

# **Bab I   Pendahuluan**

## **I.1   Latar Belakang**

Penetrasi internet di Indonesia pada saat ini cukup pesat. Berdasarkan survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pada tahun 2018 terdapat 171,17 juta masyarakat Indonesia merupakan pengguna internet (APJII, 2019). Artinya terdapat 64,8 persen dari keseluruhan masyarakat Indonesia telah terkoneksi internet. Hal ini merupakan angka yang besar karena terjadi kenaikan 10,12 persen pengguna internet dari tahun 2017. Apabila dibandingkan dengan pertumbuhan penduduk yang rata-rata tumbuh 1,34 persen per tahun, persentase penetrasi internet di Indonesia merupakan angka yang fantastis (Badan Pusat Statistik, 2018). Pertumbuhan angka pengguna internet juga diikuti dengan pertumbuhan penggunaan operator seluler. Sejalan dengan survei yang dilakukan APJII pada tahun 2018, sebanyak 96,6 persen pengguna terkoneksi dengan internet melalui paket data/kuota dari operator seluler di tahun 2019 (APJII, 2019).

Penggunaan internet di Indonesia didominasi oleh pelanggan di Pulau Jawa. Persentase pengguna internet dari Pulau Jawa sebesar 55% dari total pengguna internet di Indonesia (Eril, 2019). Angka tersebut merupakan angka yang sangat besar. Dari beberapa daerah yang ada di Pulau Jawa, Provinsi Daerah Khusus Ibukota Jakarta (DKI Jakarta) memiliki persentase pengguna internet per jumlah penduduk yang sangat besar. DKI Jakarta memiliki penduduk 11.063.324 jiwa berdasarkan statistik pada tahun 2019 (BPS, 2019). Di samping itu, pengguna internet di DKI Jakarta pada tahun 2019 sebesar 8,9 Juta pengguna (Katadata, 2019). Artinya persentase pengguna internet di DKI Jakarta mencapai lebih dari 80 persen. Angka ini cukup besar dibandingkan dengan luas wilayah DKI Jakarta yang hanya 662,33 km<sup>2</sup> (BPS, 2019).

Penyedia layanan internet berbasis jaringan seluler telepon, PT. Telekomunikasi Selular (Telkomsel) telah berdiri sejak tahun 1995. Menurut laporan detik.com, pada Mei 2019 jumlah pelanggan Telkomsel mencapai 168 juta pelanggan dan telah didukung oleh 197.000 infrastruktur BTS (Suryo, 2019). Dengan angka yang sangat besar tersebut, Telkomsel seluler harus memikirkan secara matang strategi

pemasaran yang efektif. Istilah pemasaran didefinisikan oleh Kotler sebagai proses perusahaan menciptakan nilai bagi pelanggan dan membangun hubungan pelanggan yang kuat untuk menangkap nilai dari pelanggan sebagai imbalan (Kotler, 2011). Pemasaran ini bertujuan untuk menjaga pelanggan Telkomsel agar tetap setia dalam menggunakan jasa Telkomsel. Pelanggan yang dimaksud adalah pelanggan yang berada di wilayah DKI Jakarta dikarenakan persentase pengguna internet yang sangat besar.

Sebagai anak perusahaan dari PT. Telkom Indonesia, Telkomsel melayani masyarakat dengan berbagai jenis produk seluler yang ditawarkan. Beberapa jenis produk seluler yang ditawarkan adalah Simpati, Halo, ByU, AS dan LOOP. Setiap produk yang ditawarkan oleh Telkomsel memiliki peminat dan pelanggannya masing-masing. Di dalam setiap produk pengguna memiliki karakteristik dan kebiasaan yang berbeda-beda. Hal ini menjadi masalah baru ketika Telkomsel ingin memahami pelanggannya. Pemahaman pelanggan ini menjadi kunci dalam mengetahui preferensi pelanggan. Preferensi pelanggan sering kali dilakukan dengan cara segmentasi / pengelompokan (Zhao et al., 2010). Berdasarkan preferensi tersebut internal dari Telkomsel dapat melakukan berbagai hal antara lain dapat mengetahui posisi produk di pasar sehingga dapat melakukan penawaran produk secara efektif (Kashwan & Velu, 2013). Dengan karakteristik dan jumlah pelanggan yang sangat besar dan beragam, sulit dilakukan segmentasi secara manual. Secara tradisional, perusahaan yang bergerak di bidang telekomunikasi menggunakan atribut satu dimensi untuk segmentasi pelanggan yaitu kontribusi terhadap pendapatan perusahaan (Zhao et al., 2010).

Berdasarkan laporan kanal berita CNN Indonesia, pesaing terdekat dari Telkomsel datang dari perusahaan XL Axiata. Perusahaan ini mencatatkan penetrasi pengguna ponsel pintar, yang merupakan pelanggan XL, naik sebesar 86 persen pada tahun 2019 (Cnnindonesia.com, 2020). Telkomsel hanya mencatatkan penetrasi pengguna ponsel pintar sebanyak 66 persen. Selain itu, XL Axiata juga mengklaim bahwa 89 persen pendapatan layanan berasal dari pendapatan data. Pendapatan data yang dimaksud adalah pendapatan hasil dari pembelian paket data oleh pelanggan atau pembayaran layanan pasca bayar. Berlawanan dengan klaim tersebut, perusahaan Telkomsel justru hanya mencatat 56 persen pendapatan layanan berasal

dari pendapatan data. Angka ini cukup jauh dan membahayakan posisi Telkomsel sebagai operator penyedia layanan internet di Indonesia.

Usaha yang dapat dilakukan Telkomsel untuk mempertahankan pasar mereka adalah segmentasi pelanggan. Segmentasi dapat membantu Telkomsel dalam melakukan kustomisasi penawaran layanan dan promo dari pihak pemasaran Telkomsel. Hal ini sejalan dengan pernyataan Morteza Namvar, beberapa keuntungan segmentasi pelanggan adalah: (1) meningkatkan retensi dan loyalitas pelanggan, (2) kustomisasi produk dan layanan, (3) pembentukan nilai unik kepada pelanggan (Namvar et al., 2010). Hasil dari segmentasi ini diharapkan mampu membantu Telkomsel dalam perumusan layanan dan promo sehingga dapat mempertahankan pangsa pasar Telkomsel.

Seiring perkembangan teknologi dan ilmu pengetahuan, segmentasi pelanggan dapat dilakukan secara otomatis melalui teknik penambangan data. Penambangan data atau biasa disebut *data mining* adalah proses menemukan pola di dalam kumpulan data (Albert Verasius, 2019). Penerapan penambangan data di dalam dunia pemasaran khususnya segmentasi pelanggan digunakan dalam banyak literatur (Kashwan & Velu, 2013; Namvar et al., 2010; Wu & Lin, 2005; Zhao et al., 2010). Salah satu teknik penambangan data adalah pembelajaran mesin *unsupervised learning*. Teknik ini digunakan dalam data yang tidak memiliki variabel terikat sebagai target. Dalam prosesnya, teknik *unsupervised learning* memungkinkan untuk menemukan pola di dalam data tanpa adanya supervisi dari manusia yang berupa target. Salah satu jenis dari teknik *unsupervised learning* adalah *clustering*. Dengan *clustering* dapat memungkinkan manusia untuk mendapat pola dengan cara pengelompokan data dengan algoritma tertentu. *Clustering* dibagi menjadi dua hal yaitu hierarki *clustering* dan non-hierarki *clustering* (Khairani, 2019).

Algoritma yang umum digunakan untuk melakukan *data mining* sebagai upaya segmentasi pelanggan dengan *clustering* adalah *K-Means*. *K-Means* merupakan algoritma *clustering* non-hierarki yang cocok digunakan untuk permasalahan dalam pembentukan kelompok dari data-data yang tersedia (Khairani, 2019). Algoritma ini akan mengelompokkan data berdasarkan kemiripan yang dihitung melalui jarak

(*distance*). Perhitungan jarak antar data akan menghasilkan nilai jarak *inter*-klaster dan *intra*-klaster. *Inter*-klaster adalah jarak antar dua klaster yang berbeda. Sedangkan *intra*-klaster adalah jarak antar dua data pada klaster yang sama (Krishna & Murty, 1999). Dengan meminimalkan rata-rata jarak *intra*-klaster dan memaksimalkan rata-rata jarak *inter*-klaster maka akan diperoleh hasil *clustering* yang baik.

Terdapat kekurangan dari algoritma *K-Means* sebagai algoritma yang umum digunakan untuk *clustering*. Kekurangan terbesar adalah algoritma ini secara umum tidak didesain untuk menerima data latih bertipe data kategori. Padahal dalam dunia nyata banyak sekali permasalahan yang harus diselesaikan dengan menggunakan *clustering* namun data memiliki tipe data campuran (Huang, 1998). Penelitian yang dihasilkan oleh Zhexue Huang pada tahun 1998 memberikan alternatif solusi untuk kekurangan ini. Huang memodifikasi algoritma *k-means* dengan memberikan rumus baru agar algoritma dapat menerima data latih berupa kategori. Untuk data latih yang bertipe data kategori, Huang memberikan opsi algoritma *K-modes*. Namun untuk data latih bertipe data campuran (numerik dan kategori), Huang memberikan opsi algoritma *K-prototypes*. Kedua algoritma ini sejatinya menggunakan prinsip yang sama dengan *K-means*. Perbedaan kedua algoritma ini dengan *K-means* hanya terletak pada rumus fungsi biaya untuk penentuan *centroid*.

Pada dasarnya *K-means*, *K-modes* dan *K-prototypes* membutuhkan jumlah klaster sebagai parameter awal agar algoritma dapat melakukan perhitungan klaster. Parameter ini dapat menjadi hambatan apabila data yang dimiliki sangat besar dan tim bisnis belum mengetahui berapa banyak segmen yang harus dihasilkan dalam proses segmentasi. Maka dari itu, perlu ada metode tambahan sebagai cara mengatasi keterbatasan ini. Salah satu metode yang umum digunakan antara lain; metode *elbow*, *silhouette score* dan *davies bouldin index*. Ketiga metode ini digunakan sebagai metrik evaluasi untuk memperoleh jumlah klaster terbaik (Xu & Tian, 2015).

Metode *elbow* merupakan metode untuk mengetahui nilai optimal dari jumlah klaster yang ingin dibentuk melalui algoritma clustering seperti *K-means*, *K-modes* dan *K-prototypes* (Yuan & Yang, 2019). Metode ini dirancang untuk mengetahui

konsistensi atas klaster yang terbentuk. Konsistensi diukur melalui nilai fungsi (*cost function*) yang dihasilkan melalui eksperimen dengan jumlah klaster atau nilai  $k$  yang berbeda (Liu & Deng, 2020). Nilai  $k$  yang meningkat akan menghasilkan rata-rata distorsi yang berkurang. Namun, penurunan dari nilai rata-rata distorsi akan melambat pada suatu titik. Titik tersebut disebut dengan titik optimal dari nilai  $k$ . Sehingga, penurunan nilai rata-rata distorsi yang curam disebut dengan siku (*elbow*). Metode ini banyak digunakan pada penelitian yang menggunakan data cukup besar, sehingga penentuan nilai  $k$  sulit dilakukan secara manual. Hasil dari *clustering* dengan nilai  $k$  yang optimal akan dilakukan analisis untuk interpretasi makna tiap klaster.

Penelitian ini tidak terlepas dari penelitian-penelitian sebelumnya. Beberapa penelitian telah dilakukan untuk mengatasi persoalan segmentasi pelanggan dengan *clustering* namun dengan studi kasus dan data yang berbeda. Penelitian serupa berjudul “*Telecom Customer Segmentation with K-Means Clustering*” merupakan prosiding dalam konferensi di Melbourne yang ditulis oleh Luo Ye (Ye et al., 2012). Perbedaan penelitian ini dengan penelitian yang dilakukan oleh penulis adalah jenis data dan cara pemrosesan data. Pada prosiding yang ditulis Luo Ye, data yang digunakan adalah data historis penggunaan internet pada perusahaan telekomunikasi. Data ini diproses dengan metode analisa *Recency, Frequency, Monetary* (RFM *analysis*). Sedangkan penulis menggunakan data karakteristik pelanggan perusahaan Telkomsel seperti; tipe gawai, merek gawai, sistem operasi yang digunakan dan beberapa data lainnya. Data karakteristik pelanggan ini merupakan data dengan tipe kategori sehingga algoritma yang digunakan adalah ekstensi *K-means* yaitu K-Prototype. Penelitian lain yang serupa dengan penelitian yang dilakukan penulis adalah “Peta Persaingan Produk Kartu Seluler Berdasarkan Segmen Konsumen” ditulis oleh Nur Indrianti menjelaskan tentang persaingan penyedia layanan internet di Indonesia (Indrianti, 2013). Penelitian ini menggunakan objek seluruh penyedia operator seluler seperti Telkomsel, Indosat dan Flexi untuk menghasilkan analisa persaingan. Hal ini berbeda dengan penelitian yang dilakukan penulis yang hanya berfokus pada Telkomsel dengan permasalahan pemasaran produk yang dilakukan oleh Telkomsel kepada pelanggannya.

Berdasarkan penelitian yang sudah ada serta permasalahan yang dialami oleh Telkomsel, maka terdapat urgensi untuk melakukan penelitian terkait segmentasi pelanggan di perusahaan Telkomsel khususnya wilayah DKI Jakarta. Pemilihan wilayah DKI Jakarta mempertimbangkan rasio pengguna internet per jumlah penduduk di wilayah tersebut yang besar. Dikarenakan belum ada penelitian yang spesifik membahas segmentasi pelanggan Telkomsel di DKI Jakarta dengan menggunakan data karakteristik pelanggan maka penelitian ini dirasa perlu untuk segera dilakukan. Selain itu, penelitian yang menggunakan metode elbow dan algoritma *K-means* dengan objek penelitian data telekomunikasi masih sedikit. Dengan sebab-sebab yang telah disebutkan, penulis merancang dan membuat penelitian yang berjudul “Segmentasi Pelanggan PT. Telekomunikasi Seluler Indonesia Menggunakan *Clustering* Algoritma *K-prototypes* Dan Metode *Elbow*”. Alasan penggunaan algoritma *k-prototypes* adalah data yang digunakan penulis berbentuk tabel dengan nilai kategori dan numerik. Sehingga *k-prototypes* cocok untuk diimplementasikan pada data seperti ini. Selain itu, metode *elbow* juga termasuk metode yang tergolong populer untuk menemukan nilai parameter kluster yang tepat pada algoritma *k-prototype*.

## **I.2 Rumusan Masalah**

Rumusan masalah pada penelitian ini dapat dibagi menjadi dua yaitu:

1. Bagaimana penerapan *clustering* dengan algoritma *k-prototypes* sebagai cara segmentasi pelanggan Telkomsel?
2. Bagaimana penerapan metode *elbow* pada algoritma *k-prototypes* untuk menentukan nilai kluster terbaik?
3. Bagaimana interpretasi tiap kluster terhadap data asli pelanggan Telkomsel?

## **I.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan penelitian ini dapat dijabarkan sebagai berikut:

1. Menerapkan *clustering* dengan algoritma *k-prototypes* untuk segmentasi pelanggan Telkomsel.
2. Mendapatkan nilai kluster terbaik dengan menerapkan metode *elbow* pada algoritma *k-prototypes*.

3. Mendapatkan interpretasi tiap klaster terhadap data pelanggan Telkomsel yang dapat digunakan sebagai landasan pemasaran.

#### **I.4 Manfaat Penelitian**

Manfaat dari penelitian ini dapat dibagi menjadi tiga aspek sebagai berikut:

##### **A. Manfaat bagi keilmuan**

1. Menjadi tambahan literasi dalam menambah khazanah pengetahuan dalam rumpun ilmu Sistem Informasi khususnya di bidang *data mining* dan *machine learning*.
2. Menjadi referensi rujukan untuk penelitian di bidang segmentasi pelanggan dengan *clustering* di masa yang akan datang.

##### **B. Manfaat bagi perusahaan**

1. Dapat menghasilkan model *clustering* menggunakan algoritma *k-prototypes* dengan jumlah klaster yang optimal melalui evaluasi metode *elbow*. Model ini akan digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik yang dimiliki sehingga dapat digunakan untuk merumuskan strategi bisnis Telkomsel.

##### **C. Manfaat bagi peneliti**

1. Mendapatkan pengalaman di bidang pengolahan data khususnya pengolahan data telekomunikasi milik PT. Telekomunikasi Seluler Indonesia dengan menggunakan metode *clustering* untuk menyelesaikan permasalahan nyata bisnis.

#### **I.5 Batasan Penelitian**

Batasan dari penelitian ini adalah sebagai berikut:

1. Penelitian hanya dilakukan pada sampel data pengguna Telkomsel di Daerah Khusus Ibukota Jakarta yang diambil pada tanggal 19 Juli 2020.
2. Penelitian ini menggunakan percobaan sebanyak 15 kali dengan jumlah klaster dimulai dari 1 dan diakhiri dengan 15.
3. Data pelanggan Telkomsel yang bersifat konfidensial dilakukan penyamaran data (*masking*).

## **I.6 Sistematika Penulisan**

Sistematika penulisan dalam penelitian ini dapat dibagi menjadi enam bab. Pembagian tersebut dapat dijabarkan sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini berisi penjelasan mengenai latar belakang penelitian, rumusan masalah sebagai rangkuman permasalahan, tujuan penelitian, manfaat penelitian, batasan masalah, rangkuman metodologi penelitian dan sistematika penulisan.

### **BAB II KAJIAN TEORI**

Bab ini berisi teori-teori yang berkaitan dengan permasalahan yang akan diteliti. Teori yang ada digunakan untuk menunjang pelaksanaan penelitian.

### **BAB III METODOLOGI PENELITIAN**

Bab ini berisi penjelasan langkah-langkah penelitian secara rinci meliputi metode dan model konseptual yang digunakan pada penelitian.

### **BAB IV ANALISIS DAN PERANCANGAN PENGOLAHAN DATA**

Bab ini berisi tentang tahap *preprocessing* data mentah yang didapatkan. Keluaran dari tahap ini berupa data yang digunakan dalam penelitian.

### **BAB V IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisi tentang implementasi pengolahan data menggunakan model konseptual dan metode yang telah dijelaskan. Selain itu, pada bab ini akan dilakukan pengujian hasil yang didapatkan menggunakan matriks evaluasi statistik.

### **BAB VI KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian yang dilakukan. Pada bab ini juga akan dijelaskan saran yang dapat dipertimbangkan untuk pengembangan dan penelitian selanjutnya.