

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring berkembangnya teknologi informasi saat ini maka berkembang pula perangkat – perangkat yang mendukung guna akses informasi yang semakin dibutuhkan dalam waktu yang cepat dan dalam bentuk yang semakin fleksibel seperti komputer tablet, *smartphone* , televisi, dan beberapa perangkat lain yang terhubung ke internet. Dalam prakteknya pengguna perangkat yang menggunakan sistem operasi android sangat berkembang dengan sangat signifikan, menurut penelitian *comscore reports* April 2012 (Table 1.1) operating sistem dari google yang berbasis *open source* ini menempati urutan teratas mengungguli pesaingnya Apple, RIM, Microsoft dan Symbian.

Tabel 1. 1 Daftar Platforms Smartphone comscore reports April 2012

Top Smartphone Platforms 3 Month Avg. Ending Apr. 2012 vs. 3 Month Avg. Ending Jan. 2012 Total U.S. Smartphone Subscribers Ages 13+ Source: comScore MobiLens			
	Share (%) of Smartphone Subscribers		
	Jan-12	Apr-12	Point Change
Total Smartphone Subscribers	100.0	100.0	N/A
Google	48.6	50.8	2.2
Apple	29.5	31.4	1.9
RIM	15.2	11.6	-3.6
Microsoft	4.4	4.0	-0.4
Symbian	1.5	1.3	-0.2

Pada perangkat yang menggunakan sistem operasi android, user diberikan layanan dimana user dapat mengunduh aplikasi sesuai apa yang diinginkan dan sudah diberikan suatu portal guna mengunduh aplikasi secara online baik yang berbayar maupun gratis. Dengan adanya ribuan aplikasi yang ada didalam portal serta banyak aplikasi yang pada dasarnya sama namun memberikan fasilitas atau fitur – fitur

tambahan yang berbeda – beda maka sebelum mengunduh atau menginstal aplikasi sebaiknya membaca user review dari para pengguna aplikasi tersebut sebelumnya.

Dengan adanya sentimen analisis pada review pengguna aplikasi android maka membantu orang menemukan informasi berharga dari sejumlah banyak data yang tersusun secara tidak beraturan atau tidak terstruktur.

Google telah menyediakan suatu portal yang dapat digunakan untuk mengunduh atau menginstal aplikasi android sesuai keinginan baik yang gratis maupun yang berbayar melalui android market atau *google play* hanya dengan menggunakan akun yang terkoneksi dengan gmail. Disana juga telah disediakan fitur *user review* yang dapat dimanfaatkan oleh pengguna untuk memberikan hasil penilaiannya terhadap aplikasi tersebut secara langsung sehingga pengguna berikutnya dapat melihat penilaian tersebut lebih dahulu sebelum memutuskan mengunduh atau menginstal aplikasi tersebut.

Karena data berasal dari banyak pengguna yang memberikan pendapatnya maka akan sangat bervariasi dan akan menggunakan berbagai macam kata maupun kalimat yang pada intinya akan memberikan penilaian yang positif dan negatif terhadap aplikasi tersebut, oleh karena itu metode data mining dan information retrieval bisa digunakan untuk mengklasifikasikan data sentimen analisis tersebut.

Ada banyak metode data mining yang digunakan dalam sentimen analisis atau opinion mining salah satunya adalah *naive bayes* sesuai dengan penelitian [1] dan penelitian yang dilakukan Alec Go dkk [2] melakukan percobaan mengklasifikasi sentiment menggunakan *naive bayes*, *maximum entropy*, dan *support vector machines* pada tweet dengan mengekstrak kalimat menjadi *unigram*, *bigram*, *unigram* dan *bigram*, kemudian membagi tiga buah sentiment dengan tambahan netral. Percobaannya adalah menggunakan *emoticon* sebagai *noisy* label untuk data latih mendapatkan hasil yang efektif. *naive bayes*, *maximum entropy classification*, dan *support vector machines* memiliki akurasi yang tinggi dengan metode ini.

Jonathon Read menyatakan pentingnya domain, topik, waktu, dan gaya bahasa dalam melakukan analisis sentimen. Read mencoba dengan memanfaatkan emoticon

yang ada pada tiap – tiap sentimen, dan hasil yang diperoleh tidak terlalu signifikan [3].

Penelitian yang dilakukan oleh Parikh dan Movassate [4] yang mengklasifikasikan sentimen analisis dari media jejaring sosial dengan menggunakan metode *unigram* naïve bayes, *Multinomial* naïve bayes, dan *Maximum Entropy Classification*. Hasilnya memperlihatkan bahwa naïve bayes menunjukkan performa yang lebih baik dari *Maximum Entropy* dan memerlukan preproses text terlebih dahulu guna hasil yang lebih baik.

Prasad [5] menggunakan naïve bayes dengan berbagai macam pendekatan diantaranya: *Bernoulli*, *Bernoulli Chi Square*, *Multinomial Unigram*, *Linear Bigram*, *Backoff Bigram*, *Empirical Bigram*, dan *Weighted-Normalized Complement Naïve Bayes* (WCNB). Hasilnya *Multinomial Unigram*, *Bernouli Chi Square*, dan *Linear Bigram* menunjukan hasil yang cenderung lebih baik dari pendekatan lain.

Mengacu dari penelitian – penelitian sebelumnya maka bisa diambil kesimpulan bahwa untuk mengklasifikasikan sentimen analisis diperlukan pemrosesan awal data yang baik untuk dapat dihasilkan pengklasifikasian opini berdasarkan sentimennya dengan tingkat akurasi yang tinggi. Dan menurut Ali dkk [6] menyatakan bahwa algoritma naïve bayes adalah algoritma yang memiliki popularitas besar dan telah sukses diterapkan di banyak domain serta naïve bayes memiliki berbagai kelebihan diantaranya Kuat terhadap pengisolasi gangguan pada data, Jika terjadi kasus *missing value* ketika proses komputasi sedang berlangsung, maka objek tersebut akan diabaikan, dapat digunakan untuk data yang tidak relevan [7]. Oleh karena itu pada penelitian ini penulis mengusulkan penggunaan algorima naïve bayes.

Dengan demikian dibutuhkan pemrosesan data awal secara lebih baik sebelum diklasifikasikan sesuai dengan sentimennya dan perlu dilakukan seleksi fitur untuk mengurangi atribut dari data yang tidak diperlukan atau tidak berpengaruh dengan hanya mengambil komponen yang relevan dari data awal, sehingga mengurangi dimensi data yang di proses untuk menghasilkan akurasi dan komputasi yang lebih baik. Sesuai dengan penelitian Pechenizkiy [8] yang menyatakan bahwa ekstraksi

fitur secara signifikan dapat meningkatkan akurasi serta menghasilkan ruang representasi yang lebih baik dan menyelesaikan masalah dimensi.

Sesuai penelitian yang dilakukan Mohajerani dan Ebrahimi [9] bahwa pembobotan atribut (*attribute weighting*) atau pemilihan fitur (*feature selection*) akan dapat dimungkinkan pengurangan ukuran dataset dan akan mempermudah dalam mengelola untuk dapat dijadikan suatu model.

Oleh karena itu pada penelitian ini peneliti mengintegrasikan beberapa metode pembobotan atribut yang diterapkan sebelum mengklasifikasikan review pengguna terhadap aplikasi android.

1.2. Rumusan Masalah

Dari latar belakang masalah di atas, dapat disimpulkan bahwa diperlukan pengolahan awal data untuk menghasilkan proses klasifikasi yang akurat untuk analisis sentimen.

Sedangkan pertanyaan penelitian (*research questions*) pada penelitian ini adalah: “seberapa besar akurasi yang dihasilkan dari pembobotan atribut (*attribute weighting*) terhadap algoritma naive bayes untuk analisis sentimen ?”

1.3. Tujuan Penelitian

Jadi tujuan penelitian ini adalah evaluasi pembobotan atribut pada pengolahan awal data diharapkan hasil klasifikasi pada data review aplikasi android menggunakan algoritma naïve bayes menjadi semakin akurat.

1.4. Manfaat Penelitian

Adapun penelitian ini dapat memberikan manfaat antara lain sebagai berikut:

a. Manfaat praktisi

Dari penelitian ini diharapkan dapat diterapkan kedalam website / portal yang menyediakan review dari konsumen atau user yang secara otomatis mampu memberikan referensi kepada pembaca.

b. Manfaat teoritis

Dapat dari penelitian ini dapat memberikan sumbangsih terhadap metode pengolahan awal data dan klasifikasi *opinion mining* atau analisis sentimen.

c. Manfaat penelitian

Penelitian ini dapat digunakan sebagai acuan pada penelitian selanjutnya, khususnya dalam hal yang memiliki keterkaitan dalam hal analisis sentimen, serta algoritma lain yang belum tercantum disini

d. Manfaat akademis

Dengan penelitian ini diharapkan dapat membantu peneliti memperoleh gelar Master di Universitas Dian Nuswantoro Semarang.