

METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE UNTUK MENGANALISIS SENTIMEN PADA MEDIA SOSIAL SEBAGAI REKOMENDASI OLEH-OLEH FAVORIT

Hidayatulah Himawan¹⁾, Debby Gybson Putri D.D.R²⁾, Wilis Kaswidjanti³⁾

^{1,2,3)}Program Studi Teknik Informatika UPN "Veteran" Yogyakarta

Jl. Babarsari no 2 Tambakbayan 55281 Yogyakarta Telp (0274)-485323

e-mail : if.iwan@upnyk.ac.id¹⁾, debbygybsonpoetri@gmail.com²⁾, wilisk@upnyk.ac.id³⁾

Abstrak

Media sosial pada saat ini mengalami perkembangan yang semakin pesat. Pengguna media sosialpun setiap hari semakin meningkat. Media sosial sendiri dapat dijadikan sebagai sumber informasi. Salah satu informasi yang didapatkan dari media sosial adalah opini pengguna media sosial pada suatu hal. Opini-opini tersebut dapat bernilai positif, negatif atau netral. Salah satu opini yang dapat dimanfaatkan adalah opini mengenai oleh-oleh favorit khususnya yang berada di daerah Yogyakarta. Opini tentang oleh-oleh tersebut dapat dianalisis sehingga menghasilkan sebuah rekomendasi oleh-oleh favorit. Oleh karena itu, dibutuhkan analisis sentimen yang dapat menilai sentimen dari opini yang disampaikan pengguna media sosial tersebut. Analisis sentimen adalah ilmu yang mempelajari bagaimana proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Pada penelitian ini, proses analisis sentimen menggunakan algoritma Lexicon Based dan Support Vector Machine. Data yang digunakan adalah data sentimen media sosial Twitter dan Instagram yang diambil melalui layanan API media sosial tersebut. Data diambil sebanyak 1000 data sentimen untuk data latih. Sedangkan data uji untuk proses pengujian sebanyak 50 data sentimen. Dari hasil pengujian didapatkan hasil akurasi terbesar adalah menggunakan sebesar Lexicon Based 87.78%, kemudian hasil presisi terbesar adalah menggunakan Lexicon Based sebesar 94.23%, sedangkan untuk hasil recall terbesar adalah menggunakan Support Vector Machine sebesar 100%.

Kata kunci : Rekomendasi Oleh-Oleh, Analisis Sentimen, Media Sosial, Twitter, Instagram, Lexicon Based, Support Vector Machine.

1. PENDAHULUAN

Media sosial adalah sebuah media daring, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi [blog](#), [jejaring sosial](#), wiki, forum dan [dunia virtual](#). Media sosial punya punya kekuatan super untuk mempengaruhi banyak orang melaui opini-opini yang disampaikan di media sosial. Opini merupakan pendapat pribadi dari pengguna media sosial mengenai suatu hal atau kejadian. Pengguna media sosial sering memberikan opini mengenai tempat wisata, kuliner dan tak terkecuali opini tentang oleh-oleh khas berupa makanan ataupun pernak-pernik dari suatu daerah melalui Twitter dan Instagram. Opini tersebut dapat diolah sehingga menghasilkan rekomendasi oleh-oleh khususnya di Kota Jogja. Opini yang terdapat di Twitter dan Instagram dianalisis yang disebut dengan analisis sentimen. Analisis sentimen adalah proses yang digunakan untuk menentukan opini, emosi, dan sikap yang dicerminkan melalui teks, dan biasanya diklasifikasikan menjadi opini positif dan opini negative (Coletta dkk, 2014). Tujuan dari analisis sentimen adalah bagaimana komputer dapat mengenali tingkat emosional positif maupun negatif dari dokumen teks (Mejova 2009). Pada proses analisis sentimen, dilakukan penerapan algoritma *machine learning* yang dapat membantu proses klasifikasi. Algoritma machine learning yang digunakan pada penelitian ini adalah algoritma *Support Vectore Machine* (SVM) dan *Lexicon Based*.

Oleh karena itu, pada penelitian kali ini yang akan dilakukan analisis sentimen pada opini berupa sentimen yang didapatkan dari media sosial Twitter dan Instagram menggunakan metode SVM dan Lexicon Based sehingga dapat diterapkan untuk mengetahui oleh-oleh mana yang memiliki review positif yang paling banyak dari masyarakat atau para wisatawan yang pernah membeli oleh-oleh tersebut, serta perbandingan penggunaan metode *Lexicon Based* dan *Support Vector Machine*(SVM) yang tidak dikombinasikan, dan membandingkan hasil akurasi performa pengujian sistem yang dihasilkan dari masing-masing metode sehingga dapat diketahui performa mana yang lebih unggul dari kedua metode tersebut.

2. TINJAUAN PUSTAKA

Media sosial atau jejaring sosial merupakan sebuah layanan berbasis *website* yang memungkinkan pengguna untuk membuat profil pribadi yang dapat diakses publik atau pun profil pribadi yang bersifat publik dengan pembatasan akses yang dapat diatur oleh pengguna (Boyd dan Ellison , 2007). Penelitian yang dilakukan oleh Rabson dkk, (2008) menemukan bahwa informasi yang disampaikan orang dalam media sosial *online* dapat berupa rekomendasi atau hal-hal yang mungkin bisa menguatkan prinsip seorang pembaca atau malah mempengaruhinya. Oleh karena itu kebenaran kualitas informasi dapat memberikan efek pengaruh yang kuat

bagi seorang pembaca. Penelitian lainnya yaitu memanfaatkan media sosial sebagai media informasi dan komunikasi untuk Usaha Kecil Menengah (UKM). Jenis media sosial yang digunakan oleh subyek penelitian ini yaitu *website*, Facebook, Twitter, dan YouTube. Penggunaan media sosial pada UKM memberikan manfaat diantaranya sebagai sarana kontak personal dengan konsumen, sebagai sarana promosi atau *advertising*, untuk mendata kebutuhan konsumen, sebagai media untuk menyampaikan respon ke konsumen dan sebagai dasar pengambilan keputusan bisnis. Di samping itu, media sosial juga bermanfaat sebagai forum diskusi *online*, memantau pelanggan secara *online*, survei pelanggan, mendata kebutuhan penyalur, mendata kebutuhan pemasok serta untuk menampilkan galeri produk. Hasil yang didapatkan pada penelitian ini yaitu media sosial merupakan sarana komunikasi efektif yang dirasakan oleh UKM. Media sosial juga dapat meningkatkan pangsa pasar. Penggunaan media sosial oleh UKM dapat meningkatkan volume penjualan hingga lebih dari 100% bila dilakukan *update* informasi setiap hari dan secara konsisten (Priambada, 2015).

Twitter adalah sebuah media sosial dan layanan *microblogging* yang diciptakan oleh Jack Dorsey ditahun 2006 untuk difokuskan terhadap kebutuhan komunikasi yang cepat. *Microblogging* sendiri memiliki arti sebuah situs dimana penggunaanya dapat menyiarkan pesan atau tulisan dalam ukuran lebih kecil. *Tweet* adalah sebuah pesan pendek dengan panjang karakter yang dibatasi hanya sampai 140 karakter. Dikarenakan keterbatasan karakter yang bisa dituliskan, sebuah *tweet* seringkali mengandung singkatan, bahasa *slang* maupun kesalahan pengejaan (Bennett dan Johnson 2014). Penelitian sebelumnya mengenai analisis media sosial Twitter dilakukan oleh Cha dkk, (2010) untuk mengukur tingkat pengaruh pengguna pada Twitter melalui *follower*, *retweet* dan *mention* pengguna. Penelitian serupa juga dilakukan oleh Bakshy dkk, (2011) dalam "Everyone's an influencer : *quantifying influence on Twitter*" menyimpulkan bahwa semua orang adalah pembawa pengaruh. Penelitian sebelumnya tentang Twitter juga dilakukan oleh Monarizqa (2014) dengan judul "Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating". Penelitian ini bertujuan untuk melakukan analisis sentimen pada Twitter berbahasa Indonesia sebagai pemberi rating. Terdapat dua tahap pada penelitian ini, pertama membentuk *core* aplikasi berupa konfigurasi metode analisis sentimen dan tahap kedua adalah membuat aplikasi analisis sentimen. *Core* aplikasi dikembangkan menggunakan pustaka *Lucene* sebagai alat *stemming* dan algoritme *Support Vector Machine* (SVM). Nilai akurasi dari konfigurasi algoritma SVM dengan *preprocessing* yang diterapkan penelitian ini untuk analisis sentimen pada teks berupa *tweet* berbahasa Indonesia adalah sebesar 73.43%. Konfigurasi ini kemudian menjadi *core* aplikasi analisis sentimen. Aplikasi analisis sentimen dibangun dengan menggunakan Twitter API untuk mengambil data real-time, pustaka *Lucene* sebagai alat *stemming* pada *core* aplikasi, *SVMLight*, dan Java. Ketika konfigurasi diterapkan dalam aplikasi analisis sentimen yang dibuat, nilai akurasi ketika menggunakan kata kunci "Jokowi" sebesar 68%, "Prabowo" sebesar 56%, "kalimilk" sebesar 70% dan "sunmor" sebesar 74%. Penelitian yang dilakukan oleh Kurniawati dan Edi Winarko (2016) dengan judul "Penentuan Destinasi Wisata Favorit Berbasis Aturan Dan Analisis Sentimen Pada Tweet Berbahasa Indonesia". Pada penelitian ini dibuat sistem yang dapat menentukan destinasi wisata favorit di suatu kota, memberikan informasi tren masyarakat berkaitan dengan pariwisata melalui *hashtag* favorit, memberikan informasi berupa opini masyarakat maupun wisatawan terhadap pariwisata melalui analisis sentimen, serta memberikan informasi mengenai waktu favorit pengguna Twitter di kota tersebut. Penelitian ini mengembangkan leksikon sentimen untuk pariwisata berbasis aturan dengan pendekatan kamus. Pada penelitian ini menggunakan metode *Lexicon Based* untuk mengolah informasi sentimen berbahasa Indonesia. Dalam pengujian performa sistem dalam mengenali *tweet* pariwisata dilakukan tiga kali percobaan dengan tiga *search query* berbeda yaitu #Yogyakarta, #Bali, dan #Jakarta. Hasilnya didapat rata-rata *recall* 73,1%, *precision* 100%, *f-measure* 84,1%, dan akurasi 93,2%. Akan tetapi penelitian ini masih menggunakan media Twitter dan hanya terbatas dengan *tweet* yang berbahasa Indonesia saja sebagai obyek dalam pengambilan data.

Penelitian yang dilakukan Puspitarani (2015) mengenai analisis sentimen instagram dilakukan oleh yaitu memanfaatkan *sentimen analysis* terhadap komentar-komentar pada foto dan *caption* di akun Instagram *online shop* yang bersangkutan. Proses *sentiment* yang dilakukan dengan menggabungkan pendekatan *lexicon* dan *machine learning*. Data yang digunakan adalah komentar-komentar terhadap foto dan *caption* beberapa *online shop* di Instagram. Dengan menggunakan model tersebut, didapatkan hasil klasifikasi sentimen menghasilkan akurasi yang sangat baik.

3. METODOLOGI PENELITIAN

Metodologi yang digunakan pada penelitian ini adalah kuantitatif. Tahapan yang dilakukan dalam penelitian ini: Studi Literatur, Analisis Sistem, Perancangan Sistem, Penulisan Program, Pengujian Sistem, dan Penyusunan Laporan (gambar 1).



Gambar 1 Tahapan Metodologi Penelitian

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah metode *waterfall* yang telah dikembangkan oleh Pressman (Pressman, 2012) dengan tahapan: Analisa Kebutuhan, Desain Sistem, Penulisan Kode Program, dan Pengujian Program.

3.1 Analisis Sistem

Data yang dibutuhkan sebagai bahan analisis penelitian yaitu data sentimen yang berasal dari tweet dan caption dari media sosial Twitter dan Instagram. Data sentimen didapatkan dengan meng-*crawling* data Twitter dan Instagram untuk mendapatkan sentimen yang dijadikan bahan penelitian. Proses *crawling* data ini memanfaatkan Twitter *streaming* API dan Instagram *explore*.

Analisis Kebutuhan

Adapun analisis kebutuhan yang akan digunakan untuk mengembangkan sistem ini adalah sebagai berikut :

1. Data input: Data Admin, Data Setimen Twitter dan Instagram (Data Uji), Data Stopword, Data Kamus *Lexicon*, Data kata dasar (Dictionary), dan Data Kata Sinonim.

Stopword merupakan kata-kata umum dan sering digunakan pada suatu bahasa namun kata-kata umum tersebut tidak mempunyai pengaruh dalam klasifikasi teks, seperti: “untuk”, “kapan”, “yang”, “di”, “iya”, “itu”, dan lain sebagainya. Data kamus *stopword* diperoleh dari penelitian (Tala, 2003). *Sentiment Dictionaries* atau kamus sentiment merupakan kumpulan kata-kata yang mengandung kata kunci positif dan kata kunci negatif yang pada penelitian ini digunakan sebagai kamus *lexicon*. *Sentiment Dictionaries* tersebut di peroleh dari penelitian (Rofiqoh, Perdana, & Fauzi, 2017) yang sudah memiliki kata yang mengandung positif dan negatif. Data kata dasar didapat dari kamus bahasa Indonesia *online*. Data kata dasar meliputi id kata dasar, id kata, *stopword* kata dasar, sentimen kata dasar, dan status kata dasar. Kata dasar diperoleh dari penelitian (Saragih, 2018). Data Sinonim merupakan beberapa kata yang mempunyai arti sama atau hampir sama. Pada penelitian ini data sinonim berisi kumpulan kata-kata yang disusun berdasarkan kata yang memiliki makna yang hampir sama. Sehingga dapat mempercepat dalam proses *preprocessing* kata. Data sinonim diperoleh dari kamus pada web <http://bsd.pendidikan.id>.

2. Data Proses: Proses *Text Preprocessing*, Proses Ekstraksi Fitur, Proses Latih dan Uji Data

3. Data Output: Data komentar bersih hasil *preprocessing* yang siap untuk diklasifikasikan dan diolah untuk penentuan opini atau sentiment, Statistik berisi jumlah data pada setiap kategori sentimen berupa grafik, dan Pengujian opini atau sentimen yang terdiri dari akurasi opini atau sentimen terhadap sistem yang telah dibangun.

Analisis Sentimen

Tahapan analisis sentimen yang akan digunakan pada sistem ini adalah : Analisis Teks *Preprocessing*, *Cleansing*, *Tokenizing*, Pembuangan *Stopword* (*Stopword Removal*), dan *Stemming*. Tahap pelatihan dilakukan

proses ekstraksi fitur pada komentar yaitu dengan menggunakan *Term Frequency* dan *TF-IDF (Term Frequency-Inverse Document Frequency)*. Langkah-langkah dalam tahap ini: Hitung *Term Frequency* dan Hitung *TF-IDF (Term Frequency-Inverse Document Frequency)*.

Analisis Penerapan Support Vector Machine (SVM)

Pengklasifikasian menggunakan *Support Vectore Machine* dibagi kedalam dua proses, yaitu proses *training* (latih) dan *testing* (uji). Proses *training* digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentimen dengan data *testing* atau data mentah yang baru. Langkah-langkah algoritma klasifikasi sentimen menggunakan *Support Vectore Machine*:

1. Pelatihan. Pada proses pelatihan akan dilakukan dengan menggunakan fitur *TF-IDF*. Menurut Habibi (2017) langkah-langkahnya adalah sebagai berikut: (1) Hitung bobot dan faktor bias untuk bidang pembatas; (2) Langkah selanjutnya, data komentar diubah formatnya kedalam format support vector; (3) Kernelisasi menggunakan fungsi *kernel linear*; (4) Pencarian nilai Lagrance Multiplier; (5) Perhitungan untuk data uji.
2. Perhitungan untuk data uji.

Proses selanjutnya ialah proses klasifikasi pada *support vector machine* dalam penentuan kelas menggunakan fungsi keputusan $f(x)$ yaitu fungsi $sign()$.

Analisis Penerapan Lexicon Based

Tahapan dalam penentuan kategori sentimen pada media sosial Twitter dan Instagram menggunakan *Lexicon Based*.

1. Membangun kamus sentimen yang terbagi menjadi sentimen positif dan negatif. Dalam penelitian ini kamus sentimen di peroleh dari penelitian Rofiqoh et al., n.d. kemudian ditambah secara manual sesuai dengan kebutuhan peneliti.

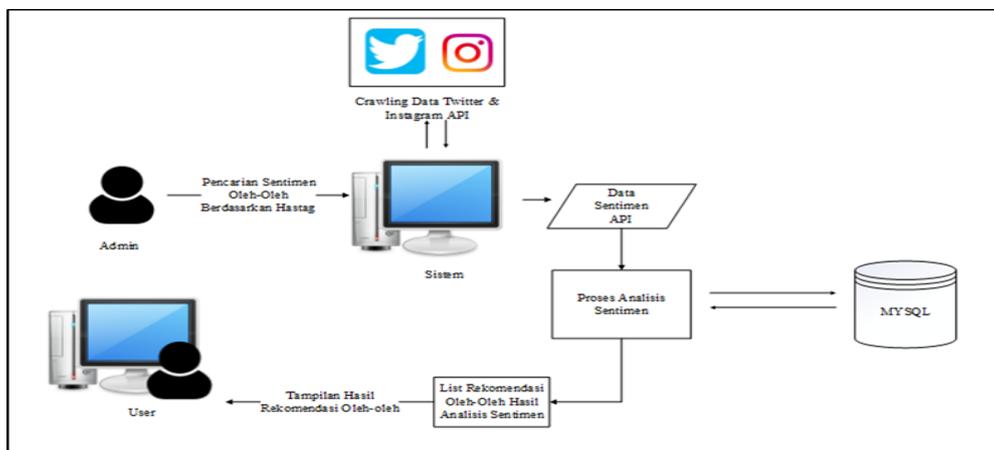
2. Preprocessing Text

Tahapan preprocessing pada analisa *lexicon* memiliki perbedaan pada analisa menggunakan *support vector machine*. Hal yang membedakan adalah tidak dilakukannya tahapan pembobotan kata atau ekstraksi fitur. Secara keseluruhan tahapan yang dilakukan adalah: *Cleansing, Tokenizing, Stopping (Stopword Removal)*, dan *Stemming*.

3.2 Perancangan Sistem

Arsitektur Sistem

Pada arsitektur sistem yang akan dibangun terdiri dari beberapa komponen yaitu terdapat admin yang akan melakukan pencarian sentimen dengan menginputkan hastag nama oleh-oleh pada sistem. Data-data yang telah diinputkan akan dikumpulkan pada proses pengumpulan data dengan menggunakan API yang telah disediakan oleh layanan media sosial tersebut. Hasil dari pengumpulan data yang merupakan data teks selanjutnya akan masuk ke tahap preprocessing. Pada tahapan ini data teks adalah tahapan untuk membersihkan serta melakukan penyeragaman kata sehingga kata tersebut siap untuk di ekstraksi ketahap berikutnya . Pada tahap selanjutnya teks tersebut akan diklasifikasikan dengan menggunakan *Lexicon Based* dan *Support Vector Machine* dan akan menghasilkan data teks yang memiliki kelas sentimen positif, negatif dan netral. Data teks yang telah diklasifikasikan kemudian digunakan untuk pembobotan dalam penentuan daftar rekomendasi oleh-oleh yang selanjutnya akan ditampilkan berupa list rekomendasi terbaik berdasarkan analisis sentimen. Gambar 2 menunjukan arsitektur sistem secara keseluruhan.



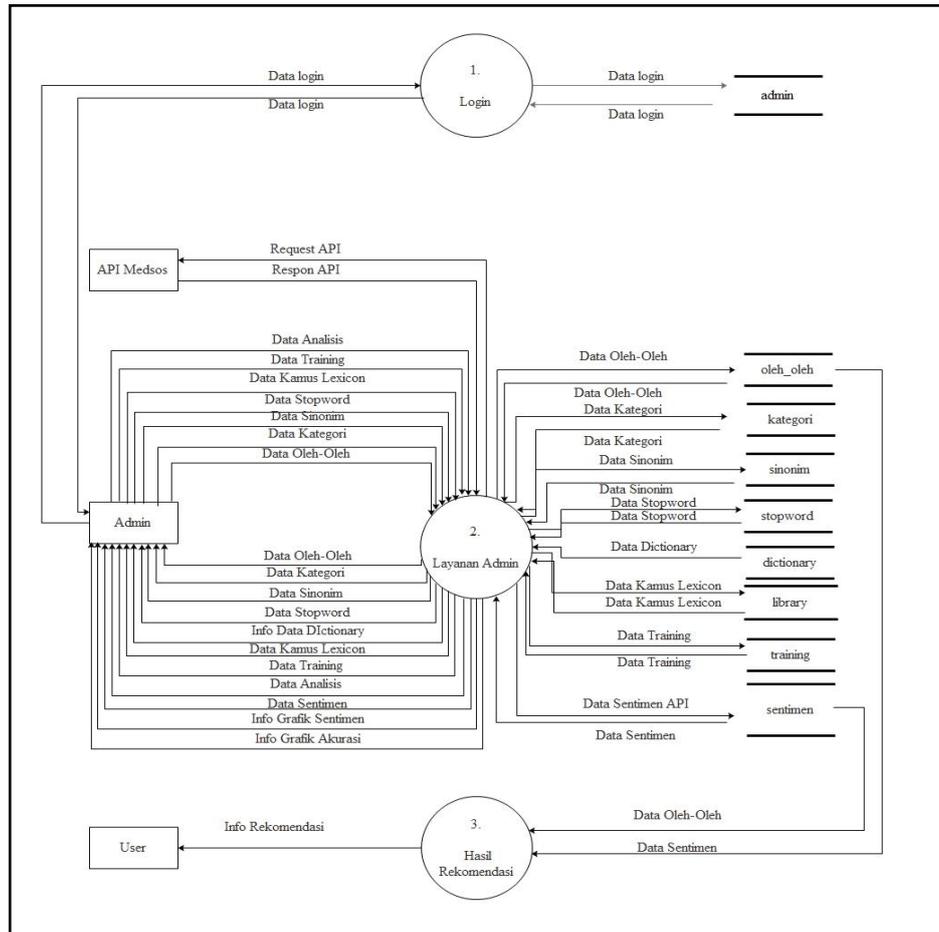
Gambar 2 Arsitektur Sistem

Perancangan Proses

Berikut ini merupakan tahapan dalam perancangan proses berupa DFD dan *Flowchart*.

DFD Level 1

DFD Level 1 dapat dilihat pada gambar 3.

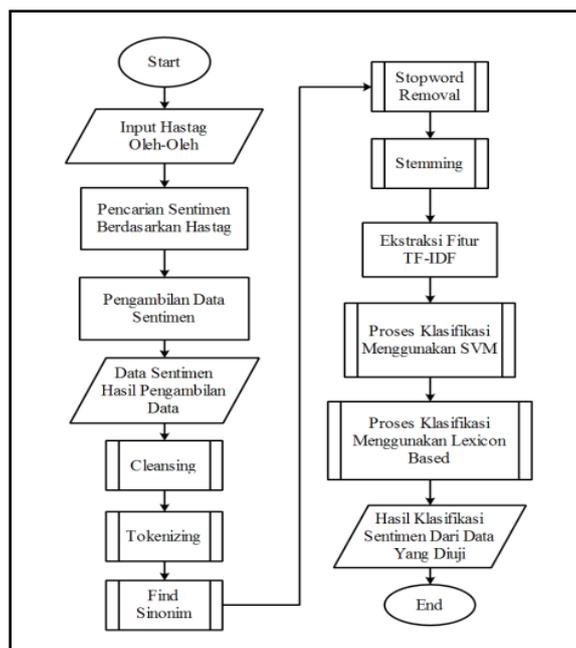


Gambar 3 DFD Level 1.

Proses layanan admin ini hanya untuk mengolah seluruh data yang berada pada *data store*, yaitu Kelola Data Oleh-Oleh, Kelola Data Kategori, Kelola Data Sinonim, Kelola Data *Stopword*, Tampil Data *Dictionary*, Kelola Kamus *Lexicon*, Kelola Data *Training*, Pengambilan Data Sentimen API dari Twitter API dan Instagram API, Pengklasifikasian Sentimen, Halaman Analisis, Tampil Grafik Sentimen dan Tampil Grafik Akurasi. Pada proses pengambilan data API membutuhkan entitas Medsos API yang selanjutnya data dari sentimen akan disimpan pada database sentimen. Selanjutnya akan masuk ke proses pengklasifikasian sentimen yang membutuhkan data dari database *dictionary*, sinonim, stopword, lexicon dan training yang selanjutnya hasil tersebut akan disimpan kembali pada database sentimen. Pada entitas API Medsos terdapat aliran *request* dan *respon* API pada sistem dan *user* terdapat aliran data yang keluar info rekomendasi.

Flowchart Analisis Sentimen

Pada *flowchart* analisis sentimen ini merupakan gambaran alur semua proses pada sistem ini. Pada proses analisis sentimen tahap pertama admin akan menginputkan hashtag sebagai keyword untuk mengambil data sentimen oleh-oleh pada sistem dari Twitter API dan Instagram API. Mekan selanjutnya pada sistem akan dilakukan pencarian data sentimen secara real-time sesuai dengan hashtag yang telah diinputkan. Setelah sentimen ditemukan maka hasil dari pengumpulan data sentimen tersebut yang merupakan data teks selanjutnya akan masuk ke tahap *preprocessing*. Pada tahapan ini data teks adalah tahapan untuk membersihkan serta melakukan penyeragaman kata sehingga kata tersebut siap untuk di ekstraksi ketahap berikutnya. Tahap selanjutnya yaitu pembobotan fitur TF-IDF, pembobotan ini digunakan sebelum melakukan tahapan klasifikasi. Setelah pembobotan fitur tahap selanjutnya teks tersebut akan diklasifikasikan dengan menggunakan *Lexicon Based* dan *Support Vector Machine* dan akan menghasilkan data teks yang memiliki kelas sentimen positif dan negatif.



Gambar 4 Flowchart Analisis Sentimen

Flowchart Klasifikasi Sentimen Menggunakan Metode SVM

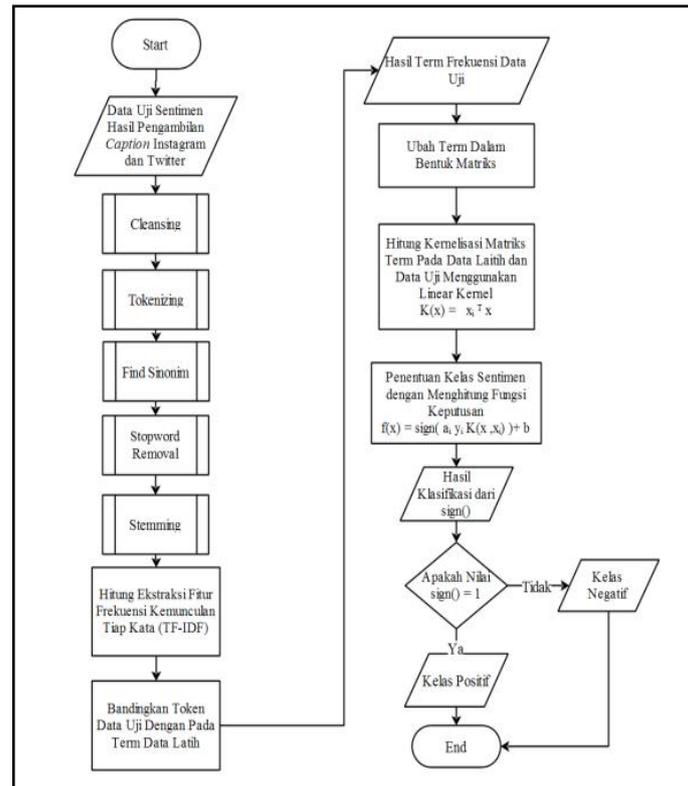
Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. *Support Vector Machine* masuk kelas *supervised learning*, dimana dalam implementasinya perlu adanya tahap pelatihan menggunakan *sequential training SVM* dan disusul tahap pengujian (Santosa, 2010). Proses klasifikasi menggunakan metode SVM dibagi menjadi 2 tahapan yaitu proses data latih dan proses data uji.

Pada *flowchart* data latih dimulai dengan menginputkan data latih yang sudah diberi label sentimen secara manual yang ada di database yang kemudian akan dilakukan tahap *preprocessing*. Selanjutnya akan dilakukan ekstraksi fitur untuk menghitung kemunculan setiap kata atau token, kemudian ubah kata kedalam format *support vector* dan beri label kelas pada setiap *term* kata. Setelah itu menghitung bobot (w) dan faktor bias (b), apabila telah didapatkan nilainya maka dilakukan kernelisasi fungsi menggunakan *linier kernel* yang akan digunakan untuk perhitungan klasifikasi pada data uji. Apabila nilai matriks telah didapat maka selanjutnya tahapan terakhir menghitung nilai *Lagrange Multiplier* α yang berguna untuk perhitungan pada data uji berikutnya.

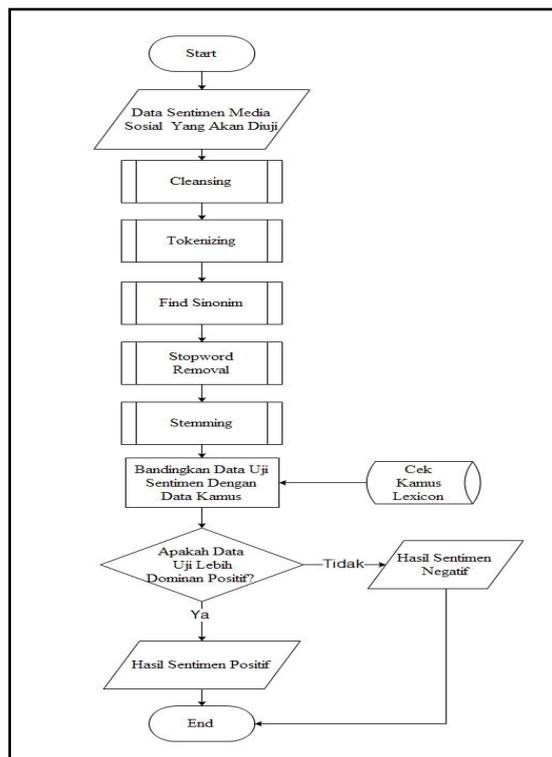
Selanjutnya pada *flowchart* proses tahapan data uji dimulai dengan inputan data uji yang diambil dari data API Instagram dan Twitter yang memiliki sentimen yang selanjutnya masuk ke tahap *preprocessing text* lalu bandingkan kata atau token dengan dengan token data latih yang disimpan dalam data pada database. Jika kata sesuai dengan kata yang ada dalam database, maka akan diberi bobot pada setiap *term*. Selanjutnya melakukan perhitungan klasifikasi menggunakan fungsi *linier kernel* dengan mengalikan matrik *term* kata baru data uji dengan matrik *term* yang merupakan dokumen pada data latih. Setelah hasil matrik didapatkan selanjutnya penentuan kelas sentimen dengan menggunakan fungsi keputusan $f(x)$ yaitu fungsi $sign()$ dengan memasukkan nilai α dan b yang didapatkan dari hasil data latih. Fungsi $sign()$ adalah fungsi normalisasi, apabila nilai x pada fungsi $sign(x) > 0$ maka fungsi tersebut memberikan nilai 1 atau positif, namun jika nilai x dalam fungsi adalah < 0 maka fungsi akan memberi nilai (-1) atau negatif, maka proses data uji selesai. Flowchart klasifikasi menggunakan SVM dapat dilihat pada gambar 5.

Flowchart Klasifikasi Sentimen Menggunakan Metode Lexicon Based

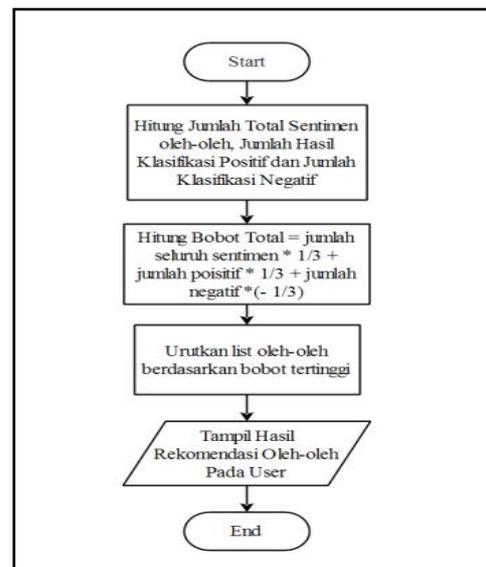
Pada *flowchart* klasifikasi *lexicon based* ini dimulai dari masukkan data uji berupa sentimen yang diambil dari Instagram dan Twitter. Kemudian lakukan tahap *preprocessing text* Setelah itu bandingkan kata hasil stemming dengan data kamus sentimen *lexicon* yang sudah terdapat pada database. Jika di dalam sebuah komentar lebih dominan positif maka sentimennya yang akan dihasilkan adalah positif, jika tidak maka hasil sentimen adalah negatif. Flowchart analisis sentimen menggunakan *lexicon based* dapat dilihat pada gambar 6.



Gambar 5 Flowchart Klasifikasi SVM (Habibi, 2017)



Gambar 6 Flowchart Klasifikasi Lexicon Based (Saragih, 2018)



Gambar 7 Flowchart Hasil Rekomendasi Pada Halaman User

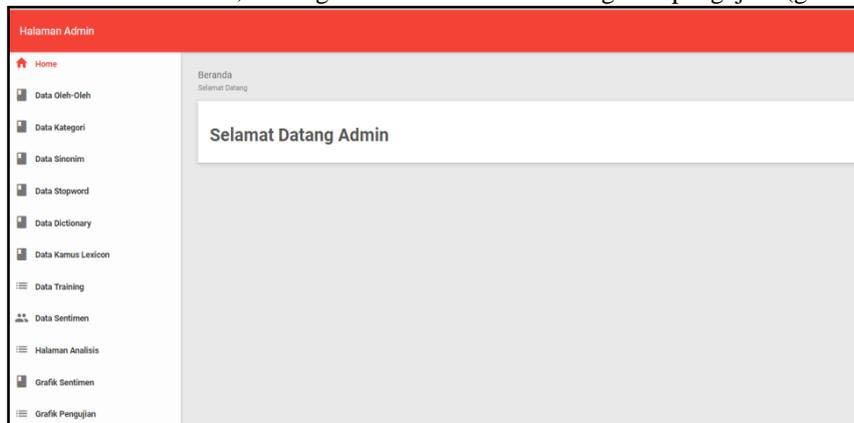
Flowchart Hasil Top Rekomendasi Oleh-Oleh (User)

Pada halaman user ditampilkan Top Rekomendasi Oleh-Oleh yang didapatkan dari penjumlahan seluruh data sentimen, data positif dan negatif SVM dan *lexicon* yang selanjutnya dihitung bobot total yang selanjutnya bobot tersebut diurutkan sebagai list rekomendasi yang dijelaskan pada gambar 7.

4. HASIL DAN PEMBAHASAN

Halaman Admin

Beberapa menu pada halaman admin yaitu menu beranda, menu data oleh-oleh, menu data kategori, menu data sinonim, menu data *stopword*, menu data *dictionary*, menu data kamus *lexicon*, menu data *training*, menu data sentimen, menu halaman analisis, menu grafik sentimen dan menu grafik pengujian (gambar 8).

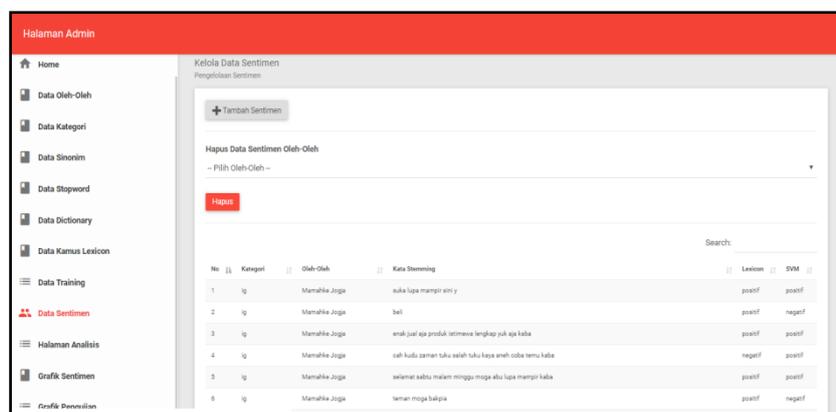


Gambar 8 Halaman Beranda Admin

Data sentimen diambil menggunakan Twitter API dan Instagram API. Proses pengambilan API diawali dengan membuka koneksi streaming API. Koneksi ini merupakan koneksi HTTP yang bersifat persisten. Setelah koneksi terbuka pengguna dapat melakukan *filtering tweet* atau *caption* Instagram dengan menginputkan kata kunci *hashtag* nama oleh-oleh. Sehingga didapatkan sentimen sesuai dengan kata kunci yang dimasukkan. Data sentimen yang diperoleh tersebut akan masuk pada tahapan *preprocessing*, pembobotan dan klasifikasi menggunakan *lexicon* dan svm. Data hasil klasifikasi akan ditampilkan pada halaman data sentimen. Pada halaman data sentimen akan ditampilkan nama oleh-oleh, sentimen hasil stemming dan kelas sentimen hasil klasifikasi serta dapat menambah sentimen dan menghapus sentimen, berikut implementasinya pada gambar 9.

Halaman Pengunjung

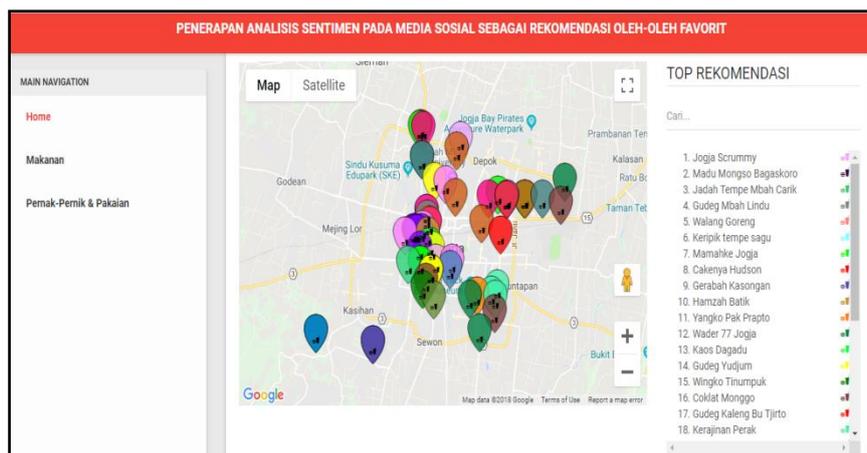
Implementasi pengunjung merupakan tampilan aplikasi yang berhubungan dengan pengunjung, dimana pengunjung dapat melihat top rekomendasi oleh-oleh berdasarkan sentimen media sosial Twitter dan Instagram beserta peta lokasi oleh-oleh tersebut, pengunjung juga dapat melakukan pencarian oleh-oleh dan dapat melihat berapa banyak review positif dan negatif dari oleh-oleh tersebut, berikut implementasinya pada gambar 10.



Gambar 9 Halaman Data Sentimen

Pengujian Sistem

Pada tahap ini akan dijelaskan hasil pengujian akurasi, presisi dan *recall* yang dihasilkan oleh metode SVM dan *lexicon based* dengan menggunakan data *training* dalam menentukan klasifikasi sentimen. Pada SVM untuk menghitung akurasi menggunakan k-fold cross validation. Data sentimen yang digunakan dalam pelatihan ini sebanyak 1450 data sentimen. Proses pengujian menggunakan pendekatan k=10, pada 10-fold cross validation data akan dibagi menjadi 10 subset dengan ukuran yang sama dan data yang berbeda. Pada setiap iterasi, satu bagian digunakan untuk data pengujian sisanya digunakan untuk data pelatihan. Misalnya pada iterasi pertama subset pertama digunakan sebagai data pengujian, subset kedua sampai sepuluh digunakan sebagai data pelatihan. Pada iterasi kedua dan subset kedua sebagai data pengujian, subset pertama, ketiga sampai dengan sepuluh sebagai data pelatihan dan begitu seterusnya hingga iterasi kesepuluh. Sedangkan pada *Lexicon based* untuk menguji akurasinya menggunakan metode *confusion matrix*. Evaluasi dilakukan untuk mengetahui performa metode sehingga dapat digunakan untuk melakukan klasifikasi.

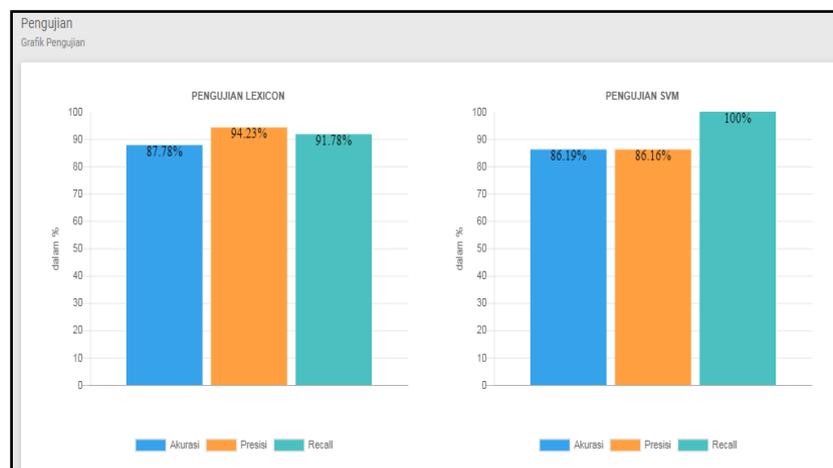


Gambar 10 Halaman Pengunjung

Berikut hasil pengujian metode SVM menggunakan *k-fold cross validation* dapat pada tabel 1 dan hasil pengujian metode *Lexicon Based* dan *Support Vector Machine* dapat dilihat pada gambar 11.

Tabel 1. Hasil Pengujian 10-Fold Cross Validation

Iterasi	Data Pelatihan		
	Akurasi	Presisi	Recall
1	91.37%	91.37%	100%
2	86.33%	86.33%	100%
3	89.93%	89.86%	100%
4	87.05%	87.05%	100%
5	81.29%	81.16%	100%
6	84.17%	84.17%	100%
7	88.49%	88.49%	100%
8	80.58%	80.58%	100%
9	86.33%	86.23%	100%
10	91.37%	91.37%	100%



Gambar 11 Grafik pengujian pada sistem

Selanjutnya akan dilakukan pengujian pada data testing akan dilakukan secara manual dengan mengambil data sampel 50 data sentimen dari API media sosial yang sebelumnya sudah ada pada sistem. Sebelum dilakukan pengujian, seluruh komentar data uji dilabeli secara manual berdasarkan jenis sentimennya. Dari hasil analisis 50 data uji tersebut dapat dibentuk sebuah tabel confusion matrix, dapat dilihat pada tabel 2.

Tabel 2. Tabel Confusion Matrix

Metode Analisis	Confusion Matrix			
	TP	FN	FP	TN
SVM	41	4	3	2
LB	40	5	1	4

Tabel *confusion matrix* digunakan sebagai acuan dalam melakukan pengujian akurasi, presisi dan recall. Melalui confusion matrix dapat diketahui *True Positif*, *False Negative*, *False Positive* dan *True Negative* yang dihasilkan oleh sistem dalam melakukan analisis sentimen. Berikut tabel hasil perhitungan menggunakan *confusion matrix* pada 50 data uji :

Tabel 3. Hasil pengujian menggunakan Support Vector Machine

Sentimen Uji	Akurasi	Presisi	Recall
50 data uji	86%	93.20%	91.11%

Tabel 4. Hasil pengujian menggunakan Lexicon Based

Sentimen Uji	Akurasi	Presisi	Recall
50 data uji	88%	97.56%	88.89%

Tabel 3 dan Tabel 4 merupakan hasil pengujian akurasi, presisi dan *recall* metode SVM dan *lexicon based* dalam melakukan pengklasifikasian sentimen. Berdasarkan hasil pengujian manual diatas memiliki nilai akurasi yang tidak jauh beda dengan hasil pengujian pada sistem.

Setelah dilakukan pengujian, dapat diketahui bahwa metode *lexicon based* memberikan akurasi dan presisi yang lebih baik daripada metode *support vector machine* dengan akurasi sebesar 87.78% dan presisi sebesar 94.23%. Sedangkan untuk hasil *recall*, metode *support vector machine* lebih baik daripada metode *lexicon based* dengan *recall* sebesar 100%.

5. KESIMPULAN

Berdasarkan hasil dari analisis dan perancangan yang telah dibuat, didapatkan beberapa kesimpulan, yaitu:

1. Hasil analisis pada sistem yang dibangun memperlihatkan bahwa analisis sentimen yang dilakukan dengan menggunakan data sentimen yang diperoleh dari media sosial Twitter dan Instagram dapat menghasilkan suatu rekomendasi oleh-oleh favorit khususnya di Yogyakarta dengan total oleh-oleh sebanyak 30 yang telah diurutkan berdasarkan oleh-oleh yang memiliki bobot tertinggi.
2. Hasil pengujian pada sistem yang dibangun memperlihatkan bahwa metode *lexicon based* memberikan akurasi dan presisi yang lebih baik daripada metode *support vector machine* dengan akurasi sebesar 87.78% dan presisi sebesar 94.23%. Sedangkan untuk hasil *recall*, metode *support vector machine* lebih baik daripada metode *lexicon based* dengan *recall* sebesar 100%.
3. Metode *lexicon based* sangat bergantung pada jumlah data kamus sentimen yang digunakan, semakin banyak jumlah data kamus yang digunakan maka semakin besar nilai keakurasiannya, dan sebaliknya semakin sedikit jumlah data kamus yang digunakan maka semakin kecil nilai keakurasiannya yang dihasilkan.

DAFTAR PUSTAKA

- Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. 2011. Everyone's an influencer. In *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11* (p. 65). New York, New York, USA: ACM Press.
- Bennett, W Lance, and Courtney N Johnson. 2014. "A Model of Crowd-Enabled Organization : Theory and Methods for Understanding the Role of Twitter in the Occupy Protests." 8: 646–72.
- Boyd, D. M., & Ellison, N. B. 2007 : *Social Network Sites: Definition, History, and Scholarship*. Dipetik September 20, 2008, dari Journal of ComputerMediated Communication: <http://jcmc.indiana.edu/vol13/issues1/boyd.ellison.html>
- Cha et al. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. 4th International AAAI Conference on Weblogs and Social Media.
- Coletta, Luiz Fernando Sommaggio, Nadia Felix Felipe da Silva, Eduardo Raul Hruschka, and Estevam Rafael Hruschka. 2014. "Combining Classification and Clustering for Tweet Sentiment Analysis." In *2014 Brazilian Conference on Intelligent Systems, IEEE*, 210–15.
- Kurniawati, N.R., Edi Winarko, D., 2016. Penentuan Destinasi Wisata Favorit Berbasis Aturan Dan Analisis Sentimen Pada Tweet Berbahasa Indonesia. Universitas Gadjah Mada.
- Mejova, Y., 2009, *Sentiment Analysis: An Overview*, Comprehensive Exam Paper, Computer Science Departement, University of Iowa.
- Pressman, R. S. 2012. *Rekayasa Perangkat Lunak: Pendekatan Praktisi*. Andi.
- Priambada, S. 2015. Manfaat Penggunaan Media Sosial Pada Usaha Kecil Menengah (UKM), (November), 2–3.
- Puspitarani, Yan. 2015. "Sentimen Analisis Terhadap Nilai Kepercayaan Sebuah Online Shop Di Instagram." *Jurnal Ilmiah Teknologi Informasi Terapan* II(1): 76–81.
- Robson, M.J., & Skarmeas, D. 2008. Determinants of Relationships Quality in Importer-Exporter Relationships. *British Journal of Management*, 19(2), 171-184.
- Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Saragih, J. 2018. Analisis Sentimen Pada Hot Messages Computer Based Information System (Cbis) FTI Upn "Veteran" Yogyakarta menggunakan Naïve Bayes Classifier Dan Lexicon Based. Universitas Pembangunan Nasional Veteran Yogyakarta.
- Tala, F. Z. 2003. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D, pp*, 39–46. <https://doi.org/10.22146/teknosains.26972>