

BAB I

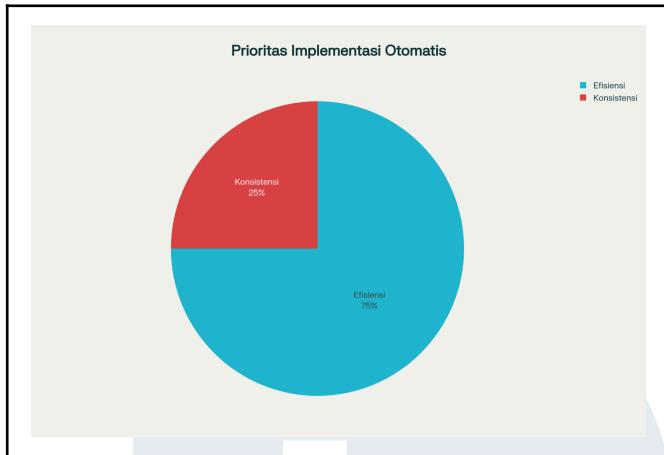
PENDAHULUAN

1.1 Latar Belakang Masalah

PT. XYZ merupakan *platform e-commerce* di Indonesia yang melayani lebih dari 25.000 pelanggan aktif setiap bulan di berbagai kategori produk seperti kebutuhan sehari-hari, makanan, minuman, dan lainnya. Dalam menghadapi dinamika perilaku konsumen yang sangat beragam dan cepat berubah, proses clustering pelanggan menjadi aspek vital untuk mendukung efektivitas strategi pemasaran serta personalisasi layanan berbasis data demografis dan transaksi.

Di awal proses penelitian ini, dilakukan wawancara dengan tim internal PT. XYZ untuk menggali secara mendalam kendala yang dihadapi saat ini. Dari wawancara ini diketahui bahwa terdapat kendala dalam proses melakukan clustering pelanggan untuk keperluan bisnis, dan hal ini yang diangkat menjadi tujuan utama penelitian. Wawancara ini menjadi tahap awal untuk memahami konteks operasional dan tantangan praktis yang dialami, seperti keterbatasan sumber daya dan lamanya waktu analisis. Berdasarkan hasil wawancara tersebut, kemudian disusun dan diberikan kuesioner kepada tim terkait untuk memperoleh data yang lebih terstruktur dan spesifik mengenai kendala tersebut. Langkah ini sekaligus membantu mengidentifikasi metrik-metrik utama yang relevan untuk dijadikan fokus penelitian dan pengembangan solusi.

Berdasarkan hasil dari kuesioner yang diberikan kepada tim dari PT. XYZ, diketahui bahwa clustering pelanggan yang saat ini dijalankan masih berbasis proses manual mengandalkan *spreadsheet* dan *query* sederhana dari database transaksi. Proses ini umumnya memerlukan waktu 1–3 hari kerja di setiap siklus, mulai dari pengumpulan, pembersihan, hingga pengelompokan data serta penyusunan laporan profil segmen. Selain memakan waktu lama, proses manual juga membutuhkan human resources dan akan terus bertambah seiring dengan bertambahnya jumlah pelanggan, kompleksitas data, dan pertumbuhan bisnis.



Gambar 1.1 Chart prioritas tujuan penelitian

Oleh karena itu, dibutuhkan solusi otomatis berbasis *machine learning* yang secara khusus mampu mengolah data campuran—baik numerik maupun kategorikal secara efisien dan *scalable*, agar dapat menghasilkan clustering pelanggan yang tidak hanya konsisten dan cepat, tetapi juga relevan, *robust*, serta dapat diandalkan dalam pengambilan keputusan [11].

Salah satu pendekatan yang sangat relevan dengan kebutuhan dan karakteristik data *e-commerce* lokal adalah implementasi algoritma Ensemble K-Prototypes *clustering*. Keunggulan utama solusi ini terletak pada kemampuannya menangani atribut numerik dan kategorikal secara *native* yang telah dibuktikan oleh penelitian yang dilakukan oleh Huang [6], dan dikombinasikan dengan metode *ensemble* yang dapat meningkatkan performa dari algoritma clustering yang telah dibuktikan oleh penelitian terdahulu [1][8]. Pendekatan ini diharapkan dapat memberikan hasil clustering yang stabil, serta secara signifikan memangkas waktu proses clustering. Dengan demikian, diharapkan dengan menggunakan solusi Ensemble K-Prototypes memungkinkan PT. XYZ memperoleh *insight* segmen pelanggan yang relevan, dalam waktu yang lebih singkat, serta tetap menjaga skalabilitas seiring pertumbuhan data dan kompleksitas bisnis.

1.2 Pertanyaan Penelitian

- 1.2.1 Berapa lama waktu yang dibutuhkan untuk melakukan clustering pelanggan menggunakan metode Ensemble K-Prototypes?
- 1.2.2 Bagaimana performa dan konsistensi Ensemble K-Prototypes *clustering* dalam menangani data pelanggan PT. XYZ dengan tipe data campuran (numerik dan kategorikal)?

1.3 Batasan Penelitian

- 1.3.1 Penelitian dilakukan menggunakan data transaksi pada perusahaan *e-commerce* PT. XYZ dengan fokus pada data pelanggan di wilayah Kota Tangerang.
- 1.3.2 Data transaksi yang digunakan untuk penelitian adalah kategori produk yang dibeli, dan data demografis yang digunakan adalah usia dan gender pelanggan.
- 1.3.3 Data transaksi yang digunakan adalah data transaksi pada bulan Juni, Juli, dan Agustus pada tahun 2025.

1.4 Tujuan Penelitian

- 1.4.1 Mengembangkan solusi clustering pelanggan otomatis berbasis Ensemble K-Prototypes yang dapat mengurangi waktu proses clustering dari 1-3 hari kerja menjadi kurang dari 1 hari kerja (8 jam) untuk *e-commerce* PT. XYZ.

1.5 Manfaat Penelitian

1. Bagi perusahaan *e-commerce* PT. XYZ:
 - Mengurangi waktu proses clustering pelanggan dari 1-3 hari kerja menjadi kurang dari 1 hari kerja, sehingga meningkatkan efisiensi operasional tim marketing dan data science secara signifikan.
 - Mendapatkan hasil clustering yang lebih akurat dan konsisten untuk mendukung pengambilan keputusan bisnis yang lebih tepat.
 - Memiliki sistem clustering yang *scalable* dan dapat dikonfigurasi ulang jika terdapat perubahan atau penambahan data di masa depan.

2. Bagi Peneliti:

- Mendapatkan pengalaman praktis dalam mengimplementasikan algoritma *machine learning* (K-Prototypes, Ensemble Learning, AdaBoost) pada kasus bisnis *e-commerce*.
- Memahami *end-to-end process* dari identifikasi masalah, perancangan solusi, dan implementasi solusi.

3. Bagi Akademisi dan Peneliti Lain:

- Memberikan kontribusi pada pengembangan metode *ensemble clustering* untuk data campuran yang masih terbatas dalam aplikasi.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut :

BAB 1 PENDAHULUAN

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

BAB 2 LANDASAN TEORI

Bab ini berisi justifikasi solusi yang ditemukan dan tinjauan teori untuk menjelaskan metode-metode yang diterapkan dalam penelitian.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi perancangan solusi yang dilakukan dan metode pengujian untuk menilai hasil dari metrik penanda keberhasilan solusi.

BAB 4 HASIL DAN DISKUSI

Bab ini berisi spesifikasi sistem, implementasi solusi, dan pengujian dan analisis solusi yang ditemukan.

BAB 5 SIMPULAN DAN SARAN

Bab ini berisi simpulan dan saran dari hasil penelitian yang dilakukan.