

BAB I

PENDAHULUAN

1.1. Latar Belakang

Peningkatan pemanfaatan sumber daya jaringan yang sangat cepat merupakan hal yang harus diperhatikan oleh pengelola jaringan. Saat ini jaringan komputer berkembang pesat dan semakin kompleks. Aliran data yang dominan membebani jaringan komputer sebesar 80% dari total lalu lintas jaringan adalah layanan peer to peer (P2P) (Callado et al., 2009) yang merupakan aplikasi file berbagi pakai serta aplikasi streaming video, contohnya seperti Youtube.

Pengalamatan host dan perangkat yang tidak sesuai persyaratan yang telah ditentukan oleh pengelola jaringan, serta perilaku pengguna jaringan yang tidak bertanggung jawab sering menyebabkan terjadinya kegagalan dalam komunikasi data. Pengguna jaringan dengan perilaku yang kompleks dalam memanfaatkan sumber daya jaringan juga berdampak besar pada pengguna jaringan lainnya dan perangkat-perangkat di dalam jaringan (Hemmati et al., 2016), tren baru seperti penggunaan streaming video dan cloud sharing juga membebani jaringan serta menyebabkan performa jaringan menjadi menurun (Maryam et al., 2017).

Pembatasan bandwidth yang diterapkan administrator kepada setiap pengguna jaringan dirasakan belum efektif dalam hal mengatur pemanfaatan sumber daya jaringan, karena kebutuhan pengguna juga meningkat tanpa diikuti dengan ketersediaan bandwidth yang ada saat ini (Badea et al., 2015). Bagi pengelola jaringan sangatlah penting dalam memahami perilaku pengguna pada

jaringan komputer yang dikelolanya, perilaku pengguna tersebut diwakili aktivitas pengguna jaringan dengan berbagai jenis aplikasi dan program yang dicirikan oleh penentuan aktivitas protokol dan port (Leng et al., 2015). Penggunaan port dan protokol adalah suatu event, semua event tersebut dicatat oleh sistem pada router untuk mendeteksi perilaku pengguna yang tidak normal. Mengingat perilaku pengguna yang tidak normal memiliki banyak cara yang rumit dan terus berevolusi guna memaksimalkan bandwidth yang diperoleh, semisal menghapus atau mengganti data sensitif.

Sebuah sistem untuk mendeteksi perilaku pengguna sangat penting, karena akan memberikan visibilitas dan kontrol yang lebih besar untuk keamanan jaringan, sistem ini memungkinkan administrator jaringan untuk mengamati kejadian secara real time, baik dari sistem operasi komputer maupun dari perangkat keamanan lainnya di dalam jaringan, peristiwa ini terpusat, berkorelasi dan tersimpan dalam database sistem untuk mendeteksi perilaku pengguna yang tidak normal.

Sebagian besar peneliti lebih serius mengamati dan melakukan pengukuran terhadap *QoS(Quality of Service)* dan *QoE(Quality of Experience)*, sedangkan perilaku dan kebiasaan pengguna jaringan merupakan salah satu hal penting yang mempengaruhi *QoS* dan *QoE*. Beberapa penelitian terhadap perilaku pengguna jaringan sebelumnya yang pernah diusulkan adalah untuk mengetahui perilaku pengguna jaringan dengan menerapkan metode TF – IDF (*Term Frequency – Inverse Document Frequency*) (Leng et al., 2015) terhadap data internet browsing yang diperoleh dari jaringan WiFi untuk menampilkan implikasi secara nyata terhadap hasil pengelompokan terhadap dokumen dan model pengguna jaringan.

Dari penerapan metode tersebut diperoleh pengelompokan perilaku pengguna yang lebih detail serta dapat mengekstraksi kebiasaan pengguna yang tersembunyi pada saat melakukan browsing dengan pemilahan kemiripan pengguna kedalam beberapa kelompok perilaku pengguna jaringan.

Data Mining merupakan suatu *knowledge* yang menggunakan beberapa teknik, seperti teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk menemukan suatu pola-pola yang dapat mengidentifikasi suatu informasi yang bermanfaat. Algoritma yang digunakan adalah *K-Means*, alasan penggunaan algoritma *K-Means* diantaranya ialah karena algoritma ini memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu algoritma *K-Means* ini tidak terpengaruh terhadap urutan objek (Ediyanto et al., 2013)

Pada metode *K-Means* terdapat langkah-langkah untuk menentukan *cluster* setiap data yaitu menentukan jumlah *cluster*, tempatkan setiap data ke *cluster* terdekat dengan menggunakan jarak Euclidean, hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang baru, dan terakhir ulangi langkah 2 dan 3 dengan memakai pusat *cluster* yang baru. Hasil akhir dari perhitungan metode ini adalah mendapatkan pengelompokan data, kemudian mendeskripsikan nilai-nilai yang sudah didapatkan untuk menghasilkan sebuah *knowledge* baru mengenai pola data pengguna internet. Proses perhitungan analisa yang dilakukan menggunakan dua cara yaitu dengan cara manual dan menggunakan aplikasi data mining (Rapidminer). Aplikasi RapidMiner dilakukan untuk pengujian hasil akhir.

Aplikasi tersebut digunakan karena langkah-langkah perhitungan algoritma yang dilakukan sama dengan perhitungan manual.

Berdasarkan penjelasan sebelumnya bahwasanya perilaku pengguna dapat digunakan sebagai acuan dalam menentukan kebiasaan penggunaan jaringan internet sehingga akan didapatkan diwaktu kapan dan situs apa yang paling banyak atau sering digunakan serta memakan bandwidth terbesar. Ketika waktu dan situs yang paling ramai didapatkan maka akan dengan mudah dalam melakukan manajemen bandwidth dan lebih terarah dalam mengalokasikan bandwidth tersebut.

1.1. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah :

1. Bagaimana mengelompokkan waktu pengguna internet di Vitazny Hotspot dengan menggunakan metode *clustering K-Means*?
2. Bagaimana menentukan situs pengguna yang ramai dari setiap konten di Vitazny Hotspot dengan menggunakan metode *clustering K-Means*?

1.2. Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut:

1. Data set yang dikaji dalam penelitian ini menggunakan data *bandwidth* dari *log RAW firewall* mikrotik yang dikumpulkan selama 1 minggu
2. RAW Data yang digunakan berdasarkan data pengguna di Vitazny Hotspot

1.3. Tujuan Penelitian

Penelitian ini mempunyai tujuan sebagai berikut:

1. Mengelompokkan perilaku pengguna berdasarkan waktu pemakaian internet di Vitazny Hotspot dengan menggunakan metode clustering K-Means
2. Menentukan situs yang sering digunakan di Vitazny Hotspot dengan menggunakan metode clustering K-Means

1.4. Manfaat Penelitian

Penelitian ini mempunyai manfaat sebagai berikut;

1. Menentukan *bandwidth* yang tepat untuk *Vitazny Hotspot*.
2. Menentukan berapa besar sumber daya internet yang dibutuhkan oleh *Vitazny Hotspot*.
3. Mengetahui situs yang diakses pengguna pada konten porno.
4. Mengelompokkan perilaku pengguna menggunakan metode clustering K-Means pada waktu pemakaian internet berdasarkan lalu lintas data yang di rekam oleh router mikrotik.

1.5. Metodologi Penelitian

Metodologi penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Pra Penelitian
 - a. Studi Pendahuluan

Dilakukan dengan cara mengambil data langsung di Vitazny

Hotspot selama satu minggu

b. Studi Pustaka

Mencari sumber referensi tentang Data Mining dan *Clustering*
Referensi bersumber dari buku, jurnal dan tugas akhir mahasiswa
lain yang penelitiannya berhubungan dengan data mining.

2. Analisis Data Mining

Metode analisis data mining sebagai tahapan untuk menganalisa data dalam menerapkan ilmu data mining menggunakan tahapan *Knowledge Discovery in Database* (KDD) yang terdiri dari *data cleaning, data integration, data selection, data transformation, data mining, knowledge presentation.*

1.6. Sistematika Penulisan

Sistematika dalam penulisan ini disusun untuk mempermudah pembacaan dan gambaran umum tentang penelitian yang akan dilakukan dengan sistematika penulisan penelitian ini sebagai berikut :

BAB 1 PENDAHULUAN

Membahas tentang latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian yang ingin dicapai, manfaat penelitian, metodologi penelitian yang merupakan langkah untuk menyelesaikan masalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Membahas berbagai konsep dasar teori-teori yang berkaitan dengan penelitian yaitu teori mengenai data mining, metode *clustering*, metode klasifikasi, algoritma K-Means , dan RapidMiner

BAB III METODOLOGI

Menjelaskan kebutuhan perangkat lunak yang akan digunakan serta metode yang akan digunakan dalam penelitian, dengan berbagai analisis yang akan dijabarkan. Hasil dari analisis ini digunakan untuk melakukan proses penelitian.

BAB IV HASIL DAN PEMBAHASAN

Berisi hasil dari proses yang telah dilakukan dari penelitian ini. Lebih tepatnya adalah membahas implementasi dari tools dan metode yang digunakan serta proses pengolahan data set serta menampilkan hasil dari proses tersebut, disertai dengan pengujian untuk mengetahui apakah penelitian ini berjalan dengan baik atau tidak.

BAB V KESIMPULAN DAN SARAN

Berisi tentang penarikan kesimpulan dari hasil penelitian yang telah dilakukan, lebih membahas tentang kesimpulan keterkaitan masalah dan tujuan penelitian yang dilakukan. Juga saran yang bersifat membangun untuk penelitian selanjutnya tentang penelitian ini agar dapat menutupi kekurangan-kekurangan yang ada.