

# KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN LAYANAN SMS “SUARA KONSUMEN” TERHADAP PRODUK MENGGUNAKAN METODE *K-NN*

David Addiwijaya Hanz<sup>(1)</sup>  
hanz.dave@gmail.com

Budi Susanto<sup>(2)</sup>  
budsus@ukdw.ac.id

Lukas Chrisantyo<sup>(3)</sup>  
lukaschris@ukdw.ac.id

## ***Abstract:***

*Limited number of characters in an SMS causes the use of words and the structure of solid and compact but obviously become very important in presenting the content and purpose of the sender of the SMS. This research studied how the use of SMS classify structures and limited words to get the sentiment of an SMS service "Suara Konsumen" through text mining approach. This study uses K-NN algorithm in the classification, besides the K-NN algorithm of this study will be weighted words using TF-IDF methods, Feature Selection in the selection of words also Cosine Similiarity in measuring the degree of proximity between documents. In determining the success, measured the accuracy of the documents the trial and the results are very accurate with an average value of 89% accuracy on the variation of K and Feature Selection.*

**Keywords:** *K-Nearest Neighbor, Sentiment Analyst, Document Classification, Short Message Service, Cosine Similiarity*

## **1. Pendahuluan**

Kepuasan konsumen menjadi salah satu faktor utama untuk dapat meningkatkan mutu produk tersebut. Beberapa produsen telah memberi sebuah layanan SMS yang disebut “SUARA KONSUMEN”. Layanan ini bertujuan untuk dapat mendapatkan saran, kritik, maupun pengaduan terhadap produk dari produsen tersebut. Terbatasnya jumlahnya karakter menjadi salah satu masalah utama dalam SMS, hal ini menyebabkan penggunaan kata-kata dan struktur yang ringkas namun padat dan jelas menjadi sangat penting dalam menyampaikan isi dan maksud dari pengirim SMS.

Penelitian ini akan meneliti bagaimana mengklasifikasikan SMS yang menggunakan struktur dan kata-kata yang terbatas untuk mendapatkan sentimen dari sebuah layanan SMS “Suara Konsumen” melalui pendekatan text mining sehingga diperlukan sebuah sistem yang mampu mengklasifikasikan setiap komentar yang masuk untuk dikelompokkan kedalam sentimen positif dan sentimen negatif. Dalam penelitian ini akan digunakan metode K-Nearest Neighbor (*K-NN*). dalam mengklasifikasikan dokumen, metode ini akan dibantu dengan penggunaan *feature selection* dalam melakukan pemilihan kata-kata, pembobotan tiap kata menggunakan metode TF-IDF dan mencari tingkatan kedekatan antar dokumen menggunakan *Cosine Similiarity*. Dalam mengukur akurasi Algoritma *K-NN* dalam mengklasifikasikan dokumen akan dibandingkan dengan penggunaan variasi nilai K dan Feature Selection sehingga dapat diketahui berapa nilai K dan Feature Selection yang optimal dalam mengklasifikasikan dokumen yang masuk.

---

<sup>1</sup>*Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.*

<sup>2</sup>*Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.*

<sup>3</sup>*Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.*

## 2. Metode Pembelajaran Mesin Dalam Sentimen Analisis

Tujuan bagian penelitian ini adalah bagaimana mesin dapat melihat setiap komentar yang telah masuk kedalam sistem dapat diproses sehingga sistem dapat mengenali komentar tersebut untuk diklasifikasikan kedalam 2 kategori yaitu sentimen positif dan sentimen negatif dalam hal ini terdapat 2 metode penting yaitu bagaimana sebuah komentar dapat dikatakan sebuah sentiment dan bagaimana Algoritma dalam mengenali sebuah komentar dan mengklasifikasikannya kedalam 2 kategori.

### 2.1. Sentimen Analisis

Sentimen Analisis merupakan proses menambang pendapat, merupakan bidang studi yang menganalisa pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka. (Liu, 2012). Kosakata sentimen merupakan bagian dari kata-kata yang memiliki makna frase dan idiom sentimen, terdapat dua kosa kata sentimen yaitu :

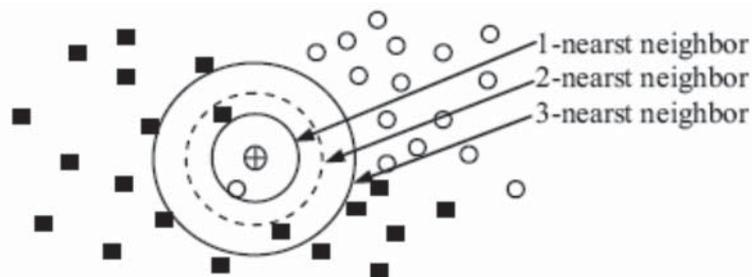
1. Kata sentimen positif yang digunakan untuk menunjukkan keadaan atau kualitas yang dikehendaki, Contoh kata sentimen positif yang cantik, indah, dan menakjubkan.
2. Kata sentimen negatif digunakan untuk mengekspresikan beberapa keadaan atau kualitas yang tidak diinginkan. Contoh kata sentimen negatif yang buruk, mengerikan, dan jelek.

### 2.2. Algoritma K-Nearest Neighbor

*K-NN* adalah metode lazy learning dalam arti bahwa tidak ada model yang dipelajari dari data pelatihan. Belajar hanya terjadi bila contoh uji harus diklasifikasikan (Liu, 2007). Berikut ini merupakan Algoritma KNN ( $D, d, k$ ) :

1. Hitunglah jarak antara  $d$  dan setiap contoh di  $D$ ;
2. Pilih contoh  $k$  di  $D$  yang terdekat  $d$ , menyatakan himpunan oleh  $P (\subseteq D)$ ;
3. Tetapkan  $d$  kelas itu adalah kelas yang paling sering di  $P$  (atau kelas mayoritas)

Komponen kunci dari algoritma KNN adalah fungsi jarak/kesamaan, yang dipilih berdasarkan aplikasi dan sifat data. Untuk data relasional, jarak *Euclidean* yang umum digunakan. Untuk dokumen teks, *cosine similarity* adalah pilihan populer.



Gambar 1. Ilustrasi Klasifikasi Algoritma K-Nearest Neighbor

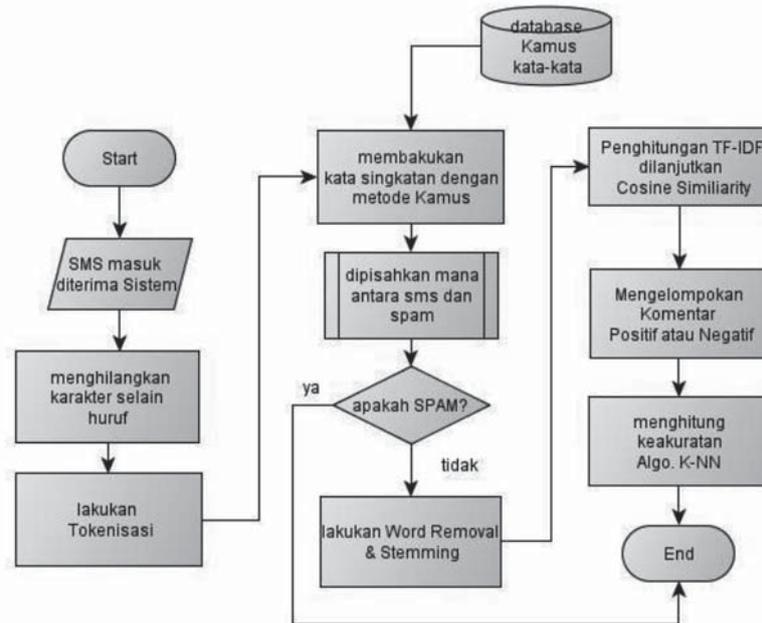
Dikutip dari Liu, Bing (2007). Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data (Web Data Mining\_ Exploring Hyperlinks, C - Bing Liu.pdf)

## 3. Hasil dan Pembahasan

Hasil dan pembahasan berisi bagaimana penelitian ini akan dibuat kedalam sebuah sistem yang telah dirancang berdasarkan teori diatas dan bagaimana Hasil dari penyelesaian masalah yang menjadi fokus utama kami dalam penelitian ini.

### 3.1. Alur Kerja Sistem

Gambar 2 memberikan gambaran arsitektur sistem klasifikasi dibentuk dan penerapan algoritma yang digunakan dalam menentukan komentar masuk kedalam kategori-kategori yang telah ditetapkan.



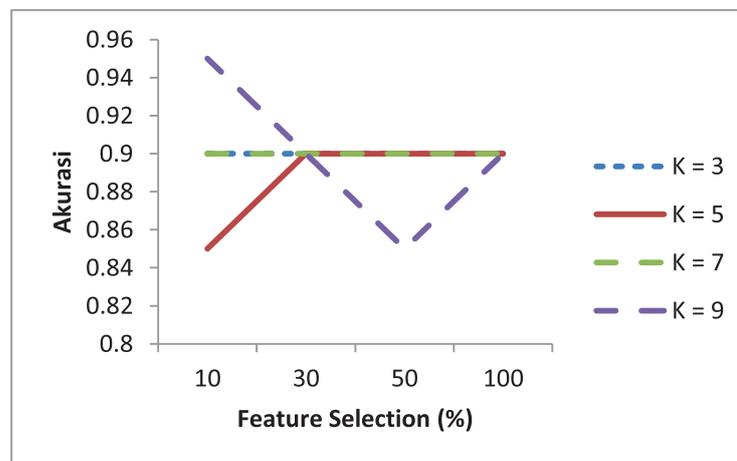
Gambar 2. Alur Sistem

Berdasarkan gambar diatas alur sistem dibagi menjadi 3 tahapan utama yaitu:

1. Pemrosesan data mentah menjadi data yang siap untuk dilakukan pembobotan yaitu dengan cara tokenisasi, *Expansion*, dan *Stopword Removal*. Prose ini sering disebut dengan *Preprocessing*.
2. Pembobotan dan Pengukuran tingkat kedekatan dokumen. Dalam hal ini sistem akan melakukan pembobotan yang menggunakan teknik TF-IDF, pemilihan kata-kata yang penting atau memiliki bobot tertinggi menggunakan *Feature Selection* dan mencari nilai kedekatan antar dokumen menggunakan *Cosine Similarity* sehingga hasil akhirnya akan dicari mana dokumen yang memiliki nilai kedekatan tertinggi terhadap dokumen uji menggunakan algoritma KNN.
3. Evaluasi Sistem. Evaluasi akan dilakukan kepada sistem dengan menggunakan beberapa dokumen uji yang didapatkan secara acak dan sudah ditentukan kategorinya.

### 3.2. Hasil Evaluasi

Dari hasil pengujian dan evaluasi dengan 4 variasi nilai K dan 4 variasi nilai Feature Selection, Nilai yang didapatkan dapat dibandingkan untuk mengetahui titik mana nilai yang memiliki keakuratan yang paling tinggi.



Gambar 3. Grafik Perbandingan Nilai Akurasi

Dari keseluruhan tabel 1 dapat diketahui bahwa :

1. Rata-rata nilai akurasi tertinggi berdasarkan nilai K terdapat pada  $K = 5$  dan rata-rata akurasi terendah terdapat pada  $K = 3$ . Nilai *Akurasi* yang paling optimal terdapat pada nilai  $K = 5$  dengan Feature Selection 30% dan 50% dan  $K = 7$  pada Feature Selection 50%. Nilai *Akurasi* terkecil atau kurang optimal terdapat pada nilai  $K = 3$  dengan Feature Selection 10 %, 30% dan 50%.
2. Nilai Nilai Precision maupun nilai *Recall* pada  $K = 9$  tidak mengalami perubahan walaupun dilakukan variasi menggunakan nilai Feature Selection 10%, 30%, 50%, dan 100%.
3. Sistem klasifikasi komentar SMS dengan metode *K-Nearest Neighbor* memberikan nilai yang cukup tinggi memiliki nilai rata-rata precision yaitu 0,68 pada rentang nilai 0 – 1 dan memiliki kecenderungan mengklasifikasikan komentar Negative dengan pola naik pada nilai  $K = 5$  dan  $K = 7$  lalu turun pada  $K = 9$ .

#### 4. Kesimpulan

Berdasarkan Analisis dan implementasi sistem yang telah dibuat maka diperoleh kesimpulan bahwa algoritma *K-NN* dapat diimplementasikan ke dalam sistem untuk digunakan dalam mengklasifikasi sentimen positif atau sentimen negatif konsumen terhadap suatu produk yang berbentuk pesan SMS dalam layanan Suara Konsumen dan sistem klasifikasi komentar SMS dengan metode *K-NN* memberikan nilai akurasi yang cukup tinggi dengan nilai rata-rata 89% serta nilai keakuratan akan optimal ketika menggunakan  $K = 3$  dan  $K = 7$  yang nilai keakuratannya mencapai 90%.

#### 5. Penutup

Demikianlah penelitian yang telah dilakukan, penulisan ini yang diperuntukkan bagi penulis yang akan mengirimkan Tugas Akhir dan semoga bermanfaat bagi yang membacanya. Terima Kasih.

#### Daftar Pustaka

- El-Halees, A. (2011). Arabic Opinion Mining using Combined Classification Approach.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concepts and Techniques*. Morgan Kaufmann.
- Handoko, B. Y. (2008). Perbandingan Metode Naive Bayes Classifier (NBC) dengan K-Nearest Neighbor (*K-NN*) untuk Klasifikasi Dokumen.
- Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews.
- Kowalski, G. (2011). *Information Retrieval Architecture and Algorithms*. Ashburn: Springer.
- Kusumo, G. D. (2013). Klasifikasi Sentimen Pembaca berdasar Komentar terhadap Berita menggunakan Metode Naive Bayes Classifier.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining* . Morgan & Claypool Publishers.
- Liu, B. (2007). *Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data*. Chicago: Springer.
- Pang, B., Lee, L., & Vaithyanathan, S. (2003). Thumbs up? Sentiment Classification using Machine Learning.
- Wiebe, J., Bruce, R., & O'Hara, T. (1999). Development and use of a gold standard data set for. *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)* , 246–253.