

BAB I

PENDAHULUAN

1.1 Latar Belakang

Adaptasi cepat dan berkelanjutan terhadap perkembangan teknologi merupakan salah satu cara bagi perusahaan untuk tetap dapat bertahan di era pesatnya perkembangan teknologi informasi (Arifin, 2017). Penelitian yang dilakukan oleh *McKinsey Global Institute* pada tahun 2017 menyatakan bahwa perusahaan yang dapat beradaptasi dan mengadopsi teknologi memiliki performa yang lebih baik dan dapat mengungguli pesaingnya (Manyika dkk., 2017).

PT. XYZ adalah perusahaan yang bergerak pada industri minuman sehat di Indonesia sejak 2016. Saat ini PT. XYZ dihadapkan dengan munculnya kompetitor-kompetitor baru, serta pesatnya perkembangan teknologi informasi yang mengakibatkan perusahaan harus tetap dapat mengikuti perkembangan tersebut. Oleh karena itu, perusahaan perlu menjaga agar strategi pemasaran perusahaan tetap relevan dan efektif serta terus berinovasi dan beradaptasi dengan perkembangan pasar sehingga perusahaan dapat tetap bersaing dan unggul (Tidd & Pavitt, 2011). Berdasarkan informasi yang diperoleh dari salah satu *stakeholder* pada PT. XYZ, hingga saat ini perusahaan masih belum memaksimalkan strategi pemasaran, perusahaan masih melakukan promosi yang sama kepada semua konsumen tanpa membedakan mana konsumen yang memiliki potensial untuk melakukan pembelian dan mana konsumen yang tidak berpotensi untuk melakukan pembelian. Hal tersebut dikarenakan perusahaan belum dapat melakukan segmentasi konsumen dengan baik dan secara akurat. Penelitian menyebutkan bahwa memberikan *treatment* dan perlakuan yang berbeda kepada setiap konsumen sesuai dengan karakteristik konsumen adalah hal yang sangat penting dilakukan (Grubor & Jaksa, 2018).

Sebagai salah satu solusi dari permasalahan tersebut, penerapan *data mining* dapat membantu perusahaan untuk mendapatkan informasi strategis sehingga perusahaan dapat melakukan manajemen konsumen dengan baik (Guo & Qin, 2017). Salah satu manfaat *data mining* yaitu dapat digunakan adalah klasterisasi dan klasifikasi konsumen berdasarkan pengelolaan data transaksi konsumen. Berdasarkan

informasi yang diperoleh dari *stakeholder* PT. XYZ hingga saat ini memang belum menggunakan teknik *mining* data dalam mengelola data transaksi perusahaan, dan kegiatan pemasaran pun masih memperlakukan hal yang sama kepada konsumennya.

Dalam penelitian ini dilakukan segmentasi pelanggan dari *dataset* transaksi dan *dataset* konsumen berdasarkan karakteristik dari masing-masing konsumen pada perusahaan, dan menganalisis segmentasi tersebut untuk diketahui masing-masing *cluster* setiap konsumen, konsumen mana yang memiliki potensi pembelian yang bagus ataupun sebaliknya. Teknik dapat digunakan untuk melakukan segmentasi konsumen yaitu dengan teknik *clustering* menggunakan algoritma yang umum seperti Algoritma *K-Means*, *K-Medoids*, *Fuzzy C-Means* (FCM), dan algoritma *Self Organizing Map* (SOM) (Aggarwal & Reddy, 2013). *K-Medoids* merupakan suatu algoritma yang dapat menemukan medoids dalam suatu kelompok data, *K-Medoids* memiliki kelebihan pada prediksi *non-linear*, kemampuan *parallel-processing* yang bagus serta mampu untuk mentoleransi kesalahan, namun memiliki kelemahan pada *training* data yang besar, *overfitting* yakni nilai akurasi pada data *training* tinggi dengan nilai akurasi *data testing* rendah, lambat dalam melakukan konvergensi serta sifatnya yang *local optimum* (Sabzi dkk., 2011). *Self Organizing Map* (SOM) merupakan metode pembelajaran tanpa pengawasan (*unsupervised learning*) yang mengubah bobot interkoneksinya berdasarkan masukan yang diberikan tanpa memerlukan target jawaban (Kohonen, 2001). Kohonen (2001) juga menyebutkan SOM memiliki prediksi dengan akurasi yang tinggi meskipun menggunakan sumber data yang sedikit, akan tetapi SOM memiliki kelemahan prediksi data yang sifatnya fluktuatif seperti pada data transaksi perusahaan. *Fuzzy C-Means* (FCM) adalah suatu algoritma yang menggunakan model *fuzzy* untuk melakukan pengelompokan data, di mana setiap data memiliki tingkat keanggotaan yang berkisar antara 0 hingga 1 terhadap setiap kelas atau *cluster* yang terbentuk. Tingkat keanggotaan ini menunjukkan sejauh mana suatu data menjadi anggota dari masing-masing kelas, kelemahan algoritma *Fuzzy C-Means* (FCM) yaitu memerlukan penentuan jumlah kelas yang tepat dan matriks keanggotaan kelompok awal yang harus ditetapkan

sebelumnya, matriks keanggotaan awal diinisialisasi secara acak, yang dapat menyebabkan masalah inkonsistensi pada metode *Fuzzy C-Means* (Sabzi dkk., 2011).

K-Means merupakan suatu algoritma non hirarki dan bersifat *unsupervised learning* yang dapat mempartisi objek kedalam beberapa *cluster* atau kelompok berdasarkan kemiripan karakteristik masing-masing data (Sinaga & Yang, 2020). *K-Means* memiliki kelemahan yaitu sensitif terhadap penentuan titik pusat (*centroid*) awal (Arbin dkk., 2016), solusinya dapat melakukan iterasi pencarian *centroid* hingga mendapatkan model paling optimal (Sawant & Bhogan, 2016). Algoritma *K-Means* memiliki beberapa keunggulan dibanding algoritma klustering lainnya, diantaranya memiliki akurasi yang tinggi, dapat mengelompokkan data yang besar dengan cepat, dan mudah untuk diimplementasikan (Handayani, 2022). Selain itu, *K-Means* dapat memecahkan masalah yang dimiliki metode FCM, *K-Medoids* dan SOM, yaitu *overfitting*, lambatnya konvergensi, *training* data yang sedikit, serta dapat memecahkan masalah pada data yang memiliki sifat fluktuatif atau data yang memiliki nilai naik turun yang signifikan (Arbin dkk., 2016). Sehingga, penggunaan metode *K-Means* ini sangat relevan untuk karakteristik data perusahaan pada penelitian ini. Herdiana (2021) pada penelitiannya tentang segmentasi klustering data UMKM menggunakan algoritma *K-Means* menyebutkan bahwa *K-Means* berpengaruh positif dalam peningkatan waktu pemrosesan klasifikasi karena *K-Means* mampu mengatasi masalah persebaran data pada *dataset*. Selanjutnya pada penelitian yang membandingkan algoritma *K-Means* dan *K-Medoids* ditemukan *K-Means* lebih baik daripada *k-medoids* dalam waktu eksekusi (Bajracharya, 2020). Sehingga pada penelitian ini peneliti menggunakan algoritma *K-Means* klustering menggunakan metode perhitungan jarak yang memiliki akurasi paling tinggi dan cepat yaitu metode *Euclidean Distance* (Nishom, 2019), serta model RFM sebagai metrik perhitungan untuk menentukan kelompok *cluster*.

Metrik RFM (*Recency*, *Frequency* dan *Monetary*) merupakan model yang menunjukkan kapan terakhir kali konsumen melakukan transaksi (*Recency*), jumlah transaksi yang dilakukan (*Frequency*), dan total nilai transaksi (*Monetary*) dalam kurun waktu yang ditentukan. Model RFM memiliki tingkat akurasi yang tinggi

dengan mengolah data riwayat transaksi konsumen di masa lalu, perusahaan dapat melakukan penilaian terhadap 3 (tiga) aspek dimensi RFM. Model RFM merupakan teknik *mining data clustering* yang efektif untuk mengetahui perilaku dan *pattern* konsumen, segmentasi konsumen berdasarkan model RFM mampu menghasilkan segmentasi dengan tingkat keberhasilan antara 75% hingga 85%, RFM dapat membantu mengidentifikasi pelanggan yang berharga dan mengembangkan strategi pemasaran yang efektif (J.-T. Wei dkk., 2010). Sehingga, pada penelitian ini model RFM dipilih sebagai basis dan atribut dalam melakukan *clustering*.

Dengan mengetahui *cluster* konsumen melalui Algoritma *K-Means* dengan menggunakan model RFM ini diharapkan menjadi solusi permasalahan yang dialami perusahaan saat ini, sehingga perusahaan dapat menentukan strategi pemasaran yang tepat dan terarah sesuai dengan perilaku konsumen di setiap *cluster*, sehingga kegiatan pemasaran yang dilakukan dapat lebih tepat sasaran, efektif dan efisien, sehingga dapat meningkatkan profit perusahaan. Dari uraian tersebut, peneliti tertarik dalam melakukan penelitian yang digunakan sebagai tugas akhir dengan judul **“Analisis Segmentasi Konsumen Menggunakan Algoritma *K-Means Clustering* dan Model RFM Sebagai Rekomendasi Strategi Pemasaran”**.

1.2 Rumusan Masalah

Berdasarkan penjelasan dasar masalah tersebut, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana segmentasi konsumen pada PT. XYZ menggunakan algoritma *K-Means clustering* berdasarkan model RFM?
2. Bagaimana gambaran rekomendasi strategi pemasaran berdasarkan analisis segmentasi konsumen pada PT. XYZ?
3. Bagaimana visualisasi segmentasi konsumen berdasarkan analisis data yang dilakukan pada PT. XYZ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan umum dari penelitian ini adalah sebagai berikut:

1. Menghasilkan segmentasi konsumen dengan algoritma *K-Means clustering* berdasarkan model RFM.
2. Memberikan gambaran rekomendasi strategi pemasaran bagi perusahaan berdasarkan analisis data segmentasi konsumen.
3. Menghasilkan *dashboard* visualisasi segmentasi konsumen berdasarkan analisis data yang dilakukan.

1.4 Manfaat Penelitian

Berdasarkan pemaparan diatas, maka manfaat penelitian ini baik secara teoritis maupun praktis adalah sebagai berikut:

1.4.1 Manfaat Teoritis

Hasil penelitian ini dapat dijadikan sebagai referensi dan acuan dalam penggunaan algoritma *K-Means* untuk melakukan *clustering* berdasarkan model RFM (*Recency, Frequency, Monetary*) dalam membuat segmentasi konsumen melalui penggalian data transaksi dan data konsumen.

1.4.2 Manfaat Praktis

Tugas akhir ini diharapkan dapat memberikan manfaat nyata bagi perusahaan PT. XYZ dalam memberikan gambaran segmentasi konsumen sebagai bentuk masukan dan rekomendasi bagi perusahaan dalam menyusun strategi pemasaran yang terarah sesuai masing-masing segmen konsumen sehingga kegiatan pemasaran lebih efektif dan efisien serta dapat meningkatkan profitabilitas perusahaan.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Tidak membahas mengenai bagaimana cara penggunaan dan pengoperasian *software* Rapid Miner Studio.
2. Penelitian ini hanya membahas salah satu metode *clustering* pada *data mining* yaitu metode *K-Means Clustering*.
3. Penelitian ini menggunakan *dataset* transaksi penjualan dari website perusahaan pada rentang waktu dari 1 Januari 2023 hingga 30 Juni 2023.

1.6 Struktur Organisasi Skripsi

Struktur organisasi skripsi berisi keseluruhan isi skripsi dan pembahasannya. Struktur organisasi skripsi dapat dijabarkan dan dijelaskan dengan sistematika penulisan yang runtun. Struktur organisasi skripsi berisi tentang urutan penulisan dari setiap bab. Struktur organisasi skripsi di mulai dari BAB I sampai BAB V.

BAB I PENDAHULUAN

BAB I berisikan uraian mengenai pendahuluan yang menjelaskan latar belakang, identifikasi masalah, rumusan dan batasan masalah, tujuan penelitian, manfaat penelitian, definisi operasional, dan struktur organisasi skripsi.

BAB II KAJIAN PUSTAKA

BAB II berisikan kajian pustaka dan penelitian terdahulu yang relevan, berisi pembahasan teori-teori dan konsep beserta turunannya dalam bidang yang dikaji.

BAB III METODE PENELITIAN

BAB III membahas mengenai komponen dari metode penelitian. Bab ini berisi tentang metode penelitian, desain penelitian, subjek dan objek penelitian, rancangan pengumpulan data, prosedur penelitian dan rancangan analisis data.

BAB IV HASIL DAN PEMBAHASAN

BAB IV membahas mengenai pencapaian hasil penelitian dan pembahasannya. Hasil penelitian dan pembahasan yang telah dicapai meliputi pengolahan data serta analisis temuan dan pembahasannya.

BAB V SIMPULAN, IMPLIKASI DAN REKOMENDASI

Bab V dalam penelitian ini memuat rangkuman kesimpulan dari analisis dan penelitian yang telah dilakukan, serta memuat implikasi manajerial sebagai rekomendasi yang ditujukan bagi subjek penelitian dan mengulas aspek penting yang dapat diambil dari hasil penelitian tersebut. Selain itu, bab ini juga memberikan saran-saran yang ditujukan untuk penelitian mendatang.