

BABI

PENDAHULUAN

1.1 Latar Belakang Penelitian

Salah satu media sosial yang banyak digunakan di Indonesia adalah twitter. Menurut halaman www.kominfo.go.id, Indonesia menempati peringkat 5 pengguna Twitter terbesar di dunia. Posisi Indonesia hanya kalah dari USA, Brazil, Jepang dan Inggris. Pengguna Twitter, berdasarkan data PT Bakrie Telecom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global. Twitter menjadi salah satu jejaring sosial paling besar di dunia sehingga mampu meraup keuntungan mencapai USD 145 juta.

Begitu banyak jumlah informasi yang terdapat pada twitter, tidak sedikit dari informasi yang ada berisikan opini pengguna. Pengguna twitter dapat dengan bebas memberikan berbagai opini, misalnya saja terkait provider atau penyedia layanan internet yang ada di Indonesia dengan melakukan diskusi antara pengguna mengenai kepuasan, kekecewaan ataupun harapan terhadap layanan yang diberikan.

Dari berbagai *tweet* atau komentar yang terkumpul dapat diperoleh gambaran mengenai opini pengguna melalui proses penggalian informasi lebih lanjut dari tiap komentar yang dibuat oleh pengguna sebagai bahan evaluasi untuk meningkatkan pelayanan terhadap pelanggan dengan melakukan proses analisis. Proses analisis yang bisa digunakan untuk melihat opini masyarakat melalui komentar adalah analisis sentimen.

Analisis sentimen atau *opinion mining* merupakan salah satu bidang penelitian dalam text mining. Analisis sentimen merupakan proses untuk melakukan identifikasi sentimen yang muncul pada suatu teks dengan mengolah data tekstual untuk memahami opini yang terkandung dalam suatu sentimen. Analisis sentimen cenderung dilakukan untuk melihat opini dari suatu sentimen

terhadap sebuah objek dengan mengidentifikasi apakah sentimen tersebut mengandung opini positif atau negatif (Darma S, Rizal Setya Perdana, 2018)

Terdapat berbagai teknik klasifikasi yang dapat diterapkan pada analisis sentimen, salah satu yang sering digunakan yaitu *Support Vector Machine (SVM)*. *SVM* merupakan algoritma klasifikasi yang pada umumnya digunakan untuk mengkategorikan teks. Algoritma *SVM* merepresentasikan contoh kasus dari kategori berbeda yang dipetakan sebagai titik dalam ruang dan dipisahkan oleh margin dengan jarak maksimal. *SVM* memberikan hasil terbaik daripada algoritma Naive Bayes dan tool klasifikasi sentimen. Dengan menemukan hyperplane yang direpresentasikan sebagai vektor w yang memisahkan vektor dokumen dalam satu kelas dari vektor di kelas lain (Bholane Savita D., 2016).

Dalam studi literatur penggunaan algoritma untuk klasifikasi dengan hasil terbaik dapat diraih menggunakan metode *SVM* (Hadna, Santosa, & Winarno, 2016). Dengan membandingkan hasil penelitian – penelitian yang sudah ada meliputi penelitian yang dilakukan oleh (Padmaja dkk., 2014) dengan membandingkan dua metode analisis sentimen, yaitu *Naive Bayes Classifier* dan *SVM*, penelitian oleh (Chandani dkk., 2014) terhadap 1000 data review film yang diambil dari situs IMDB dan penelitian yang dilakukan oleh (Taheri & Mammadov, 2013) tentang model optimasi klasifikasi juga membandingkan beberapa metode klasifikasi yang berbeda seperti *Naive Bayes*, *Augmented Naive Bayes*, *SVM*, *C4.5* dan *NN*.

Terdapat penelitian yang dilakukan dengan mengklasifikasikan teks kedalam opini positif atau negatif untuk proses pemilihan produk smartphone android dan black berry dengan menggunakan data *review* kedua *smartphone*. Algoritma klasifikasi yang digunakan adalah *SVM* dengan optimasi pemilihan fitur menggunakan algoritma genetika. Hasil yang diperoleh setelah penerapan algoritma genetika, menunjukkan peningkatan hasil akurasi, dari 71% menjadi 78.02%. Pada pengembangan penelitian, penulis mengharapkan peningkatan hasil akurasi analisis sentimen dengan melakukan penambahan fitur ataupun membandingkan dengan metode klasifikasi yang lain (Darmawan, 2015).

Penelitian analisis sentimen yang sama juga dilakukan, dengan menggunakan data twitter mengenai layanan transportasi online ternama yaitu GoJek. Data komentar yang terkumpul dihilangkan *noise*-nya, selanjutnya digunakan metode ekstraksi fitur dan algoritma *SVM* untuk metode klasifikasi. Hasilnya adalah prediksi data komentar yang dianggap sebagai sentimen positif atau negatif terhadap layanan GoJek. Dari hasil pengujian diperoleh akurasi 86%, tingkat kesalahan prediksi 14%, tingkat prediksi yang benar untuk sentimen positif 100%, dan tingkat prediksi yang benar untuk sentimen negatif 67,44% (Windasari, Uzzi, & Satoto, 2018).

Dari penelitian analisis sentimen menggunakan data twitter yang ada, sebagian besar gambaran proses dasar analisis sentimen dilakukan dengan mengumpulkan data komentar twitter terlebih dahulu. Data yang terkumpul diberikan label secara manual, dan digunakan sebagai data *training* dan *testing* pada klasifikasi *SVM*. Tahap selanjutnya adalah *text preprocessing* yaitu mengolah data komentar agar siap untuk diproses dan menghilangkan *noise* data. Hasil dari *text preprocessing* yang berupa teks diubah kedalam bentuk angka melalui proses *feature extraction* dengan melakukan perhitungan *TF-IDF*. Nilai *TF-IDF* ini yang menjadi masukan untuk algoritma *SVM* untuk mengklasifikasikan komentar tersebut masuk ke kategori komentar positif atau negatif.

Dalam penulisan komentar tentunya penggunaan bahasa tidak sangat diperhatikan. Banyak pengguna twitter menggunakan bahasa tidak baku seperti bahasa pergaulan, salah ejaan penulisan dan salah penggunaan singkatan. Seperti penulisan kata “Tidak” menjadi “Tdk” atau “Tidk”, atau penulisan kata “Bagus” menjadi “Bgs” atau salah ketik “Bguus”.

Levenshtein Distance atau yang biasa disebut dengan *edit distance* adalah suatu metode yang dapat digunakan untuk mengatasi terjadinya kesalahan ejaan yang dibuat oleh Vladimir Levenshtein. Kesalahan ejaan terjadi apabila kata yang diketik oleh pengguna tidak sesuai dengan daftar kata pada kamus yang digunakan. Perhitungan algoritma *Levenshtein Distance* didapatkan dari matriks yang digunakan untuk menghitung jumlah perbedaan string antara dua string.

Perhitungan jarak antara dua string ditentukan dari jumlah minimum operasi perubahan untuk membuat string A menjadi string B.

Pada penelitian sebelumnya implementasi algoritma *Levenshtein Distance* digunakan dalam penelitian untuk menampilkan saran perbaikan kesalahan pengetikan dokumen berbahasa Indonesia. Algoritma *Levenshtein Distance* ditambahkan pada sistem pengecekan ejaan Bahasa Indonesia berbasis web dengan menggunakan bahasa pemrograman PHP. Penerapan algoritma *Levenshtein Distance* dapat membantu mengatasi permasalahan pada kesalahan pengetikan (Adriyani, Ni Made Muni dkk, 2010)

Oleh karena itu akan dilakukan perbaikan kesalahan ejaan dalam tahap *text preprocessing* dengan menggunakan algoritma *Levenshtein Distance* untuk memperbaiki kesalahan kata pada tahap *text preprocessing*, sehingga dapat diketahui perbedaan hasil klasifikasi sebelum dan setelah digunakan algoritma *Levenshtein Distance* pada tahap *text preprocessing* dengan menggunakan *Support Vector Machine*.

1.2 Rumusan Masalah

Adapun rumusan masalah yang terdapat dalam penelitian ini meliputi:

- a. Bagaimana implementasi algoritma *Levenshtein Distance* pada tahap *text preprocessing*?
- b. Bagaimana perbedaan hasil yang didapat, sebelum dan setelah mengimplementasikan algoritma *Levenshtein Distance* pada tahap *text preprocessing* analisis sentimen data twitter menggunakan *SVM*?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini yaitu:

- a. Komentar pada twitter yang digunakan adalah komentar dengan Bahasa Indonesia.

- b. Komentar yang akan diklasifikasikan adalah komentar yang berkaitan dengan pengguna provider atau penyedia layanan internet di Indonesia meliputi Indosat, XL Axiata, Telkomsel, Smartfren, Indihome, dan Biznet.
- c. Pada proses SVM, pada penelitian ini menggunakan *libsvm php*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu;

- a. Mengetahui implementasi algoritma *Levenshtein Distance* pada tahap *text preprocessing*.
- b. Mengetahui perbedaan hasil klasifikasi *SVM* yang didapat, sebelum dan setelah menggunakan algoritma *Levenshtein Distance* pada tahap *text preprocessing*.

1.5 Manfaat Penelitian

Manfaat dari hasil penelitian ini dalam bidang akademis, diharapkan mampu memberikan solusi untuk pengembangan penelitian mengenai analisis sentimen dengan mengelola *data input* dalam proses analisis sentimen, misalnya mempersiapkan data awal menjadi data yang lebih mudah dan optimal untuk diproses sehingga dapat meningkatkan hasil penelitian. Selain itu manfaat praktis penelitian ini diharapkan dapat digunakan sebagai bahan evaluasi untuk meningkatkan pelayanan dari penyedia layanan internet sehingga akan menambah banyak pelanggan dan menopang pertumbuhan perusahaan dengan melakukan analisis sentimen data pelanggan.