

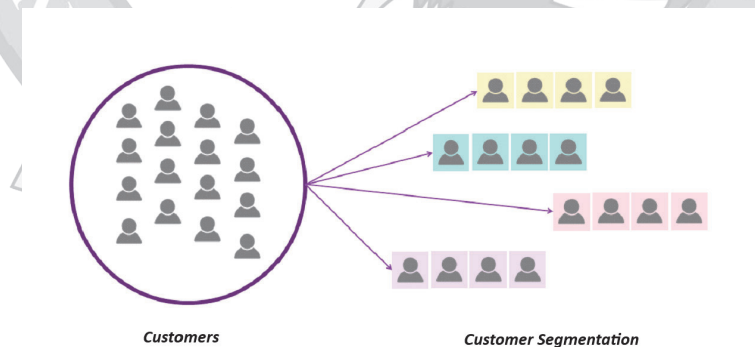
BAB I

PENDAHULUAN

1.1 Latar Belakang

Ilmu matematika telah berkembang dengan pesat dan tidak hanya dapat diaplikasikan dalam bidang matematika, tetapi dapat diaplikasikan dalam berbagai bidang termasuk ekonomi. Dengan berjalannya waktu, semakin banyak pesaing dalam dunia ekonomi yang menyebabkan pelanggan memiliki banyak pilihan. Perusahaan harus memusatkan pemasaran yang sesuai dengan kebutuhan dan keinginan pelanggan. Pemasaran memiliki peran yang penting dalam sebuah perusahaan, karena dengan berfokus pada kebutuhan pelanggan, perusahaan dapat meraih keuntungan yang berkelanjutan dalam jangka panjang [1].

Segmentasi pelanggan merupakan hal yang sangat penting dalam mengetahui minat dan kepuasan pasar sehingga dapat mengetahui langkah yang tepat dalam pemasaran [2]. Segmentasi pelanggan dapat dilihat dari berbagai aspek, seperti demografi, geografi, perilaku, dan psikografis [2]. Hal ini dapat membantu pengelola usaha untuk memusatkan pemasaran sesuai dengan kepuasan konsumen. Metode *Machine Learning* dapat digunakan untuk mengklasifikasi segmentasi pelanggan berdasarkan kebutuhan perusahaan untuk menentukan strategi pemasaran. Metode yang digunakan adalah *Decision Tree*, *Random Forest*, *XGBoost*, *SVM* dan *Hierarchical Clustering*. Pada Gambar 1.1 diberikan diagram ilustrasi mengenai segmentasi pelanggan.



Gambar 1.1: Ilustrasi Segmentasi Pelanggan
Sumber: *Machine Learning For Customer Segmentation*[3].

Metode *Decision Tree* merupakan bagian dari *supervised learning* berupa

grafik yang memberikan hasil dalam bentuk pohon yang terdiri dari simpul dan cabang. Setiap simpul menggambarkan atribut yang akan diklasifikasi dan setiap cabang memberikan nilai yang dapat diambil oleh simpul tersebut [4]. *Decision Tree* memberikan interpretabilitas yang tinggi karena memberikan hasil yang logis dan mudah dipahami. Metode ini sensitif terhadap gangguan kecil pada data yang dapat menyebabkan *overfitting*, tetapi hal ini dapat diatasi dengan bantuan metode *Random Forest* [5].

Metode *Random Forest* merupakan metode *ensemble* yang terdiri dari beberapa pohon serta model klasifikasi berkinerja tinggi. *Random Forest* menggunakan *bagging* untuk *Decision Tree* dengan mengekstraksi sampel secara acak saat melakukan *bootstrap sampling* [6]. *Variable importance* merupakan peringkat yang menunjukkan pentingnya suatu variabel untuk memprediksi hasil [5]. Sehingga, *variable importance* berperan penting dalam menentukan variabel yang paling berpengaruh dalam segmentasi pelanggan.

Metode *Extreme Gradient Boosting (XGBoost)* juga dapat digunakan dalam klasifikasi segmentasi pelanggan dengan mempelajari pola-pola kompleks dalam data pelanggan dan klasifikasi dilakukan berdasarkan pola-pola tersebut. *XGBoost (Extreme Gradient Boosting)* dikenal dengan kinerja tinggi dan efisiensi waktu komputasi, yang membuatnya sangat efektif untuk dataset besar dan kompleks. Teknik ini memanfaatkan *boosting* untuk meningkatkan akurasi melalui kombinasi dari beberapa model yang lemah.

Support Vector Machine (SVM) adalah metode dari *supervised learning* yang digunakan untuk mengatasi masalah klasifikasi dan juga regresi. Dalam klasifikasi, SVM menggunakan batas optimal atau *hyperplane* guna mengklasifikasi data ke dalam dua kelas yang berbeda. Terdapat istilah "fungsi *kernel*" yang memproyeksikan data ke dalam ruang berdimensi lebih tinggi guna melakukan transformasi pada data sebelum dapat dipisahkan oleh *hyperplane* [5]. Keunggulan SVM dalam menangani data berdimensi tinggi dan kemampuannya untuk menemukan *hyperplane* terbaik menjadikannya metode yang kuat dalam klasifikasi yang presisi.

Hierarchical Clustering merupakan metode *unsupervised learning* yang bertujuan untuk membangun hierarki kluster dalam bentuk struktur pohon. Pada umumnya, strategi *hierarchical clustering* terbagi menjadi *agglomerative* dan *divisive*. *Agglomerative* merupakan pendekatan *bottom-up* di mana setiap observasi dimulai dalam kluster itu sendiri yang kemudian digabungkan satu per satu, bergerak ke atas dalam hierarki. *Divisive* merupakan pendekatan *top-down* di mana semua observasi dimulai dalam satu kluster yang kemudian kluster tersebut

dibagi secara rekursif, bergerak ke bawah dalam hierarki [7].

Penerapan metode-metode *Machine Learning* ini memberikan keunggulan kompetitif bagi perusahaan dalam memahami dan melayani pelanggan dengan lebih baik. Dengan segmentasi yang tepat, perusahaan dapat meningkatkan loyalitas pelanggan dan meningkatkan profitabilitas. Integrasi teknologi dan analisis data dalam pengambilan keputusan bisnis membuka peluang untuk inovasi dan efisiensi yang lebih tinggi dalam berbagai aspek operasional perusahaan. Akan diketahui metode *Machine Learning* terbaik di antara *Decision Tree*, *Random Forest*, *XGBoost*, *SVM*, dan *Hierarchical Clustering* dalam menentukan segmentasi yang akurat. Analisis berdasarkan hasil segmentasi yang diberikan oleh setiap metode yang digunakan juga akan diberikan. Data yang digunakan merupakan data penjualan produk tersier dari sebuah toko *online* di Indonesia.

1.2 Rumusan Masalah

Berdasarkan penjelasan dalam latar belakang, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut.

1. Bagaimana cara menerapkan metode *Hierarchical Clustering* berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman guna mengetahui karakteristik pelanggan dari sebuah toko *online*?
2. Bagaimana cara menerapkan metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman guna mengetahui karakteristik pelanggan dari sebuah toko *online*?
3. Bagaimana membandingkan performa dan menentukan metode terbaik diantara metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* dalam mengidentifikasi karakteristik pelanggan berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman pada sebuah toko *online*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini dapat dirancang sebagai berikut.

1. Menerapkan metode *Hierarchical Clustering* berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman guna mengetahui karakteristik pelanggan dari sebuah toko *online*.
2. Menerapkan metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman guna mengetahui karakteristik pelanggan dari sebuah toko *online*.
3. Membandingkan performa dan menentukan metode terbaik diantara metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* dalam mengidentifikasi karakteristik pelanggan berdasarkan domisili pelanggan, rentang harga, interval waktu, hari pembelian, kecepatan pengiriman, dan promo gratis biaya pengiriman pada sebuah toko *online*?

1.4 Batasan Masalah

Batasan masalah dan asumsi yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan dalam penelitian ini merupakan data penjualan produk tersier dari sebuah toko *online* yang beroperasi di Indonesia.
2. Data yang digunakan mencakup hasil penjualan per transaksi dari 1 Desember 2023 - 31 Desember 2024.
3. Package R yang digunakan untuk membangun model *clustering* adalah *cluster*.
4. Package R yang digunakan untuk membangun model klasifikasi adalah *rpart*, *randomForest*, *xgboost*, dan *e1071*.
5. Metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-Score* digunakan untuk mengukur akurasi prediksi dari model klasifikasi.

1.5 Manfaat Penelitian

Beberapa manfaat secara teoretis maupun praktis adalah sebagai berikut.

1.5.1 Manfaat Teoretis

1. Menambah wawasan terhadap metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* dalam klasifikasi segmentasi pelanggan guna mengetahui karakteristik pelanggan.
2. Menambah wawasan terhadap metode *Hierarchical Clustering* dalam membagi kluster terhadap segmentasi pelanggan guna mengetahui karakteristik pelanggan.
3. Sebagai bahan kajian untuk studi selanjutnya yang berkaitan dengan metode *Machine Learning* dalam berbagai konteks bisnis lainnya.

1.5.2 Manfaat Praktis

1. Membantu penjual mengetahui strategi pemasaran yang sesuai berdasarkan karakteristik pelanggan.
2. Membantu penjual dalam memahami kebutuhan segmen pasar yang sesuai dengan minat dan kebutuhan pelanggan.

1.6 Sistematika Penulisan

Berikut merupakan penjelasan dari beberapa poin yang akan dibahas pada bab-bab selanjutnya.

1. BAB I - PENDAHULUAN

Pada Bab I akan dijelaskan latar belakang yang mendasari penelitian, cara mengidentifikasi dan merumuskan masalah berdasarkan latar belakang tersebut. Bab ini juga mencakup tujuan penelitian, batasan masalah, manfaat teoretis dan praktis dari penelitian, serta sistematika penulisan skripsi.

2. BAB II - LANDASAN TEORI

Pada Bab II akan dijelaskan notasi serta teori yang digunakan dalam skripsi. Berikutnya akan dibahas teori yang akan dipaparkan yaitu metode *Decision Tree*, *Random Forest*, *XGBoost*, *SVM*, dan *Hierarchical Clustering*. Teori mengenai metrik evaluasi juga akan dibahas dalam bab ini.

3. BAB III - METODE PENELITIAN

Pada Bab III, diuraikan langkah-langkah yang dilakukan dalam menerapkan

metode *Decision Tree*, *Random Forest*, *XGBoost*, *SVM*, dan *Hierarchical Clustering*. Diberikan sebuah diagram alur guna menjelaskan metode penelitian secara umum. Pembahasan akan dijelaskan secara rinci pada setiap subbab di bab ini.

4. **BAB IV - ANALISIS DAN PEMBAHASAN**

Pada Bab IV, disajikan hasil klasifikasi yang diberikan oleh metode *Decision Tree*, *Random Forest*, *XGBoost*, dan *SVM* serta analisis terhadap hasil klasifikasi setiap metode. Hasil klaster yang diberikan oleh metode *Hierarchical Clustering* juga akan diberikan pada bab ini. Perbandingan hasil dari setiap metode akan dibahas secara rinci dalam mengklasifikasi dan klaster segmentasi pelanggan.

5. **BAB V - PENUTUP**

Pada Bab V akan dijelaskan kesimpulan berdasarkan hasil yang didapatkan dari metode yang digunakan. Selain itu, diberikan saran yang dapat digunakan pada penelitian selanjutnya sehingga memberikan hasil yang lebih baik.

