Implementasi Metode Support Vector Machine (SVM) pada Model Prediksi Rating Obat Berdasarkan Ulasan Pasien

Wisnu Jayanata¹, Isman Kurniawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung
¹wisnuj@students.telkomuniversity.ac.id, ²ismankrn@telkomuniversity.ac.id

Abstrak

Pesatnya perkembangan internet dan media sosial mengakibatkan membaca ulasan sebelum membeli suatu produk, terutama produk obat menjadi hal yang lumrah. Namun, jumlah ulasan yang banyak dan tersebar mengakibatkan kesulitan dalam melakukan penilaian kualitas produk obat. Oleh karena itu, sistem yang dapat membantu pelanggan dalam menghadapi kendala ini sangat dibutuhkan. Pada pemodelan sistem, digunakan TF-IDF untuk mereduksi fitur dan Support Vector Machine sebagai metode klasifikasi. Dilakukan pula *Hyperparameter Tuning* untuk meningkatkan performa sistem. Pada penelitian ini didapat bahwa polynomial merupakan kernel SVM yang paling optimal untuk memprediksi rating obat berdasarkan ulasan pasien dengan akurasi mencapai 75.00% dan f-1 score sebesar 74.23%.

Kata Kunci: analisis sentimen, Support Vector Machine, TF-IDF, ulasan, obat

Abstract

Recent surges of internet and social media usage make the act of reading drug review before buying them become quite common. However, considering the huge amount of review scattered on the internet, it is quite hard to properly evaluate a drug. Therefore, a system that can help potential customer facing this problem is very much needed. The created model used TF-IDF to reduce the feature and Support Vector Machine is used as classification method. The writer has done a *Hyperparameter Tuning* to increase system performance. In this study, it is found that polynomial is the optimal SVM kernel to predict drug rating based on patient reviews with an accuracy up to 75.00% and f1-score of 74.23%.

Keywords: sentiment analysis, Support Vector Machine, TF-IDF, review, drug

1. Pendahuluan

1.1. Latar Belakang

Sentiment analysis atau bisa disebut *opinion mining*, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, perilaku dan emosi seseorang terhadap entitas seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka [1]. Perkembangan *Sentiment analysis* ini beriringan dengan berkembangnya teknologi, khususnya internet, yang mengakibatkan melimpahnya *user-generated data* seperti ulasan produk, ulasan layanan/jasa, dan penggunaan media sosial. *Sentiment analysis* sendiri telah diaplikasikan ke berbagai bidang, salah satunya adalah kesehatan. Berdasarkan Organisasi Kesehatan Dunia (WHO)[2], kesehatan dapat dideskripsikan sebagai kondisi kesejahteraan fisik, mental, dan sosial yang lengkap dan bukan sekadar tidak adanya penyakit atau kelemahan. Kondisi kesejahteraan ini sendiri bisa bervariasi antara individu yang mengalami gejala yang sama[3]. Subjektivitas pada kondisi kesehatan membuat kesehatan menjadi topik yang cocok untuk diteliti dengan *Sentiment Analysis*.

Beberapa penelitian sudah pernah dilakukan terkait analisis sentimen di bidang layanan kesehatan. Ahmet Emre Aladağ dkk melakukan penelitian terhadap pendeteksian niat bunuh diri pada forum[4]. Penelitian ini sendiri dilakukan di forum Reddit, dan sekitar 10.000 kiriman berbahasa Inggris digunakan sebagai dataset. Berbagai metode untuk ekstraksi fitur digunakan dan salah satunya adalah Term Frequency – Inverse Document Frequency (TF-IDF). Adapun metode klasifikasi yang digunakan salah satunya adalah Support Vector Machine (SVM). Akurasi yang didapat dari sistem yang dibangun dengan SVM lumayan tinggi dengan bisa mencapai 92%. Penelitian lain dilakukan oleh Abdulaziz dkk[5] untuk mengklasifikasikan kebutuhan pasien dari data yang didapat dari *Electronic Health Records (EHRs)*. Salah satu metode klasifikasi yang digunakan SVM dengan kernel yang digunakan adalah linear. Sistem yang dibangun mendapat akurasi sebesar 95%.

Namun, analisis sentimen di bidang kesehatan sayangnya masih belum cukup maju. Penelitian yang dilakukan oleh Anastazia Zunic dkk[6] menemukan bahwa *Sentiment Analysis* pada bidang kesehatan masih belum cukup maju dibandingkan dengan bidang lain, dengan penyebab yang masih dipertanyakan.

Dengan pesatnya perkembangan internet dan media sosial, ulasan-ulasan terhadap suatu obat bukanlah hal yang asing lagi. Ulasan-ulasan tersebut dapat mempengaruhi keputusan seorang calon pelanggan. Situs ulasan

dapat membantu calon pelanggan dalam mengambil keputusan dengan memberikan masukan mengenai obat yang akan mereka beli. Berdasarkan survey yang dilakukan oleh PowerReviews[7], lebih dari 99.9% konsumen mengatakan bahwa membaca ulasan saat melakukan belanja online, dan 98% konsumen merasa bahwa ulasan adalah salah satu faktor penting dalam mengambil keputusan berbelanja. Umumnya, ulasan obat terbagi menjadi dua bagian. Rating yang merepresentasikan tanggapan konsumen secara numerikal, dan komentar yang memberikan tanggapan konsumen dengan lebih detail. Namun, dengan banyaknya ulasan yang bisa dijumpai, tentunya calon pelanggan akan merasa kesulitan memnaca semua ulasan yang ada. Oleh karena itu, dibutuhkan sebuah sistem yang dapat membantu calon pelanggan dalam menghadapi kendala ini.

Penelitian ini bertujuan untuk membuat model prediksi ulasan obat berdasarkan ulasan pelanggan dengan menggunakan Support Vector Machine. Support Vector machine digunakan dalam pembangunan model dengan menggunakan kernel linear, RBF dan polynomial. Penelitian ini menggunakan SVM karena SVM memiliki keunggulan dalam mengklasifikasikan data linier dan non-linier melalui penggunaan kernel[8].

1.2. Rumusan Masalah

Berdasarkan pada latar belakang yang telah dipaparkan, dapat dirumuskan masalah-masalah berikut:

1. Bagaimana hasil dari seleksi fitur dataset rating obat berdasarkan ulasan pasien?
2. Seberapa besar peningkatan performa performa model prediksi rating obat berdasarkan ulasan pasien setelah dilakukan Hyperparameter tuning?
3. Kernel SVM mana yang memberikan performa terbaik untuk model prediksi rating obat berdasarkan ulasan pasien yang telah dibangun?

1.3. Tujuan

Berdasar pada rumusan masalah yang telah dibahas, berikut merupakan tujuan yang ingin dicapai pada penelitian:

1. Mengetahui hasil dari seleksi fitur dataset rating obat berdasarkan ulasan pasien
2. Mengetahui seberapa besar peningkatan performa performa model prediksi rating obat berdasarkan ulasan pasien setelah dilakukan Hyperparameter tuning
3. Mengetahui kernel SVM mana yang memberikan performa terbaik untuk model prediksi rating obat berdasarkan ulasan pasien yang telah dibangun

2. Metodologi

Dalam penelitian ini, dirancang sistem untuk memprediksi kepuasan pembeli obat berdasarkan ulasan. Alur pembangunan sistem diilustrasikan pada Gambar 1.