

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dengan perkembangan internet dan web 2.0 platform seperti blog, forum diskusi, jaringan *peer-to-peer*, dan berbagai macam tipe media sosial lainnya, konsumen memiliki jangkauan dan kekuatan untuk berbagi pengalaman dan opini mereka terhadap suatu produk atau layanan. Perusahaan-perusahaan besar juga semakin menyadari bahwa pendapat suatu konsumen dapat memberikan pengaruh yang sangat besar dalam pembentukan loyalitas konsumen, keputusan pembelian konsumen, *brand* perusahaan, dan pendapat konsumen lain (Zabin dkk., 2008).

Menganalisa data *consumer reviews* untuk membuat data menjadi lebih dinamis merupakan bidang yang penting sekarang. Meningkatkan algoritma berbasis pembelajaran mesin untuk membaca ribuan ulasan suatu produk akan memakan waktu. Sehingga kategorisasi ulasan dibutuhkan untuk memahami popularitasnya di antara pembeli di seluruh dunia (Johar dkk., 2020).

Keberhasilan setiap bisnis perusahaan bergantung pada kepuasan pelanggan. Perusahaan-perusahaan yang berhasil memuaskan pelanggan sepenuhnya akan tetap berada di posisi teratas di *market*. Perusahaan saat ini sudah mengetahui bahwa kepuasan pelanggan adalah kunci untuk kesuksesan bisnis dan pada saat yang sama berperan penting untuk memperluas *market value* (Khadka dkk., 2017).

Kepuasan pelanggan dipengaruhi oleh persepsi pelanggan terhadap suatu produk atau layanan. Kepuasan suatu pelanggan juga seringkali dipengaruhi oleh orang lain (Zeithaml dkk., 2003). Kepuasan pelanggan yang meningkat dapat

memberikan keuntungan bagi perusahaan seperti loyalitas pelanggan. Kepuasan pelanggan terhadap suatu produk atau layanan perusahaan bisa membuat pelanggan untuk sering membeli produk dan merekomendasikan produk atau layanan kepada pelanggan yang lain (Tao 2014).

Opinion mining mengacu pada identifikasi dan klasifikasi sudut pandang atau opini yang diekspresikan dalam teks menggunakan komputasi linguistik. Pendapat yang diungkapkan dalam sebuah topik lebih penting daripada topik itu sendiri (Jebaseeli dkk., 2012). *Opinion mining* mengekstrak informasi subjektif dari sumber material seperti opini menggunakan teknik seperti *natural language processing*, dan teks analitik. Opini berperan penting dalam pengumpulan informasi sebelum melakukan pengambilan keputusan (Jebaseeli dkk., 2012).

Lidya dkk. (2015) melakukan analisis sentimen pada teks berbahasa Indonesia. Algoritma yang digunakan adalah Naïve bayes dan Support Vector Machine. Berdasarkan hasil penelitian yang dilakukan dengan menggunakan *cross validation* dengan varian fold (4,5,6,10...15), terbukti bahwa SVM memiliki akurasi yang lebih tinggi dibandingkan dengan Naïve bayes.

Tan (2008) melakukan studi empiris mengenai kategorisasi sentimen pada *Chinese documents*. Songbo Tan menggunakan empat metode seleksi fitur (MI, IG, CHI dan DF) dan lima learning method (klasifikasi centroid, KNN, klasifikasi winnow, Naïve Bayes dan SVM). Berdasarkan hasil penelitian dapat disimpulkan bahwa IG (Information gain) menghasilkan performa terbaik dalam seleksi fitur dan SVM menunjukkan performa yang terbaik dalam klasifikasi.

Jain dkk. (2019) melakukan sebuah survey mengenai klasifikasi sentimen pada *product review*. Algoritma yang digunakan dalam proses klasifikasi adalah (Naïve

Bayes, Support Vector Machines dan Maximum Entropy). Berdasarkan hasil penelitian, Support Vector Machine menunjukkan performa terbaik dalam ketepatan presisi.

Johar dkk. (2020) menggunakan pendekatan *supervised learning* model untuk melakukan klasifikasi pada data *Amazon product review*. Fitur ekstraksi yang digunakan adalah TF-IDF dan algoritma yang digunakan untuk klasifikasi adalah Naïve bayes dan Support Vector Machine. Berdasarkan hasil penelitian, akurasi kedua algoritma tidak jauh berbeda yaitu 0.82 (Naïve bayes) dan 0.83 (SVM).

Irham dkk. (2019) melakukan klasifikasi berita bahasa Indonesia menggunakan algoritma Support Vector Machine, Neural Network, K-nearest neighbors, dan Naïve Bayes dengan menggunakan TF-IDF sebagai fitur ekstraksi dan Mutual Information sebagai fitur seleksi. Berdasarkan hasil penelitian yang dilakukan, SVM menunjukkan hasil yang paling akurat, yaitu sebesar 94,24% *accuracy*.

Berdasarkan penelitian-penelitian yang telah dilakukan bisa disimpulkan bahwa Support Vector Machine memiliki performa yang terbaik dalam melakukan klasifikasi teks. Penelitian ini akan mengadopsi metode versi Sayyed Johar, dkk yaitu menggunakan fitur ekstraksi TF-IDF dengan algoritma Support Vector Machine. Selain itu, metode tersebut akan diterapkan pada dataset *user feedback* yang didapatkan dari ulasan pada aplikasi Tokopedia untuk mengklasifikasi kepuasan pelanggan yang terdapat dalam *user feedback* tersebut. Dengan itu, diharapkan dapat membantu para perusahaan mempermudah dalam memproses opini konsumen mengenai suatu produk.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang sudah diuraikan di atas, rumusan masalah yang dipilih adalah:

1. Bagaimana menerapkan algoritma Support Vector Machine yang menggunakan fitur ekstraksi TF-IDF untuk mengklasifikasi *user feedback* berdasarkan label yang diberikan.
2. Bagaimana mengukur performa klasifikasi kepuasan pelanggan yang menggunakan algoritma Support Vector Machine dengan fitur ekstraksi TF-IDF.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dapat dijabarkan menjadi beberapa poin sebagai berikut.

1. Label yang digunakan untuk klasifikasi kepuasan pelanggan ada dua yaitu puas dan tidak puas.
2. Dataset yang digunakan untuk klasifikasi adalah dataset ulasan produk pada aplikasi Tokopedia yang diambil dari Kaggle. Berikut tautan link data.
<https://www.kaggle.com/farhan999/tokopedia-product-reviews>.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang sudah ada, tujuan penelitian dapat dirumuskan sebagai berikut.

1. Mengimplementasikan Support Vector Machine yang menggunakan fitur ekstraksi TF-IDF untuk klasifikasi kepuasan pelanggan pada *user feedback*.

2. Mengukur performa dari metode Support Vector Machine yang menggunakan fitur ekstraksi TF-IDF untuk klasifikasi kepuasan pelanggan pada *user feedback*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini dapat dijabarkan sebagai berikut.

1. Membantu perusahaan dalam melakukan klasifikasi kepuasan pelanggan terhadap suatu produk.
2. Membantu perusahaan dalam menerapkan *market strategies* menggunakan klasifikasi kepuasan pelanggan pada *user feedback*.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan laporan skripsi ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menjelaskan latar belakang dari penelitian ini, rumusan masalah, batasan masalah, tujuan masalah, dan manfaat penelitian

BAB 2 LANDASAN TEORI

Bab ini berisikan penjelasan mengenai teori teori yang digunakan dalam penelitian antara lain *user feedback*, *TF-IDF*, *Latent Semantic Indexing*, *Singular Value Decomposition*, *Support Vector Machine*, dan *Confusion Matrix*.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisikan metodologi penelitian yang diuraikan kedalam bentuk *flowchart* beserta penjelasannya.

BAB 4 HASIL DAN DISKUSI

Bab ini berisi pembahasan dari implementasi sistem yang telah dibuat serta uji coba dari skenario uji terhadap algoritma *Support Vector Machine* untuk klasifikasi kepuasan pelanggan pada *customer feedback*.

BAB 5 SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk penelitian selanjutnya berdasarkan sistem yang telah dibuat.