembangan penelitian selanjutnya.

7. DAFTAR PUSTAKA

   Memuat referensi yang digunakan selama proses penyusunan laporan, ditulis menggunakan format APA Style.