

BAB I

Pendahuluan

1.1 Latar Belakang

Sejak dimulainya era Digital Marketing (DM) dan Data Science (DS) banyak perusahaan yang menggunakan keduanya untuk meningkatkan profit perusahaan. Social media menjadi peran penting untuk kesuksesan dalam pasar yang sudah masuk kedalam digitalisasi. Perkembangan teknologi memberikan dampak yang besar pada data. Dataset yang besar mempersulit pengolahan data secara manual, maka dari itu pengolahan data menggunakan teknologi sebagai sarana pengolahan data. Untuk itu pelaku usaha memerlukan identifikasi secara mendalam menggunakan teknik Artificial Intelligence (AI) atau Machine Learning (ML) (Saura, 2021).

Menurut Asosiasi Penyelenggara Jasa Internet (APJII) pengguna layanan internet Indonesia pada tahun 2021 adalah 202,6 juta pengguna. Informasi tersebut ditanggapi oleh pelaku usaha untuk memaksimalkan kinerja perusahaan mereka dalam digitalisasi. Hal itu berdampak pada layanan digital yang mengalami peningkatan 37% (Inayati et al., 2022).

Teknologi digital telah mengubah komunikasi antara pelaku bisnis dengan konsumennya. Teknologi ini juga menurunkan biaya marketing dalam konvensional marketing dengan menggunakan media sosial. Media sosial merubah pendekatan atau cara mengajak pelaku usaha untuk menarik konsumen. Melakukan CRM yang sudah dikombinasikan dengan teknologi memiliki kesempatan lebih mengenal dan dapat menentukan strategi yang tepat (Nuseir & Refae, 2022).

CRM ini dapat dibagi ke empat langkah *customer identification*, *customer attraction*, *customer retention* dan *customer development*. Identifikasi pelanggan memiliki satu tahap penting yaitu segmentasi pelanggan (Tavakoli et al., 2018a).

Segmentasi pada sekumpulan konsumen merupakan suatu hal yang membantu pelaku usaha untuk dapat melakukan pendekatan kepada konsumen. Segmentasi ini berfungsi membantu pelaku usaha untuk mengambil langkah atau strategi yang tepat dalam pendekatan konsumen sesuai dengan ciri-ciri atau sifat konsumen tersebut (Zhou et al., 2021).

Segmentasi perilaku merupakan salah satu jenis segmentasi yang sering dilakukan. Segmentasi ini berfokus pada cara pelanggan berinteraksi dengan produk yang diberikan. Hasil segmentasi ini membantu perusahaan untuk menjaga konsumen agar tetap memakai produk yang pelaku usaha berikan (Kadir & Achyar, 2019).

Tidak luput juga badan amil zakat untuk mengumpulkan donasi yang sudah bergerak dalam digitalisasi. Pengelolaan data transaksi dan data donatur akan dilakukan segmentasi. Segmentasi ini bertujuan untuk mengelompokkan donatur kedalam beberapa kelompok. Kelompok ini didasarkan kesamaan perilaku donaturnya. Teknik pengelompokan ini dilakukan dalam clustering (Fang & Liu, 2021).

Salah satu segmentasi yang sering digunakan adalah metode Recency Frequency Monetary (RFM) analisis. Metode ini berfokus pada kebiasaan seorang konsumen, karena kebiasaan konsumen yang berubah-ubah setiap saat. Jika perusahaan memiliki suatu produk baru yang akan berdampak dengan pola pembelian. Model RFM difokuskan pada kebiasaan konsumen dan relasi antara *Frequency* dan *Monetary* (Tavakoli et al., 2018b).

Dalam melakukan RFM analisis dilakukan dengan tiga tahap. Tahap pertama melakukan ekstraksi dataset ke bentuk RFM. Tahap kedua melakukan segmentasi dengan clustering. Tahap ketiga melakukan analisis pada hasil cluster (Bachtiar, 2018).

Mendapatkan hasil clustering dengan memisahkan atau membagi data ke dalam beberapa cluster yang sudah ditentukan dan hasilnya akan dievaluasi (Golalipour et al., 2021). Tujuan utama dari clustering antara lain: (1) untuk mendapatkan pemahaman yang berguna dari data (2) mengidentifikasi tingkat kesamaan (3) Untuk mengatur data dan mendapatkan pemahaman tentang data melalui analisis clustering (Ghosal et al., 2020).

Teknik clustering lebih sulit daripada klasifikasi (*supervised*) karena tidak mengetahui label dari dataset tersebut. Label pada dataset supervised menentukan instance tersebut termasuk kelompok mana. Clustering akan menjadi sulit untuk menentukan kelompoknya karena ada beberapa parameter yang menentukan kelompok clustering. Semakin banyak *feature* (kolom) dan banyak data meningkatkan *cost computation* dan memberikan effect konsistensi algoritma pada data (Saxena et al., 2017).

Dua metode yang sering digunakan dalam melakukan clustering k-means dan hierarchical agglomerative clustering. Hierarchical clustering dapat mendukung macam-macam informasi tetapi pengembangan yang harus efisien karena algoritma untuk komputasinya yang begitu penting pada beberapa bagian (Cohen-addad, Kanade, Mallmann-trenn, & Mathieu, 2019).

Sebenarnya metode hierarchical clustering terbagi menjadi dua agglomerative dan divisive. Perbedaanya hanya pada jika agglomerative mengumpulkan data ke dalam suatu group dan divisive membagi data ke dalam beberapa group. *Ward's method* merupakan salah satu algoritma agglomerative. Algoritma yang memiliki akurasi yang tinggi tetapi sangat sensitif terhadap pencilan. Banyak simulasi yang sudah dilakukan pada

beberapa dataset dan *Ward's method* memiliki performa yang signifikan dari yang lainnya (Eszergár-Kiss & Caesar, 2017).

Time complexity dan *space complexity* dalam hierarchical clustering yang cukup tinggi tidak cukup baik pada dataset yang berukuran besar. *Time complexity* yang dimiliki hierarchical clustering sebanding dengan $O(N^3)$ dan *space complexity* sebanding dengan $O(N^2)$, nilai N merupakan banyak record (Purnama, 2019).

Badan amil zakat yang biasanya melakukan pengumpulan dana dengan cara konvensional sekarang dapat dilakukan secara digital. Kelebihan digital ini adalah orang-orang dengan mudah melakukan zakat, infaq dan sebagainya. Data transaksi dan data donatur menjadi sangat penting untuk dikelola lebih lanjut. Optimasi hasil dari data tersebut digunakan untuk memaksimalkan kinerja badan amil zakat. Dataset yang digunakan data transaksi 277.442 record dan 10 kolom. Kegiatan ini akan menambahkan bantuan pada orang-orang yang memang membutuhkan bantuan. Pengelolaan data yang tepat akan menemukan pemahaman dan pembentukan pengetahuan baru tentang suatu kejadian peristiwa (Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021).

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas penulis mengidentifikasi masalah menjadi beberapa poin sebagai berikut:

- Dataset yang besar untuk HAC yang memiliki time complexity dan space complexity yang tinggi.
- Pengaruh pencilan untuk segmentasi.
- Karakteristik untuk setiap cluster pada setiap segmentasinya.

1.3 Batasan Masalah

Berdasarkan identifikasi masalah di atas maka tidak semua masalah dapat penulis atasi karena keterbatasan penulis dalam segi biaya, waktu dan

kemampuan. Dengan demikian penulis membatasi masalah hanya dalam melakukan segmentasi metode RFM yang dikombinasikan dengan *Agglomerative Ward's Method*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah untuk mengetahui:

1. Proses *data mining* sampai menghasilkan output yang diinginkan.
2. Pengaruh dataset pada algoritma HAC.
3. Banyak cluster yang dibuat pada segmentasi.
4. Mengetahui karakteristik pada setiap cluster

1.5 Kontribusi

Dalam segi bisnis hasil penelitian ini memiliki manfaat dalam segi marketing. Bagaimana pelaku usaha melakukan pendekatan kepada konsumennya dan cara menentukan strategi yang tepat dalam melakukan CRM. Untuk kelanjutannya dapat melakukan evaluasi kebiasaan konsumen dalam melakukan transaksi.

Dalam segi akademis penelitian ini membantu di dalam melakukan tahapan atau proses melakukan segmentasi. Menentukan apakah metode ini cocok atau tidak dalam melakukan permasalahan yang dikerjakan dan sebagai informasi pembelajaran.