

Analisis Sentimen untuk Komentar pada Sistem Pencarian Kost Menggunakan Metode *Support Vector Machine* (SVM)

Fandy Oktavianus Suryanputra¹, Banu W. Yohanes², Saptadi Nugroho³

¹Program Studi Teknik Elektro,
Fakultas Teknik Elektronika dan Komputer,
Universitas Kristen Satya Wacana, Salatiga
¹fandyoktavianus@yahoo.co.id

^{2,3}Program Studi Sistem Komputer,
Fakultas Teknik Elektronika dan Komputer,
Universitas Kristen Satya Wacana, Salatiga
²banu.yohanes@staff.uksw.edu, ³saptadi_nugroho@yahoo.com

Ringkasan

Dengan terbatasnya informasi bagi mahasiswa dalam mencari kost, diperlukannya usaha dan waktu untuk mendapatkan informasi. Maka sistem pencarian kost akan sangat dibutuhkan untuk mempermudah dalam mencari kost. Atas dasar inilah diusulkan pembuatan sistem pencarian kost dengan sistem informasi rating dan rekomendasi kost di sekitar UKSW yang akan memudahkan dalam pencarian kost. Untuk sistem rating pada tulisan ini diajukan sistem analisis sentimen dengan menggunakan metode Support Vector Machine (SVM). Dari rating tersebut maka akan muncul rekomendasi – rekomendasi kost dengan rating tertinggi yang akan dimunculkan pada halaman utama web. Pengujian dilakukan pada 5 tempat kost yang berbeda yang akan dianalisa oleh SVM untuk menentukan termasuk kalimat positif atau negatif. Dari pengujian SVM tersebut menggunakan parameter $C = 2.5$ dan $\gamma = 0.3$ sehingga kost Turen II menghasilkan akurasi 92.30% , Dipo 88 mendapat akurasi 81.81%, Kemiri 2 mendapat 92.30%, Graha Widya mendapat 54.54%, dan Wisma Mawar mendapatkan akurasi 90.90%.

Kata kunci: SVM, rating, akurasi, rekomendasi

1. Pendahuluan

Bagi mahasiswa yang berasal dari luar pulau khususnya bagi calon mahasiswa, pencarian tempat kost merupakan sebuah masalah yang penting. Dengan terbatasnya informasi yang mereka punya pencarian kost menjadi sebuah tantangan bagi mereka. Cara tradisional diperlukannya usaha dan waktu untuk mendapatkan informasi yang belum tentu detail dan menyeluruh.

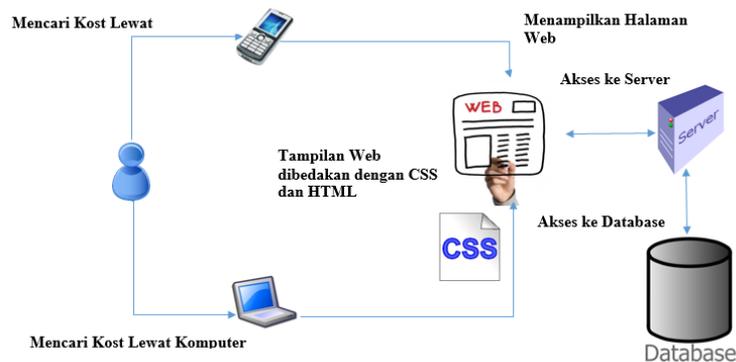
Hal ini menunjukkan cara tersebut tidak praktis. Oleh karena itu maka penulis mengusulkan sebuah sistem berbasis web yang dapat diakses dari berbagai perangkat. Sistem yang berisi tentang informasi mengenai lokasi, harga, serta fasilitas yang ditawarkan oleh tempat penyedia kost di sekitar kampus UKSW. Pada sistem web terdapat sistem rating yang berasal dari hasil pengolahan SVM. Di sini SVM berfungsi untuk menentukan apakah komentar pada tiap kost bersifat positif atau negatif. Kemudian hasil tersebut akan dibandingkan dengan dataset SVM yang sudah dibuat

sebelumnya. Dan menghasilkan akurasi. Akurasi dapat dioptimalkan dengan mengganti parameter C dan γ dengan yang paling optimal. Jika sudah dapat akurasi terbaik maka SVM dapat menentukan rating kost melalui jumlah komentar positif yang sudah diklasifikasi, sehingga nilai rating ini yang akan membantu para pencari kost. Dengan nilai rating yang baik maka kost tersebut akan tertampil pada halaman utama dengan rekomendasi.

2. Perancangan

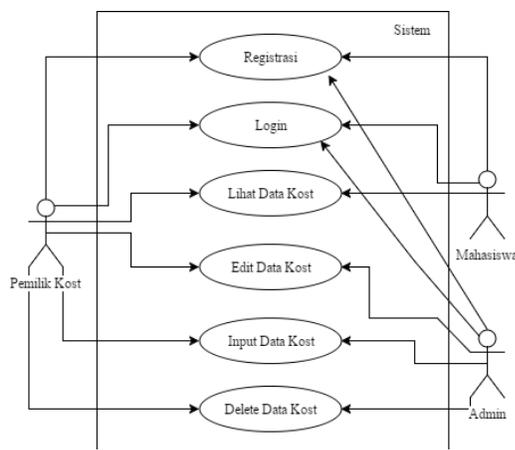
2.1. Perancangan Umum

Seperti yang telah dijabarkan sebelumnya, aplikasi ini akan dibuat berbasis web agar aplikasi ini dapat diakses di mana dan kapan saja selama pemakai terkoneksi dengan internet, tanpa perlu menginstall-nya. Dalam perancangan dan pengimplementasian aplikasi berbasis web ini, akan dipakai perangkat lunak Microsoft Visual Studio sebagai tools IDE dalam merancang halaman website dan Microsoft SQL Server sebagai tools dalam membuat database yang memuat informasi tempat kost. Kemudian untuk mempermudah dalam pencarian lokasi, aplikasi ini akan dilengkapi dengan google maps. Jadi pemakai aplikasi dapat lebih mudah dalam menentukan lokasi dari tempat kost.



Gambar 1 Arsitektur sistem yang dibangun

2.2. Diagram Use Case Sistem



Gambar 2 Diagram Use Case dari Sistem

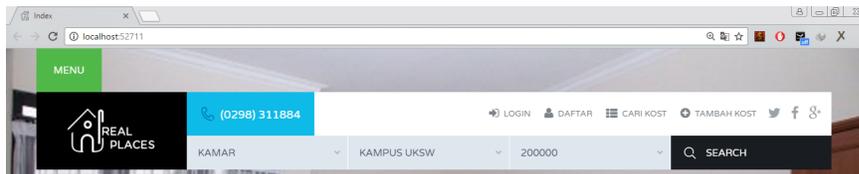
Diagram use case menggambarkan aktifitas yang dilakukan oleh suatu sistem dari sudut pandang pengamatan luar, dapat ditunjukkan seperti pada Gambar 2. Dalam sistem pengelolaan data kost terdapat 3 pengguna yang terlibat, antara lain Admin, Pemilik Kost dan Mahasiswa. Admin bertugas untuk mengecek keaslian dari data tiap kost dan dapat mengatur seluruh aktivitas web yang ada. Lalu pemilik kost bisa mempromosikan tempat kostnya dengan menginput data kost melalui persetujuan dari admin. Dan kemudian mahasiswa dapat melakukan pencarian kost dengan web yang sudah diisi data beberapa kost yang tersedia, kemudian jika mahasiswa yang sudah kost di tempat kost tersebut dapat melakukan review dengan memberikan komentar dan bintang.

2.3. Perancangan *Interface* Halaman Web

Berikut merupakan perancangan interface dari halaman web yang sedang dibuat:

a. Halaman utama pencarian kost

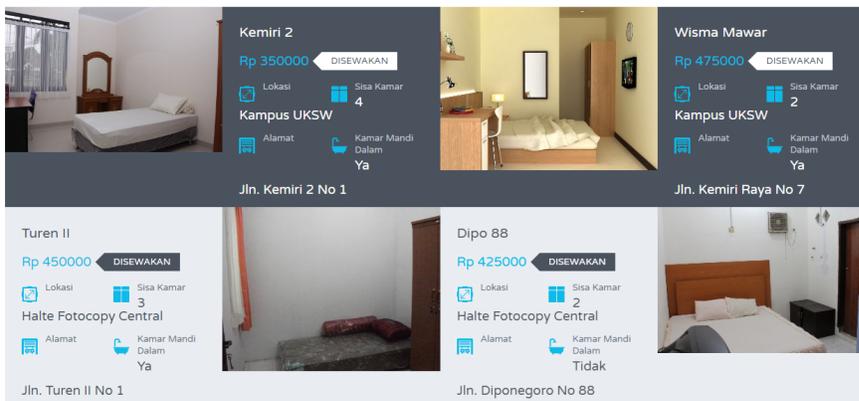
Halaman ini merupakan halaman yang tampil di dalam web browser ketika pertama kali user membuka alamat web tersebut. Halaman ini terdiri beberapa menu di antaranya di bagian header berisi menu Login, Daftar, Cari Kost, dan Tambah Kost, kemudian di bagian bawahnya terdapat menu pencarian kost berdasarkan jenis, lokasi dan harga yang dapat diubah sesuai dengan keinginan user. Ketika user klik tombol search, maka web akan menampilkan halaman web berupa daftar kost – kost yang diinginkan oleh user sesuai kategori yang diinginkan.



Gambar 3 Halaman Utama Pencarian Kost

b. Halaman utama rekomendasi

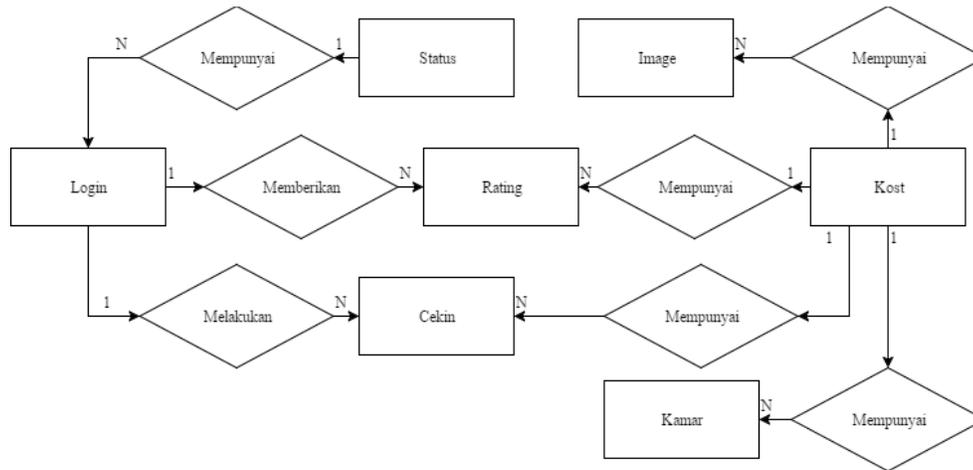
Halaman ini akan menampilkan beberapa kost yang memiliki rating tertinggi yang tersedia dalam web tersebut. Sehingga akan memudahkan user untuk memilih kost yang terbaik.



Gambar 4 Halaman Utama Rekomendasi

2.4. Perancangan Database MySQL

Database yang dirancang memiliki 7 tabel yaitu Login, Status, Rating, Kost, Cekin, Kamar dan Image. Tabel Login dirancang untuk menyimpan daftar informasi user yang tersedia. Tabel Status dirancang untuk membedakan tipe user berupa Admin, Pemilik Kost, dan Mahasiswa. Tabel Rating dirancang untuk menyimpan nilai rating dan komentar dari mahasiswa yang sudah melakukan review. Tabel Kost dirancang untuk menyimpan daftar kost yang tersedia. Tabel Kamar juga dirancang untuk menyimpan daftar kamar yang tersedia pada setiap kost yang terdaftar. Dan yang terakhir tabel Image dirancang untuk menyimpan gambar yang dimiliki setiap kost. Semua tabel tersebut memiliki relasi seperti pada Gambar 6.



Gambar 5. Diagram relasi secara umum

2.5. Perancangan Analisis Sentimen

Berikut ini adalah proses analisis sentimen yang digunakan dalam penelitian ini.

2.5.1. Pengumpulan Data

Penelitian dilakukan untuk komentar pada website pencari kost yang sudah dibuat. Data terbagi atas komentar bersifat positif dan bersifat negatif. Serta dipisah untuk data *training* dan *testing*.

2.5.2. Ekstraksi Fitur dan Pembobotan

Ekstraksi fitur dilakukan dengan pembobotan pada tiap kata di dalam sekumpulan komentar yang dimasukkan.

2.5.3. SVM

Kemudian SVM akan mengubah komentar menjadi data vektor. SVM mencoba untuk menemukan garis yang terbaik membagi dua kelas, dan kemudian mengklasifikasikan komentar uji berdasarkan di sisi mana dari garis tersebut mereka muncul.

2.5.4 Training

Pada tahap training data – data yang berasal dari komentar uji akan dicek dengan database set dari SVM sehingga didapatkan nilai akurasi, yang kemudian hasilnya akan digunakan untuk mencari nilai parameter C dan gamma yang optimal.

2.5.5 Testing

Pada tahap testing data – data yang berasal dari komentar baru akan dicek dengan training set apakah sesuai atau tidak, yang kemudian hasilnya akan diolah menjadi rating.

3. Pengujian

3.1. Pengujian Akurasi SVM

Tabel 1 Tabel Hasil Pengujian Akurasi SVM

Nilai Parameter C dan γ	Akurasi SVM
C = 1 dan $\gamma = 1$ (default)	Turen II = 84.61% Dipo 88 = 27.27% Kemiri 2 = 30.76% Graha Widya = 63.63% Wisma Mawar = 36.36%
C = 1 dan $\gamma = 0.3$	Turen II = 69.23% Dipo 88 = 27.27% Kemiri 2 = 15.38% Graha Widya = 45.45% Wisma Mawar = 18.18%
C = 1.5 dan $\gamma = 0.3$	Turen II = 92.30% Dipo 88 = 45.45% Kemiri 2 = 38.46% Graha Widya = 63.63% Wisma Mawar = 36.36%
C = 2.5 dan $\gamma = 0.3$	Turen II = 92.30% Dipo 88 = 81.81% Kemiri 2 = 92.30% Graha Widya = 54.54% Wisma Mawar = 90.90%
C = 3.5 dan $\gamma = 0.3$	Turen II = 92.30% Dipo 88 = 81.81% Kemiri 2 = 92.30% Graha Widya = 54.54% Wisma Mawar = 90.90%
C = 2 dan $\gamma = 0.6$	Turen II = 92.30% Dipo 88 = 81.81% Kemiri 2 = 92.30% Graha Widya = 54.54% Wisma Mawar = 90.90%

Tujuan dari pengujian di Tabel 1 adalah untuk melihat nilai akurasi pada klasifikasi komentar dalam menentukan komentar tersebut positif atau negatif. Akurasi diukur dari perbandingan antara data training atau data yang sudah dilabel sebelumnya dengan hasil klasifikasi dengan menggunakan pembelajaran SVM.

Pada Tabel 1 dilakukan percobaan mencari parameter terbaik dengan rentang nilai $1 \leq C \leq 10$ dan $0.1 \leq \gamma \leq 10$. Karena untuk parameter lebih atau kurang dari nilai tersebut menghasilkan akurasi yang sama. Sehingga hanya digunakan beberapa kemungkinan pada parameter tersebut.

Padapengujian di atas didapat nilai akurasi terbaik terdapat pada C = 2.5, $\gamma = 0.3$, C = 3.5, $\gamma = 0.3$ dan C = 2, $\gamma = 0.6$ dengan nilai akurasi tertinggi 92.30% dan nilai terendah 54.54%. Nilai akurasi 54.54% kecil dikarenakan kata – kata yang terdapat pada komentar kost Graha Widya pada saat proses training memiliki kelas yang berbeda dengan yang

diujikan. Contoh pada komentar “Harga kost mahal dan banyak serangga”, kata banyak dalam proses training dikategorikan sebagai kata positif, padahal kata banyak pada komentar tersebut memiliki arti yang negatif. Sehingga dapat disimpulkan bahwa sistem masih belum bisa menentukan polaritas sebuah komentar yang berisi perbandingan.

3.2. Pengujian SVM (C = 2.5 γ = 0.3)

Tabel 2 Tabel Hasil Pengujian pada Kost Turen II

Komentar (Rating)	Komentar(Data)
Kost ini <u>terbaik</u> (4)	+1 2:1
Kost lumayan <u>baik baik</u> (3)	+1 2:2
kost ini <u>jelek jelek</u> (3)	-1 1:2
Kost ini <u>nyaman</u> dan <u>dekat</u> dengan kampus selain itu juga banyak tempat <u>makan</u> disekitarnya (4)	+1 9:1 10:1 21:1 -1 2:1 12:1 13:1 32:1
Rumah Kost disini dilihat dari sisi lingkungan sosial <u>tidak</u> terlalu <u>baik</u> , karena banyak aktivitas <u>kenakalan</u> remaja disekitar kost. Sehingga membuat <u>risih</u> . (2)	-1 12:1 15:1 29:1 34:1 +1 2:1 3:1 11:1 16:1
Ibu Kost disini <u>galak</u> selain itu juga ada <u>peraturan</u> untuk <u>tidak</u> boleh membawa teman. Dan ada <u>jam</u> malam. (1)	+1 9:1 17:1 +1 10:1 18:1 21:1
Fasilitas di sini lumayan <u>bagus</u> , <u>kebersihan terjaga</u> dan ibu kost disini <u>baik</u> (4)	-1 12:1 24:1 33:1 34:1
Kost disini <u>nyaman enak</u> untuk belajar (5)	+1 16:1
Tempat kost <u>dekat</u> dengan kampus dan untuk cari <u>makan mudah</u> (3)	+1 10:1 16:1 19:1 20:1
Di dalam kost terdapat rumah ibu kos sehingga <u>tidak bebas</u> banyak <u>peraturan</u> dan parkir <u>sempit</u> (2)	+1 4:1 9:1
Kost ini <u>bersih</u> (3)	
<u>Enak</u> buat kumpul, kamar mandi <u>bersih</u> , internet <u>lancar</u> , parkir <u>luas</u> dan dekat kampus. (3)	
Kost ini <u>nyaman</u> dan <u>indah</u> . (4)	

Pada pengujian SVM dengan menggunakan parameter C = 2.5 dan parameter γ = 0.3 menghasilkan akurasi tertinggi yaitu 92.30% dan terkecil 54.54% dan rating tertinggi 4.615 dan terkecil 3.076. Hasil tersebut didapat dengan menggunakan proses training sebanyak 15 data dan testing sebanyak 59 data.

Kemudian saat dibalik data training menjadi 59 data dan data testing menjadi 15 data, didapat akurasi sebesar 84.61% di kost Turen II dengan rating SVM 4.230, pada kost Dipo 88 mendapatkan akurasi 90.90% dengan rating SVM sebesar 5, kemudian pada kost Graha Widya menghasilkan akurasi 100% dengan rating SVM sebesar 2.72, dan pada kost Wisma Mawar mendapatkan akurasi 91.67 dengan rating SVM sebesar 4.583.

Komentar – komentar yang berasal dari data kost Turen II yang berbentuk kata akan diubah menjadi bentuk data yang dapat diproses oleh SVM. Perubahan bentuknya berdasarkan kemunculan kata yang terdapat dalam database.

“Kost lumayan baik baik”.

Kata baik memiliki ID 2.

Baik muncul 2 kali sehingga menjadi 2:2.

Baik memiliki bobot +1 dan dominan di kalimat tersebut.

Sehingga menjadi +1 2:2

Untuk akurasi tiap kost didapat dari SVM yang mengecek kecocokan antara data yang diuji yang berasal dari komentar dengan data yang sudah diset sebelumnya. Dan akurasi juga didapat dari sejumlah data yang memiliki akurasi terbaik sehingga didapat parameter C = 2.5 dan γ yang bernilai 0.3 yang paling optimal dalam sistem ini.

Sedangkan untuk rating bintang didapat dari hasil rata – rata dari user yang telah memberikan review pada kost tersebut.

$$\text{Rating Bintang} = \frac{\text{Total Bintang}}{\text{Total Komentar}} = \frac{41}{13} = 3.153 \quad (1)$$

Dan untuk rating dari SVM sebuah perhitungan. Rating dibuat dalam skala 5.

$$\text{Rating SVM} = \frac{\text{Total Komentar Positif}}{\text{Total Komentar}} \times 5 = \frac{8}{11} \times 5 = 3.076 \quad (2)$$

4. Kesimpulan

Dengan membuat sistem menggunakan data kost rekaan akan membantu mahasiswa maupun orang lain dalam mencari kost di daerah UKSW berdasarkan kategori jenis, lokasi, dan harga serta memberikan informasi tentang kost yang diinginkan. Dan sistem ini akan membantu para pemilik kost untuk mempromosikan kost miliknya.

Dari pengujian SVM pada 5 tempat kost berbeda dengan menggunakan parameter $C = 2.5$ dan $\gamma = 0.3$ didapat rating tertinggi sebesar 4.615 pada kost Kemiri 2 dan yang terkecil sebesar 3.076 pada kost Turen II dan untuk akurasi didapat pada tempat kost Turen II menghasilkan akurasi 92.30% , Dipo 88 mendapat akurasi 81.81%, Kemiri 2 mendapat 92.30%, Graha Widya mendapat 54.54%, dan Wisma Mawar mendapatkan akurasi 90.90%. Terdapat akurasi yang bernilai kecil yaitu 54.54% dikarenakan sistem masih belum bisa menentukan polaritas sebuah komentar yang berisi perbandingan.

Daftar Pustaka

- [1] Yosanny, A., & Halim, A. Pengembangan Aplikasi Mobile untuk Penyediaan Informasi Sekitar Kampus Binus Berbasis Web. *Jurnal ComTech*, 4(02), 2013.
- [2] Maricha, Erry. Perancangan Aplikasi Mobile Pencarian Lokasi Penyewaan Rumah Kost Berbasis Android Di Kota Yogyakarta, 2013.
- [3] Nugroho, AntoSatriyo., Arief Budi Witarto, danDwiHandoko.Support Vector Machine, TeoridanAplikasinyadalamBioinformatika, 2003
- [4] Liu, B. *Opinion Mining*. Chicago, United States of America, 2012.
- [5] Triawati, C. *Text Mining*. Bandung, Jawa Barat, Indonesia, 2009.
- [6] Sulistyawan, Rubianto, Rahmad Saleh, Modifikasi Blog Multiply dengan CSS, halaman 32. *Elex Media Komputindo*
- [7] Andi Sunyoto, M. Kom, Ajax Membangun Web denganTeknologiAsynchronouse JavaScript & XML, ANDI.
- [8] Sholiq. *Pemodelan Sistem Informasi Berorientasi Objek dengan UML*.Yogyakarta: Graha Ilmu, 2006.
- [9] M. Nur dan D. Santika, "Analisis Sentimen pada Dokumen Berbahasa Indonesia dengan Pendekatan Support Vector Machine," dalam *Konferensi Nasional Sistem dan Informatika*, 2011.

