

**Optimasi *Cluster* pada K-Means menggunakan DBSCAN untuk Segmentasi
Pelanggan**

TESIS

Oleh

Ihsan Satriawan

23513008



**PROGRAM STUDI MAGISTER INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG**

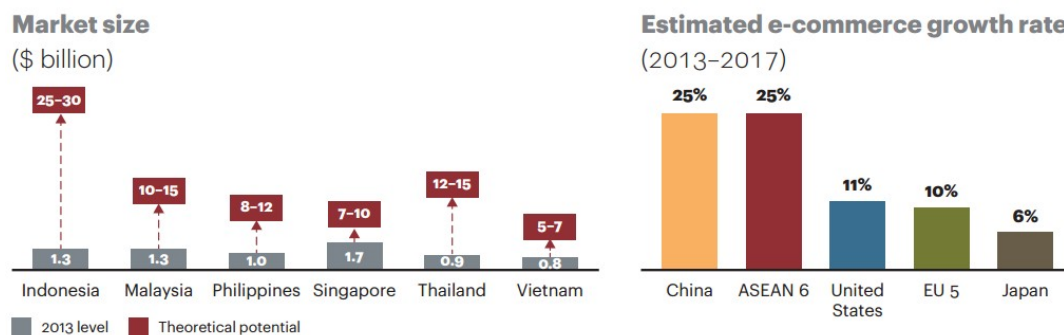
2016

BAB I

Pendahuluan

I.1. Latar Belakang

E-commerce merupakan salah satu sektor perdagangan yang sedang berkembang pesat ditengah gencarnya penetrasi internet dimasyarakat Indonesia. Gambar 1.1, menunjukkan bahwa Indonesia memiliki potensi pasar yang besar dibandingkan negara ASEAN lainnya dengan pesatnya penetrasi internet serta cukup besarnya nilai pasar *e-commerce* di Indonesia, tidak aneh jika bermunculan pemain-pemain disektor ini seperti Hijup, Bukalapak, Tokopedia, Kaskus, Elevenia, MatahariMall, dll. Dengan banyaknya pemain disektor ini. mengakibatkan ketat nya persaingan yang terjadi antar pemain. Dengan kondisi persaingan yang begitu kompetitif membuat perusahaan menyadari betapa pentingnya peran pelanggan, karena pelanggan yang menjadi alasan tetap bertahannya suatu perusahaan, oleh karena itu banyak perusahaan yang mempertahankan pelanggan melalui program pengembangan segmentasi pelanggan.



Gambar 1.1. ASEAN market potencial (sumber : A.T. Kearney analysis)

CRM (*Customer Relationship Management*) merupakan strategi untuk membangun, mengelola, serta memperkuat hubungan perusahaan dengan pelanggan yang loyal agar dapat tercipta hubungan yang tahan lama. CRM harus menggunakan pendekatan berpusat pada pelanggan (*customer-centric*). Hal tersebut dilakukan untuk

mengidentifikasi dan memahami perbedaan kebutuhan, referensi, serta perilaku dari setiap pelanggan (Tsiptsis dan Chorianopoulos, 2009). Pemanfaatan *data mining* dalam kerangka kerja CRM dapat memberikan pandangan dari sisi pelanggan yang merupakan elemen penting dalam membangun CRM yang efektif. Tsiptsis dan Chorianopoulos. (2009) mengatakan bahwa segmentasi pelanggan merupakan salah satu aktivitas CRM yang dapat didukung dengan *data mining*. *Clustering* merupakan salah satu teknik *data mining* yang bisa digunakan dalam proses segmentasi pelanggan (Tsiptsis dan Chorianopoulos, 2009).

Algoritma K-Means merupakan salah satu algoritma untuk melakukan *clustering* segmentasi pelanggan yang paling populer digunakan karena memiliki kelebihan mudah di implementasikan serta memiliki kompleksitas metode yang cukup efisien $O(nkl)$ (Aggarwal dan Aggarwal, 2012; Bunnak dkk, 2015; Hajiha dkk, 2011; Huang dan Song, 2014). Kekurangan yang terdapat dalam algoritma K-Means yakni mempunyai masalah dalam penentuan awal jumlah *cluster* serta masalah dalam penentuan nilai awal pusat *cluster* (*centroid*) (Zhang dan Fang, 2013). Pada kenyataannya, penentuan awal jumlah serta nilai awal pusat *cluster* sulit, hal ini dikarenakan penentuan awal jumlah serta nilai awal pusat *cluster* yang berbeda dapat menghasilkan *cluster* yang berbeda.

Beberapa penelitian yang telah dilakukan untuk menangani hal tersebut antara lain penelitian yang dilakukan oleh (Zhang dan Fang, 2013) mengoptimalkan inisialisasi pusat cluster awal dengan cara menemukan satu set data yang mencerminkan karakteristik distribusi data sebagai pusat awal cluster untuk mendukung pembagian data ke batas yang terbaik. (Xu dkk, 2015) menggunakan hierarchical clustering sebelum memasukkan ke dalam k-means. Level iterasi yang dipilih peneliti pada tahap hierarchical clustering menjadi jumlah k cluster pada k-means, serta rata-rata point setiap cluster pada hierarchical clustering menjadi nilai pusat awal (*centroid*) pada k-means. Keterbatasan solusi yang diusulkan oleh (Zhang dan Fang, 2013) adalah masih perlu nya untuk menentukan jumlah k cluster yang akan dibentuk, sedangkan solusi yang diusulkan oleh (Xu dkk, 2015) perlu nya menentukan sampai iterasi seberapa

dalam proses hierarchical clustering dalam menentukan jumlah k cluster yang akan digunakan. Oleh karena itu diperlukan pendekatan lain untuk dapat mengoptimasi *cluster* pada K-Means dengan penentuan jumlah k cluster serta *centroid* nya bersamaan.

DBSCAN merupakan salah satu metode *clustering* berdasarkan kepadatan atau yang biasa disebut *density based clustering* dan dapat mendeteksi data yang menyimpang (*noise*). Metode DBSCAN dapat mengelompokkan kumpulan data tanpa perlu mengatur jumlah *cluster* yang akan dibentuk, namun perlu mengatur nilai *minpts* yang merupakan jumlah minimal objek data di dalam suatu *cluster* dan nilai *epsilon* yang merupakan nilai untuk jarak antar objek data (Hu dkk, 2014). Berdasarkan hal tersebut maka DBSCAN dapat menghasilkan jumlah *cluster* sesuai dengan nilai *minpts* dan nilai *epsilon* sehingga dapat dimanfaatkan untuk mengoptimasi *cluster* pada K-Means.

Penelitian ini mengusulkan sebuah optimasi *cluster* pada K-Means untuk *customer segment*. Optimasi yang diusulkan dengan memanfaatkan keluaran dari DBSCAN untuk masukan K-Means berupa jumlah *cluster* serta nilai *centroid*.

I.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka permasalahan utama dalam penelitian dapat dirumuskan dalam *research question* sebagai berikut :

1. Bagaimana teknik *clustering* dengan DBSCAN dapat digunakan untuk mengoptimasi *cluster* pada K-Means ?
2. Bagaimana hasil atau kinerja teknik *clustering* dengan optimasi *cluster* pada K-Means dalam melakukan proses segmentasi pelanggan ?

I.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui hasil dan kinerja K-Means setelah dilakukan optimasi *cluster* menggunakan DBSCAN dalam melakukan proses segmentasi pelanggan.

I.4. Batasan Masalah

Dalam upaya agar penelitian ini tetap fokus pada tujuan utamanya maka perlu ditambahkan batasan persoalan yang ingin diselesaikan. Berikut adalah batasan yang ditetapkan dalam penelitian ini:

1. Proses *clustering* menggunakan metode DBSCAN dan K-Means dan model RFM
2. Data yang digunakan pada proses segmentasi pelanggan menggunakan data transaksi pembelian pada salah satu *e-commerce* di Indonesia diatas tahun 2015

I.5. Metodologi

Metodologi penelitian ini adalah eksperimen dengan tahapan pengerjaan sebagai berikut :

1. Studi Pustaka

Tahap studi pustaka bertujuan untuk memperkaya pemahaman mengenai konsep dari hal-hal yang berkaitan dengan *Customer Relationship Management*, segmentasi pelanggan, serta *clustering* untuk segmentasi pelanggan.

2. Pengumpulan Data

Data yang digunakan merupakan data yang dapat merepresentasikan transaksi pelanggan sehingga dapat ditransformasikan ke dalam model RFM.

3. Praproses Data

Tahapan praproses data menggunakan *query* untuk mendapatkan nilai RFM dari setiap pelanggan lalu mentransformasikan nilai RFM tersebut ke dalam skala yang ditentukan.

4. Implementasi Metode

a) Melakukan pembentukan *cluster* menggunakan DBSCAN

b) Melakukan segmentasi pelanggan menggunakan K-Means yang sudah dioptimasi

5. Analisis Hasil dan Kesimpulan

(a) Analisis pembentukan *cluster* menggunakan DBSCAN

(b) Analisis pembentukan segmentasi pelanggan menggunakan K-Means yang sudah dioptimasi

6. Pembuatan Laporan

Penulisan laporan penelitian berdasarkan sumber referensi yang menjadi teori dasar penelitian dan hasil penelitian berupa tahap-tahap dari setiap kegiatan yang dilakukan.

I.6. Sistematika Penulisan

Penulisan laporan tesis ini dilakukan secara sistematika dengan dibagi ke dalam beberapa bab seperti di bawah ini :

1. Bab I Pendahuluan

Bab ini menjelaskan tentang beberapa hal, yakni : latar belakang tesis, rumusan masalah, tujuan penelitian, batasan masalah, metode penelitian dan sistematika penulisan

2. Bab II Tinjauan Pustaka

Bab studi literatur menjelaskan beberapa teori atau konsep dasar yang mendukung pengembangan tesis, yakni konsep CRM, segmentasi pelanggan K-Means Clustering, DBSCAN, serta tentang penelitian sebelumnya yang berkaitan.

3. Bab III Perancangan dan Implementasi

Bab ini menjelaskan tentang perancangan dan implementasi dari optimasi *cluster* pada K-Means menggunakan DBSCAN untuk segmentasi pelanggan

4. Bab IV Hasil dan Analisis

Bab ini menjelaskan hasil dan analisis dari pembentukan graf dan pengaruh perhitungan kualitas modularity MC pada algoritma Girvan-Newman terhadap hasil deteksi komunitas berdasarkan skenario pengujian untuk dianalisis perbandingan hasilnya dengan modularity NG.

Bab ini menjelaskan hasil dan analisis dari pengaruh DBSCAN dalam optimasi *cluster* pada K-Means terhadap hasil segmentasi pelanggan berdasarkan skenario pengujian untuk dianalisis perbandingan hasilnya dengan segmentasi pelanggan menggunakan K-Means biasa.

5. Bab V Kesimpulan dan Saran

Pada bagian ini akan dijelaskan kesimpulan dari hasil penelitian yang telah dilakukan, serta saran untuk melakukan pengembangan dari hasil penelitian yang di dapat.

DAFTAR REFERENSI

- Aggarwal, N., Aggarwal, Ki., & Gupta, K. (n.d.). Comparative Analysis of k-means and Enhanced K-means clustering algorithm for data mining.pdf.
- Bunnak, P., Thammaboosadee, S., & Kiattisin, S. (2015). Applying Data Mining Techniques and Extended RFM Model in Customer Loyalty Measurement. *Journal of Advances in Information Technology*, 238–242. <http://doi.org/10.12720/jait.6.4.238-242>
- Hajiha, A., Radfar, R., & Malayeri, S. S. (2011). Data mining application for customer segmentation based on loyalty: An Iranian food industry case study. In *Industrial Engineering and Engineering Management (IEEM), 2011 IEEE International Conference on* (pp. 504–508). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6117968
- Huang, X., & Song, Z. (2014). Clustering Analysis on E-commerce Transaction Based on K-means Clustering. *Journal of Networks*, 9(2). <http://doi.org/10.4304/jnw.9.2.443-450>
- Hu, X., Zhang, H., Wu, X., Chen, J., Xiao, Y., Xue, Y., ... Zhao, H. (2014). A New Customer Segmentation Framework Based on Biclustering Analysis. *Journal of Software*, 9(6). <http://doi.org/10.4304/jsw.9.6.1359-1366>
- Tsiptsis, K., & Chorianopoulos, A. (2009). *Data mining techniques in CRM: inside customer segmentation*. Chichester, West Sussex, U.K: Wiley.
- Xu, T.-S., Chiang, H.-D., Liu, G.-Y., & Tan, C.-W. (2015). Hierarchical K-means Method for Clustering Large-Scale Advanced Metering Infrastructure Data. *IEEE Transactions on Power Delivery*, 1–1. <http://doi.org/10.1109/TPWRD.2015.2479941>
- Zhang, C., & Fang, Z. (2013). An Improved K-means Clustering Algorithm.pdf.