

ANALISIS SENTIMEN TERHADAP PENERAPAN SISTEM GANJIL GENAP MENGGUNAKAN METODE K - NEAREST NEIGHBOR (KNN)

Fadillah Ali Rohmansyah¹, Bayu Bintoro², Imam Santoso³

¹STIKOM Cipta Karya Informatika, ²STIKOM Cipta Karya Informatika, ³UTM Jakarta
fadillah@stikomcki.ac.id, bayu@stikomcki.ac.id, imam.santoso@utmj.ac.id

ABSTRAK

Penerapan sistem Ganjil Genap pada rambu lalu lintas merupakan kebijakan pemerintah terutama untuk mengurangi kemacetan di ibu kota, dengan cara tersebut pemerintah memberikan solusi kepada pengguna jalan terutama angkutan umum karena di ibu kota banyak pengguna kendaraan pribadi terutama mobil yang bisa membuat kemacetan parah, karena semakin banyak para pengguna kendaraan pribadi pada saat hari kerja. Dalam penelitian ini mencoba untuk menganalisis hasil sentimen pada platform medi sosial untuk kebijakan ganjil genap tersebut, sehingga dapat mengelompokkan pengguna platform media sosial menjadi opini positif dan negatif, penelitian ini menggunakan metode data mining untuk klasifikasi dengan algoritma K- Nearest Neighbor. K- Nearest Neighbor merupakan K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasi berdasarkan mayoritas dari kategori pada KNN. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap, dalam penelitian ini menghasilkan accuracy 75,36%.
Kata kunci : Analisis Sentimen, Penerapan Sistem Ganjil Genap, Metode K- Nearest Neighbor

ABSTRACT

The application of the odd-even system to traffic signs is a government policy, especially to reduce congestion in the capital city. private vehicle users during weekdays. In this study trying to analyze the results of sentiment on social media platforms for the odd-even policy, so that users of social media platforms can be grouped into positive and negative opinions, this research uses data mining methods for classification with the K-Nearest Neighbor algorithm. K-Nearest Neighbor is K-Nearest Neighbor (KNN) which is a method that uses a supervised algorithm where the results of the new query instance are classified based on the majority of the categories in the KNN. This research produces text classifications in positive and negative form for the application of odd-even traffic , in this study produced an accuracy of 75.36%.

Keywords: Sentiment Analysis, Application of the Odd-Even System, K-Nearest Neighbor Method

1. PENDAHULUAN

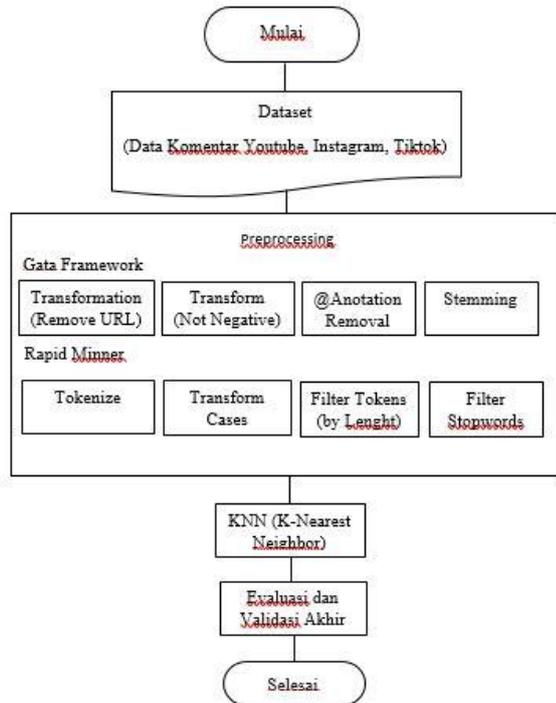
Aktivitas ibu kota Jakarta sangat padat dan juga sangat ramai lalu lintasnya pada saat orang-orang melakukan aktivitas di pagi hari. Dengan demikian menyebabkan kemacetan lalu lintas. Dikarenakan juga kurang kesadaran para pengguna kendaraan pribadi terhadap transportasi umum, sehingga kemacetan di Jakarta sangat tidak bisa dihindari. Kemacetan yang terjadi di Jakarta hampir disetiap ruas jalan di Jakarta dikarenakan sarana dan prasarana transportasi umum kurang memadai. Untuk masalah transportasi dalam kemacetan karena buruknya layanan angkutan umum inilah yang memicu peningkatan pengguna kendaraan pribadi di Jakarta. Selain itu juga kurangnya kesadaran pribadi untuk menggunakan kendaraan umum.

Kebijakan penerapan sistem ganjil/genap di Jakarta untuk mengurangi kepadatan lalu lintas. Dengan

angka penjualan mobil yang ada, otomatis pengendara pribadi semakin meningkat untuk mengantisipasinya maka diberlakukan untuk sistem ganjil /genap pada setiap plat nomor kendaraan khususnya mobil. Dan tingkat kedisiplinan serta kesadaran yang semakin hari semakin tidak beraturan sehingga tidak terkendalinya. Penataan dan pengaturan penggunaan jalan dan fasilitas pendukungnya yang belum maksimal untuk para pengguna jalan. Dengan adanya penerapan sistem ganjil/genap banyak sekali tanggapan masyarakat ada yang berpendapat positif dan ada yang berpendapat negatif, terutama pada platform media sosial yang orang bebas berpendapat untuk berpendapat apa saja. Hasil yang didapat dari penelitian tersebut K-Nearest Neighbor 75.36 % untuk Accuracy, serta -3,81 untuk micro average.

2. METODOLOGI

Metode yang digunakan dalam penelitian ini adalah metode *Knowledge Discovery in Databases* yang dianggap efektif dan tepat dalam proses sehingga format yang diambil benar. Adapun langkah-langkah metode penelitian yang dijelaskan pada gambar dibawah ini :



Gambar 1. Metode Penelitian

Penelitian ini bertujuan untuk mengklasifikasi sentimen pada platform media sosial terhadap penerapan sistem ganjil /genap sehingga dapat bisa memberikan hasil apakah penerapan ganjil/genap tanggapan masyarakat baik atau buruk di media sosial twitter.

A. Pengumpulan dan Pengolahan Data

Proses pengambilan data dilakukan secara manual satu persatu. data yang digunakan dalam penelitian ini sebanyak 350 data yang terdiri dari 128 data *review* positif dan 221 data *review* negatif.

Setelah melakukan pengumpulan data, dilakukan proses pengolahan data atau *preprocessing*, tahapan ini mencakup kegiatan membangun data dan dilanjutkan kegiatan pembersihan data agar dapat dilakukan pengelolaan ketahap selanjutnya. Berikut tahapan *preprocessing* data, antara lain :

1. Remove URL

Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil

pengumpulan data. Biasanya digunakan untuk merujuk pada *source* atau sumber dari suatu berita/informasi.

2. Transform (Not Negative)

Proses mengubah kata – kata yang bermakna negatif yang akan disatukan dengan tanda garis bawah (_).

3. Anotation Removal

Menghapus tanda Mention (@) beserta teks yang ada dibelakangnya. Pada media sosial biasanya berfungsi untuk memanggil user/pengguna lainnya.

4. Stemming

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

5. Tokenize

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

6. Transforms Cases

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

7. Filter Tokens (By Length)

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

8. Filter Stopword

Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Preprocessing pertama dilakukan menggunakan *gataframework*, yaitu pengolahan data teks berbasis *web*, dengan menggunakan bahasa pemrograman *php*. *Gataframework* dapat diakses di alamat link www.gataframework.com/textmining. Banyak yang tersedia di *Gataframework*, antara lain *anotation removal*, *remove url* dan lainnya. Peneliti menggunakan *Gataframework* karena memiliki keunggulan dalam melakukan proses *Stemming* bahasa Indonesia. Pengolahan data selanjutnya dilakukan menggunakan tools *Rapidminer*.

3. LANDASAN TEORI

A. Analisis Sentimen

Sentiment Analysis, atau yang biasa disebut juga *Opinion Mining*, adalah bidang studi yang bertujuan untuk menganalisis opini, sentimen, penilaian, evaluasi, sikap dan emosi publik terhadap suatu entitas dari produk, pelayanan, suatu permasalahan, organisasi,

peristiwa tertentu, topik yang hangat dibicarakan dan atributnya.

Ada beberapa macam metode yang bisa digunakan pada analisis sentimen, yaitu antara lain, NB (*Naive Bayes*), *Decision Tree*, KNN (*K-Nearest Neighbor*), .

B. Algoritma K-Nearest Neighbor

KNN Algorithm, adalah algoritma machine learning yang bersifat non-parametric dan lazy learning. Metode yang bersifat non-parametric memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, terlepas data tersebut berukuran kecil ataupun besar.

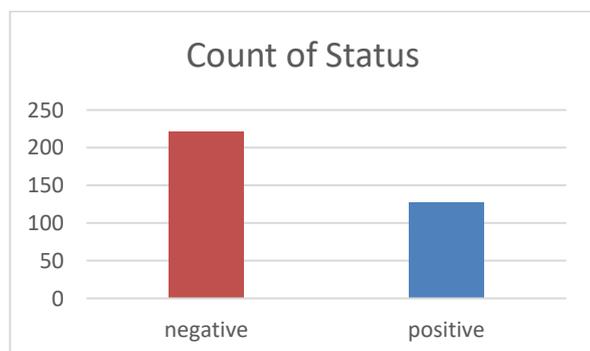
Algoritma KNN juga bersifat lazy learning, yang artinya tidak menggunakan titik data training untuk membuat model. Singkatnya pada algoritma KNN tidak ada fase training, walaupun ada juga sangat minim. Semua data training digunakan pada tahap testing. Hal ini membuat proses training lebih cepat dan tahap testing lebih lambat dan cenderung ‘mahal’ atau membutuhkan banyak cost dari sisi waktu dan memori. Algoritma KNN mengasumsikan bahwa sesuatu yang mirip akan ada dalam jarak yang berdekatan atau bertetangga. Artinya data-data yang cenderung serupa akan dekat satu sama lain.

4. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar video youtube mengenai peresmian pabrik mobil esemka, peneliti mengumpulkan 350 data opini pengguna youtube, twitter, instagram. Kemudian data tersebut kumpulkan dalam format file xls, setelah itu data dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif.

Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga kemudian dihasilkan 128 opini Positif dan 221 opini negatif.



Gambar 2. Statistik hasil pengumpulan data.

B. Pengolahan Data

Setelah pengumpulan data dan membagi data menjadi komentar positif dan komentar negatif, selanjutnya dilakukan proses pengolahan data atau *preprocessing*. Berikut ini adalah tahapan *preprocessing* :

1. Remove URL

Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil pengumpulan data. Berikut contoh perbedaan sebelum dan sesudah proses.

Tabel 2. Hasil *Remove URL*.

Sebelum	Sesudah
Banyak yang mengikuti link ini https://youtu.be/OKr97KPrMJ0 jadi banyak yang sama	Banyak yang mengikuti link ini jadi banyak yang sama

2. Transform (Not Negative)

Proses mengubah kata – kata yang bermakna negatif yang akan disatukan dengan tanda garis bawah (_).

Tabel 3. Hasil *Transform Negative*.

Sebelum	Sesudah
Mendukung bukan dihujat	Mendukung bukan_dihujat

3. Anotation Removal

Menghapus tanda *Mention* (@) beserta teks yang ada dibelakangnya.

Tabel 4. Hasil *Anotation Removal*.

Sebelum	Sesudah
Riviewnya kirim ke @fadillah	Riviewnya kirim ke

4. Stemming

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

Tabel 5. Hasil *Stemming*.

Sebelum	Sesudah
Mencari cara atasi macet adalah pola hidup masyarakat untuk peralihan transportasi umum	cara atas macet adalah pola hidup masyarakat untuk alih transportasi umum

5. Tokenize

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

6. Transforms Cases

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

Tabel 6. Hasil *Transforms Cases*.

Sebelum	Sesudah
Cara mengatasi kemacetan yang ada di Ibu Kota	Cara mengatasi kemacetan yang ada di ibu kota

7. Filter Tokens (By Length)

Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

8. Filter Stopword

Membuang kata – kata yang diabaikan pada sentimen analis, biasanya yang berupa kata sambung dan kata keterangan.

Tabel 7. Hasil Filter Stopword.

Sebelum	Sesudah
percuma adanya aturan	percuma ada aturan

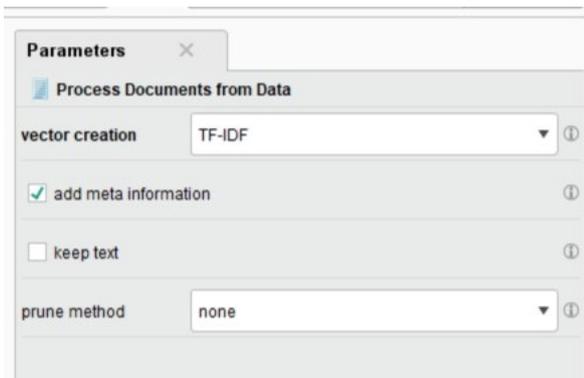
Berikut gambar proses pengolahan data pada tools rapidminer.



Gambar 3. Proses Pengolahan Data Rapidminer.

Data hasil *preprocessing* yang berupa kata akan diubah ke dalam bentuk angka dengan dilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapan ini terdapat dua bagian proses yaitu *TF* (*Term Frequency*) dan *IDF* (*Inverse Document Frequency*),

TF adalah jumlah kemunculan tiap kata pada sebuah dokumen semakin banyak kata muncul pada tiap dokumen maka semakin besar nilai *TF*. *Df* adalah jumlah nilai dokumen pada tiap kata yang berbanding terbalik yaitu apabila suatu kata jarang muncul pada sebuah dokumen maka nilai *IDF* lebih besar daripada kata yang sering muncul.

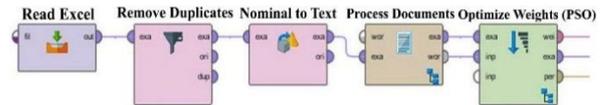


Gambar 4. Proses Tf Idf.

C. Fitur seleksi Particle Swarm Optimization.

Peneliti mengusulkan menggunakan *Particle Swarm Optimization* untuk meningkatkan akurasi dari pengklasifikasi Algoritma K- Nearest Neighbor.

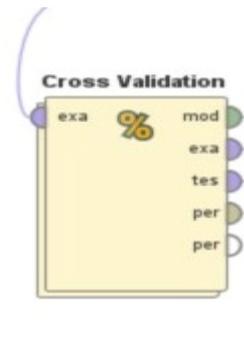
Penelitian ini nantinya menghasilkan akurasi dan nilai *AUC* dengan menggunakan aplikasi *RapidMiner* untuk hasil evaluasi.



Gambar 5. Model Pengujian PSO dengan RapidMiner.

D. Proses Validasi Algoritma.

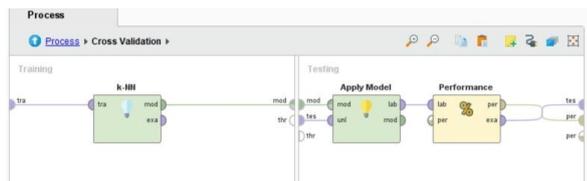
Selanjutnya adalah penggunaan algoritma K-Nearest Neighbor. untuk melakukan klasifikasi data yang dihubungkan dengan pengujian *10-fold cross validation* dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak ke dalam *10 fold* untuk mendapatkan 10 data yang sama, kemudian data tersebut digunakan *9 fold* untuk data *training* dan *1 fold* untuk data *testing*.



Gambar 6. Proses Cross Validation dengan Rapidminer.

Tahapan selanjutnya adalah proses Evaluasi, ialah mengevaluasi kinerja terhadap permodelan dan perhitungan yang telah digunakan dengan menerapkan algoritma K- Nearest Neighbor yang ditambahkan *Particle Swarm Optimization* agar dihasilkan peningkatan akurasi pada penghitungan.

Untuk mengestimasi performa dari model algoritma yang telah dipilih digunakan *Cross Validation* sehingga dapat menghasilkan nilai akurasi (Proses *Cross Validation* dapat dilihat pada Gambar 6). Dibawah ini adalah model pengujian K- Nearest Neighbor dengan menggunakan tools *Rapidminer*.



Gambar 7. Model Pengujian K- Nearest Neighbor dengan RapidMiner

E. Evaluasi

Penelitian ini menggunakan algoritma *KNN* (*K-Nearest Neighbor*) untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan *Accuracy* dan *AUC* (*Area Under Curve*).

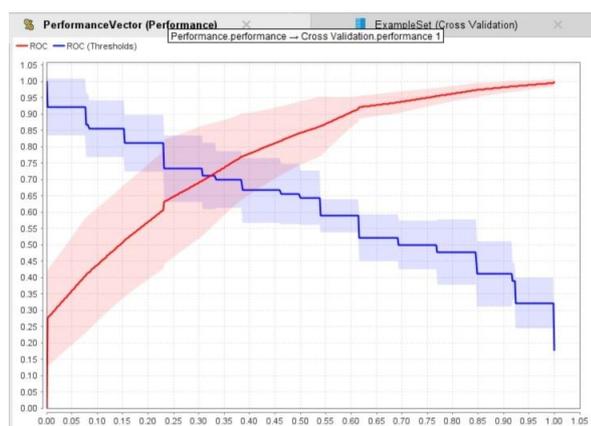
Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan *dataset* sebanyak 350 komentar opini pengguna *Youtube*, *Instagram*, *Tiktok* mengenai video pelecehan seksual di lingkup kampus, maka hasil Akurasi Algoritma *KNN* (*K-Nearest Neighbor*) tanpa menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 75,36%.

Berikut ini adalah Tabel *Confusion Matrix* Algoritma *K-Nearest Neighbor* tanpa fitur seleksi.

Tabel 8. *Confusion Matrix* Algoritma NB.

	True Positive	True Negative	
Pred Positive	68	26	72,34%
Pred Negative	60	195	76,47%
	53,12%	88,24%	

Pada tabel 8 *Confusion Matrix* dapat dilihat, sebanyak 128 data diprediksi *class* positif ternyata sesuai, yaitu masuk kedalam *class* positif, sebanyak 221 data yang diprediksi *class* negatif ternyata termasuk kedalam prediksi *class* positif.



Gambar 9. Grafik *Area Under Curve* (AUC) *KNN*.

5. KESIMPULAN

Dari hasil penelitian untuk mencari sentiment dalam hal penerapan sistem ganjil genap untuk tanggapan pada pengguna twitter dan dilakukan analisa sentiment dengan label positif dan negatif dengan metode klasifikasi naïve bayes pengujian dilakukan dengan data training dan data testing yang di ambil dari twitter menggunakan software rapid miner sehingga memudahkan memberikan hasil dari penelitian tersebut dengan hasil accuracy 75,36%.

Model terbentuk dapat diterapkan untuk melihat sentimen dari para pengguna kendaraan dalam bentuk positif dan negatif. Hal ini dapat membantu dalam penerapan kebijakan pemerintah untuk memberikan solusi terbaik dalam pengaturan lalu lintas di jalan.

DAFTAR PUSAKA

Kompas.com” Penerapan ganjil genap” <https://www.google.com/amp/s/amp.kompas.com/megapolitan/read/2022/06/08/16162421/dish-ub-dki-sebut-penerapan-ganjil-genap-di-25-ruas-jalan-jakarta-turunkan>

Trivusi.2022. “Yuk Kenali Apa itu Algoritma K-Nearest Neighbors (KNN)”, <https://www.trivusi.web.id/2022/06/algoritma-knn.html>

Gorunescu, F. (2011). *Data Mining Concepts, models and Techniques*. Springer Berlin Heidelberg.

Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.