

BAB I

PENDAHULUAN

1.1. Latar Belakang

Situs jejaring sosial merupakan gaya hidup sosial baru yang muncul seiring berkembangnya internet. Gaya hidup baru tersebut memiliki ruang lingkup yang lebih luas dan terbuka dibandingkan gaya hidup konvensional. Sebelum era jejaring sosial, publik sulit menyuarkan opininya akibat keterbatasan cara dan sarana. Sarana publik untuk menyuarkan opini seperti koran, televisi, radio, dan forum memiliki alur cara penyampaian yang rumit dan tidak memiliki jangkauan luas. Sebelum era digital, tidak semua informasi dapat tercatat dengan baik. Hal tersebut dikarenakan publik lebih sering menyuarkan opininya secara verbal.

Tanpa adanya catatan informasi dan jangkauan informasi yang sempit, membuat sulit untuk dilakukan tindak lanjut mengenai opini publik. Setelah berkembangnya internet dan jejaring sosial, publik mendapatkan cara dan sarana yang efektif untuk menyampaikan opini. Situs jejaring sosial secara global dapat menjangkau seluruh kalangan yang terhubung dengan internet dan membebaskan penggunanya untuk melakukan berbagai hal, salah satunya penyampaian opini.

Segala hal yang dilakukan pengguna jejaring sosial, termasuk opini pengguna, akan tersimpan dengan baik di dalam server penyedia situs jejaring sosial dalam jangka waktu yang telah ditentukan. Hal tersebut membuat dapat dilakukannya tindak lanjut mengenai opini publik. Beberapa bentuk tindak lanjut yang dapat dilakukan adalah menggali data opini publik dan

menganalisisnya untuk mendapatkan informasi yang berguna.

Pada kuartar ke - 3 2015 rata-rata pengguna internet menghabiskan 1.77 jam untuk menggunakan jejaring sosial. Secara khusus, pengguna internet di Indonesia menghabiskan 2,8 jam untuk menggunakan jejaring sosial (GlobalWorldIndex, 2015). Waktu interaksi yang intens di jejaring sosial, berpotensi membuat publik lebih sering menyuarakan opini.

Twitter merupakan jejaring sosial yang memiliki pangsa pasar global sebesar 43%. Dengan rincian, pengguna aktif sebesar 21% pengguna internet global. Dapat dijabarkan pula 58% pengguna aktif *Twitter* secara rutin mengakses *Twitter* (GlobalWorldIndex, 2015). Dapat disimpulkan bahwa *Twitter* adalah jejaring sosial yang besar dengan data-data yang dihimpun akan masuk kategori *big data*.

Pengamatan dan analisa dapat dilakukan dengan menangkap data-data dari situs jejaring sosial *Twitter* yang berkaitan dengan *brand*. Data diambil dari *Twitter* karena *tweet* yang berisikan opini publik dibatasi 140 karakter sehingga data mudah diolah. Data-data tersebut akan diolah menggunakan teknik data *mining* untuk menghasilkan hasil yang tepat dan performa pemrosesan yang efisien. Klasifikasi data berdasarkan sentimen akan menggunakan algoritma klasifikasi dari data *mining* yaitu *Naïve Bayes Classifier*. Klasifikasi data berdasarkan topik akan menggunakan metode perkiraan relasi kata. Pengolahan data dilakukan dengan tujuan untuk mendapatkan topik opini publik dan diklasifikasikan kembali berdasarkan sentimen positif atau negatif.

Pemilik *brand* dapat mengetahui opini publik mengenai *brand* mereka dengan memilah data dari jejaring sosial. Opini publik di jejaring sosial mengenai suatu *brand* dapat mempengaruhi citra *brand*. Penting bagi pemilik *brand* untuk menjaga citra mereka karena dapat mempengaruhi kelangsungan bisnis *brand*. Menjadi bijaksana bila pemilik *brand* untuk menjaga citra positif bagi *brand* mereka dengan mengamati dan menganalisa opini publik di jejaring sosial. Dengan menganalisa opini publik berdasarkan topik dan sentimen, pemilik *brand* akan mendapatkan informasi penting dalam menentukan kebijakan strategis *brand* dalam meningkatkan citra *brand*.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah dalam penelitian ini dapat dijabarkan rumusan masalah :

1. Bagaimana mengamati dan menganalisis opini publik mengenai *brand* di *Twitter*?
2. Bagaimana mengklasifikasikan *tweets* berdasarkan topik?
3. Bagaimana mengklasifikasikan *tweets* berdasarkan sentimen?

1.3. Batasan Masalah

Menimbang banyaknya aspek yang dapat diperhatikan dalam penelitian ini, maka diberikan batasan masalah sebagai berikut:

1. Data penelitian merupakan *tweet* pengguna *Twitter* berbahasa Indonesia.
2. Tata bahasa yang tidak teratur dan tidak baku di luar fokus penelitian.

3. *Brand* yang dibahas sebagai objek penelitian adalah produk *smartphone Asus* dengan *brand Zenfone*.
4. Metode yang digunakan untuk klasifikasi adalah metode *Naïve Bayes Classifier* dan perkiraan relasi kata.
5. Hasil dari analisa *brand* dalam penelitian ini adalah data *tweets* yang sudah terklasifikasikan berdasarkan topik bahasan dan sentimen dalam bentuk *excel* dan grafik informasi.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

1. Mengamati dan menganalisis opini publik mengenai *brand* di *Twitter*.
2. Mengkategorikan opini berdasarkan topik bahasan.
3. Mengkategorikan opini berdasarkan sentimen positif atau negatif.

1.5. Metodologi Penelitian

Metodologi penelitian merupakan penjelasan tahapan-tahapan yang dilakukan dalam penelitian.

A. Studi Literatur

Studi Literatur dilakukan dengan melakukan pencarian literatur atau sumber pustaka yang berkaitan dengan penelitian *data mining* dengan sumber data dari *Twitter*, teknik klasifikasi *Naïve Bayes Classifier*, jejaring sosial dan *brand*. Literatur yang ditemukan akan digunakan sebagai dasar penelitian. Dilakukan juga studi terhadap aplikasi yang diperlukan selama proses penelitian yaitu bahasa pemrograman C# dan Visual Studio 2008.

B. Pembangunan Perangkat Lunak

1. Analisis

Analisis dilakukan dengan mencari dan menganalisis data dan informasi mengenai kebutuhan penelitian sehingga dapat menjadi bahan pengembangan perangkat lunak. Dalam penelitian ini analisis akan menentukan data dan rincian data yang akan diolah, serta alur kerja perangkat lunak hingga menghasilkan informasi untuk dianalisa. Hasil analisis akan dituliskan dalam dokumen teknis Spesifikasi Kebutuhan Perangkat Lunak (SKPL).

2. Perancangan Perangkat Lunak

Perancangan dilakukan untuk mendapatkan deskripsi arsitektural perangkat lunak dari alat pemrosesan *data mining*, deskripsi antarmuka, deskripsi data, dan deskripsi prosedural. Hasil perancangan berupa dokumen Deskripsi Perancangan Perangkat Lunak (DPPL).

3. Implementasi Perangkat Lunak

Implementasi dilakukan dengan menerjemahkan deskripsi perancangan ke dalam bahasa pemrograman C# dengan pengelolaan data menggunakan Microsoft SQL Server. Dalam penelitian ini implementasi merupakan penerjemahan tahapan-tahapan data *mining* ke dalam bentuk perangkat lunak. Perangkat lunak tersebut akan menjadi alat pendukung untuk melakukan analisis informasi yang dihasilkan. Tahapan-tahapan data *mining* adalah sebagai berikut :

a. Seleksi Data

Metode seleksi data merupakan tahapan pertama dalam data *mining*. Dalam tahapan ini akan dikumpulkan sumber data untuk data *mining*. Pemilihan sumber data

dilakukan dengan mempertimbangkan tujuan dari penelitian. Dalam pemilihan sumber data juga ditentukan atribut data penting yang akan diproses.

Data yang digunakan dalam penelitian ini bertujuan untuk menganalisa opini publik mengenai *brand* Zenfone. Sumber data yang dipilih berasal dari *Twitter* yang dipersempit ruang lingkungannya pada *tweets* yang mengandung kata *Zenfone* dan dituliskan dalam berbahasa Indonesia. Atribut data yang digunakan adalah tanggal *tweet*, *screen name*, *full name*, teks *tweet*, *ID tweet*, *followers*, *follows*, *retweets*, *favorites*, *verified*, dan *user since*. Data diambil menggunakan *Twitter Archiver* (Agarwal, 2016) sebanyak 12.818 data yang diambil dalam rentang waktu 18 Desember 2015 - 22 Januari 2016. Hasil dari tahapan ini adalah didapatkan sumber data yang valid dengan atribut yang tepat dan berkesesuaian dengan tujuan penelitian.

b. *Preprocessing* Data

Metode data *preprocessing* merupakan tahapan data *mining* untuk mengolah data mentah menjadi data berkualitas yang siap untuk di proses secara efektif dan efisien di tahapan berikutnya. Data mentah merupakan data kotor yang belum dapat diolah karena terdapat ketidaklengkapan, *noisy* dan tidak konsisten. Data yang berkualitas dibutuhkan karena akan mendukung untuk menghasilkan hasil data *mining* berkualitas. Terdapat 5 tahapan dalam melakukan data *preprocessing* yaitu data *cleaning*, data *integration*, data *transformation*, data *reduction*, dan data *discretization*. Data *cleaning* merupakan proses untuk mengisi data kosong, menghaluskan *noise* data dan menangani data tidak konsisten. Data *integration*

merupakan proses untuk melakukan integrasi data dari beberapa basis data, kubus data atau berkas. Data *transformation* merupakan proses untuk melakukan normalisasi data agregasi data. Data *reduction* merupakan proses untuk mengurangi ukuran representasi data yang digunakan namun tetap menghasilkan hasil analisis yang sama. Data *discretization* merupakan proses untuk mengurangi data, secara khusus data numerik, untuk kepentingan tertentu.

Tahapan *Preprocessing* yang dilakukan dalam penelitian ini adalah data *cleaning* yang dilakukan untuk menghaluskan *noise* dan ketidakkonsistenan data yang muncul akibat banyaknya pengguna *Twitter* yang tidak menggunakan bahasa Indonesia yang baku dalam *tweet*. Hasil dari tahapan ini adalah dihasilkan data berkualitas yang siap di proses untuk menghasilkan informasi data *mining* yang berkualitas.

c. Transformasi Data

Metode transformasi data merupakan tahapan data *mining* untuk mengubah data menjadi bentuk yang tepat. Proses transformasi yang dapat dilakukan dalam tahap ini adalah *smoothing*, agregasi, generalisasi, normalisasi dan konstruksi atribut. Transformasi *smoothing* merupakan proses penghilangan *noise* dalam data. Transformasi agregasi merupakan proses melakukan agregasi data untuk menghasilkan informasi dan membentuk data *cube*. Transformasi generalisasi merupakan proses penalaran data untuk mengubah data ke kesimpulan umum mengenai data tersebut. Transformasi normalisasi merupakan proses penyederhanaan data untuk menghilangkan perbedaan data yang besar. Transformasi konstruksi atribut merupakan pembentukan atribut baru

yang didapatkan dengan menarik kesimpulan dari atribut yang sudah ada.

Penelitian ini melakukan transformasi data agregasi untuk menghasilkan mengenai informasi kata atau topik yang dibahas dalam sebuah *tweet*. Hasil dari tahapan ini adalah dihasilkan data yang sudah sempurna untuk dilakukan proses data *mining*.

d. *Data Mining*

Metode data *mining* merupakan tahapan penerapan metode klasifikasi data untuk mengekstraksi pola dari data. Proses data *mining* dapat dilakukan dengan berbagai metode klasifikasi. Metode *description* merupakan metode yang menggunakan penafsiran manusia untuk menemukan pola untuk mendeskripsikan data. Metode *prediction* merupakan metode yang menggunakan variabel untuk memprediksi nilai yang tidak diketahui dari variabel lain. Data hasil tahap transformasi data akan diolah dalam tahapan ini menggunakan metode dan algoritma yang sesuai untuk mendapatkan hasil informasi untuk dianalisa.

Penelitian ini menggunakan metode klasifikasi data *mining*, *Naïve Bayes Classifier*. Metode ini menghitung nilai posisi probabilitas untuk mengetahui kelompok kelas dari data. Metode ini memerlukan data *training* sebagai acuan dalam pengelompokan data. Penelitian ini akan menghitung nilai posisi probabilitas sentimen positif dan negatif dari data *tweets* yang menjadi subyek penelitian. Nilai perhitungan terbesar akan merujuk ke kelas data tersebut. Metode lain yang digunakan dalam penelitian ini adalah memperkirakan relasi kata dalam *tweets* dengan tabel referensi kategori. Jika ditemukan relasi kata maka *tweets* yang

diolah akan masuk ke dalam kelas kategori yang berelasi. Hasil dari tahapan ini adalah informasi pola dari data yang akan digunakan untuk interpretasi.

e. Interpretasi Hasil

Metode interpretasi merupakan tahapan terakhir dari data *mining*. Dalam tahapan ini informasi hasil data *mining* akan direpresentasikan sebagai pengetahuan. Untuk mendapatkan representasi data yang jelas maka informasi hasil data *mining* akan dianalisa untuk menghasilkan pengetahuan yang dibutuhkan untuk mencapai tujuan. Interpretasi data dilakukan dengan melihat tujuan dari data *mining*, kemudian melakukan penalaran terhadap informasi yang dihasilkan.

Pengetahuan yang dihasilkan dari penelitian ini akan dibuat dalam bentuk grafik jumlah pembahasan topik dan bobot topik terhadap sentimen. Hasil dari tahapan ini adalah representasi pengetahuan yang disajikan dalam bentuk visualisasi data seperti grafik dan diagram.

1.6. Sistem Penulisan Tugas Akhir

Adapun Sistematika penulisan laporan tugas akhir adalah sebagai berikut :

BAB 1 PENDAHULUAN

BAB 2 TINJAUAN PUSTAKA

BAB 3 LANDASAN TEORI

BAB 4 ANALISIS DAN PERANCANGAN SISTEM

BAB 5 IMPLEMENTASI DAN PENGUJIAN SISTEM

BAB 6 KESIMPULAN DAN SARAN