

ANALISIS SENTIMEN PUBLIK TERHADAP PENGAMBILALIHAN JALAN RUSAK DI LAMPUNG MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN)

Afyra Ar'bah Lailany ¹, Anggit Nur Hannaa Regita ², Imam Santoso, M. Kom³

^{1,2} Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, ³ Universitas Teknologi Muhammadiyah Jakarta

^{1,2} Jl. Radin Inten II, Kec. Duren Sawit, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, 13440, ³ Jl Sungai Kendal, Kec. Cilincing, Jakarta Utara, Daerah Khusus Ibukota Jakarta, 14140

¹afyraarbah@gmail.com, ²anggitnurhannaregita@gmail.com, ³imam.santoso@utmj.ac.id

ABSTRAK

Pada saat ini banyak sekali kasus yang sedang marak diperbincangkan oleh masyarakat yaitu jalan di Lampung yang rusak parah akibat sudah lama tidak di perbaiki, dimana gubernur Lampung diklaim menggunakan uang pembangunan dengan tidak sesuai sehingga membuat kita sebagai masyarakat terlebih masyarakat Lampung tentunya akan merasa dirugikan terhadap kasus tersebut. Kasus tersebut sampai mendatangkan bapak presiden, pak Jokowi, yang sudah ditentukan bahwa perbaikan jalan di Lampung akan dipindah tugas kan menjadi tanggung jawab pusat.. Topik pada kasus ini banyak menimbulkan pro dan kontra seperti yang ramai diperbincangkan di media sosial. Perlu dilakukan penelitian untuk melihat seberapa besar masyarakat yang setuju terhadap pengambilalihan jalan rusak di Lampung oleh pusat tersebut, Dari pernyataan itulah maka berdampak pada terjadinya kontroversi mengenai peran Gubernur Lampung yang tidak mampu menjalankan tugas di daerahnya sendiri. Peneliti memilih analisis sentimen karena merupakan teknik yang tepat untuk pengolahan dataset. Dari 500 opini pengguna tiktok yang berhasil dikumpulkan, didapat 422 opini positif, 48 opini negative. Kemudian, data tersebut diklasifikasikan menggunakan algoritma KNN, selanjutnya dilakukan accuracy yang menghasilkan sebesar 84,37% untuk hasil yang menggunakan algoritma KNN. Pengambilalihan jalan rusak di Lampung ke pada pusat adalah cara yang tepat dilakukan melalui uraian dan analisis secara opini publik.

Kata kunci : *Analisis Sentimen, Pengambilalihan jalan rusak di Lampung oleh Pusat, KNN (K-Nearest Neighbor), Opini masyarakat, Klasifikasi, Pengolahan Dataset.*

ABSTRACT

At this time there are many cases that are being discussed by the public, namely roads in Lampung that are badly damaged due to not being repaired for a long time, where the governor of Lampung is claimed to have used development money inappropriately, making us as a society, especially the people of Lampung, of course, will feel disadvantaged by the case. The case brought the president, Mr. Jokowi, who has determined that the repair of roads in Lampung will be transferred to the responsibility of the central government. The topic of this case has generated many pros and cons as discussed on social media. It is necessary to conduct research to see how much the public agrees with the takeover of damaged roads in Lampung by the center, from that statement, it has an impact on the controversy regarding the role of the Governor of Lampung who is unable to carry out tasks in his own region. Researchers chose sentiment analysis because it is the right technique for processing datasets. Of the 500 opinions of TikTok users that were collected, 422 positive opinions were obtained, 48 negative opinions. Then, the data is classified using the KNN algorithm, Then the accuracy is carried out which results in 84.37% for results using the KNN algorithm. Taking over damaged roads in Lampung to the center is the right way to do it through description and analysis of public opinion.

Keyword : *Sentiment Analysis, Takeover of damaged roads in Lampung by the Center, KNN (K-Nearest Neighbor), Public opinion, Classification, Dataset Processing .*

1. PENDAHULUAN

Pada akhir pekan lalu, Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR) mulai mengeksekusi pengambilalihan perbaikan sejumlah ruas jalan rusak di Provinsi Lampung. Menteri PUPR, Basuki Hadimuljono, mengatakan perbaikan jalan daerah itu mengacu kepada Instruksi Presiden (Inpres) Nomor 3 tahun 2023 tentang Percepatan Peningkatan Konektivitas Jalan daerah yang telah dikeluarkan oleh Presiden pada bulan Maret 2023. Saat mengunjungi Provinsi Lampung pada pekan lalu, Presiden Joko Widodo (Jokowi) mengungkapkan, bahwa pemerintah pusat akan mengalokasikan dana Rp 800 miliar untuk perbaikan 15 ruas jalan rusak di Lampung. Maka dari itu sekarang lebih diperhatikan lagi untuk anggaran dannaya. Topik pada kasus ini banyak menimbulkan pro dan kontra seperti yang ramai diperbincangkan di media sosial. Perlu dilakukan penelitian untuk melihat seberapa besar dampak dari pengambilalihan jalan rusak di Lampung tersebut, diharapkan dari penelitian ini dapat bermanfaat dalam penelitian mengenai opini masyarakat lainnya.

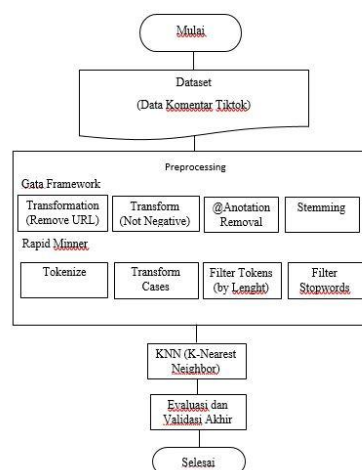
Dari pernyataan itulah maka berdampak pada terjadinya kontroversi mengenai pengambilalihan jalan rusak di Lampung, banyak masyarakat yang memberikan opini negatifnya terhadap kasus pada topik ini. Pada penelitian analisis sentimen kali ini, saya menggunakan KNN sebagai seleksi fitur yang diterapkan untuk meningkatkan peforma klasifikasi teks

Melalui penelitian ini diharapkan bisa mendapatkan hasil yang lebih baik, serta melalui penelitian ini dapat diketahui hasil dari algoritma KNN. Pada penelitian analisis sentimen ini, dilakukan untuk melihat sebuah opini seseorang atau masyarakat yang ditujukan kepada pengambilalihan jalan rusak di Lampung, sehingga dihasilkan opini yang dapat masuk ke kategori opini negatif dan opini positif. Dengan adanya penilaian maupun pandangan masyarakat yang disampaikan pada media sosial, maka peneliti berinisiatif membuat penelitian ini, peneliti

melihat belum ada penelitian tentang opinion mining sebelumnya yang mengangkat permasalahan tersebut, sehingga kedepannya diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif dan negatif ataupun lainnya.

2. METODOLOGI

Berikut adalah langkah – langkah metode penelitian yang dijelaskan pada gambar dibawah ini:



Gambar 1. Metode Penelitian

Penelitian ini menggunakan data yang didapat dari komentar tiktok berikut source link video <https://vt.tiktok.com/ZSLdf4BT2/> <https://vt.tiktok.com/ZSLdPehUE/>

Proses pengambilan data dilakukan secara manual satu persatu pada tanggal 22 Mei 2023, data yang digunakan dalam penelitian ini sebanyak 500 data yang terdiri dari data 422 review positif dan 78 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 Lenght).*
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 *annotation removal*, *remove url* dan lainnya. Peneliti menggunakan Gataframework karena memiliki keunggulan dalam melakukan proses Stemming bahasa Indonesia. Pengolahan data selanjutnya dilakukan menggunakan *tools Rapidminer*

2. 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 KNN (K-Nearest Neighbor)

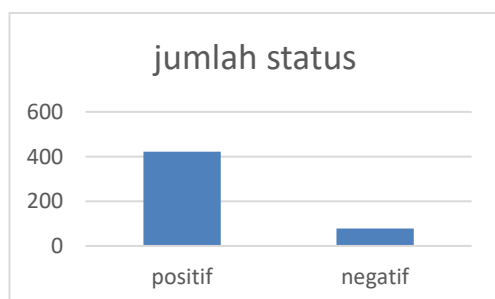
KNN Algorithm, adalah algoritma machine learning yang bersifat non-parametric dan lazy learning. Metode yang bersifat non-parametric memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi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

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar video tiktok mengenai pengambil alihan jalan rusak di Lampung, peneliti mengumpulkan 500 data opini pengguna tiktok. 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 422 opini Positif dan 78 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
rugi gubernur lampung bangun infrastruktur lampung uang negara uang, seperti link ini	rugi gubernur lampung bangun infrastruktur ur uang negara
https://dinastph.lampungprov.go.id/detail-post/pembangunan-infrastruktur-jalan-merupakan-salah-satu-program-prioritas-pemerintah-provinsi-lampung	

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
Gubernur_panic buru buru itu dana jalan	gubernur panic buru buru itu dana jalan

3. Anotation Removal.

Menghapus tanda Mention (@) beserta teks yang ada dibelakangnya

Tabel 4. Hasil Anotation Removal.

Sebelum	Sesudah
ketar ketir @gubernur	ketar ketir gubernur

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
bilanganya kerja pagi siang malam darimana	bilang kerja pagi siang malam mana

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

Lampung Utara rusak parah	lampung utara rusak parah
---------------------------	---------------------------

7. *Filter Tokens (By Lenght).*

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

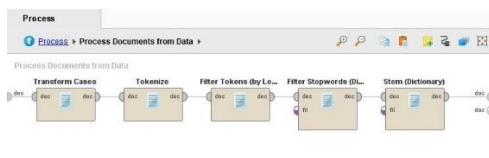
8. *Filter Stopword.*

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

Tabel 7. Hasil Filter Stopword.

Sebelum	Sesudah
tidak tahumalu jabat	tidak tahu malu
daerah jalan daerah	jabat daerah jalan
rusak pura tidak	daerah rusak tidak
tahu saat presiden	tahu saat presiden
ambil alih jabat	ambil alih jabat
daerah pura prihatin	daerah pura prihatin

Berikut gambar proses pengolahan data pada *tools rapidminer*.

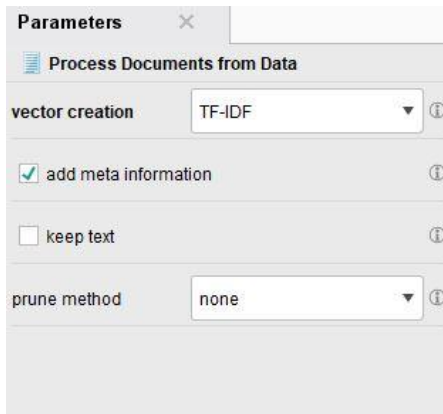


Gambar 3. Proses Pengolahan Data Rapidminer.

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

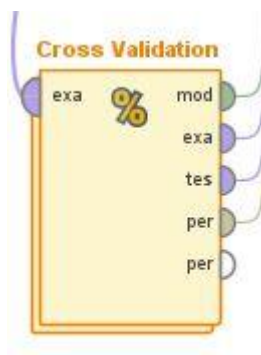
pada sebuah dokumen maka nilai IDF lebih besar daripada kata yang sering muncul.



Gambar 4. Proses Tf Idf .

C. Proses Validasi Algoritma

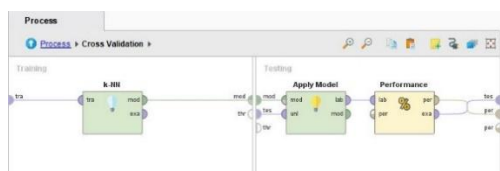
Selanjutnya adalah penggunaan algoritma KNN ((K-Nearest Neighbor) untuk melakukan klasifikasi data yang dihubungkan dengan pengujian 10-fold cross validation dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak ke dalam 10 fold untuk mendapatkan 10 data 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 KNN ((K-Nearest Neighbor). 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). Di bawah ini adalah model pengujian *KNN* ((*K-Nearest Neighbor*) dengan menggunakan *tools Rapidminer*.



Gambar 7. Model Pengujian *KNN* dengan *RapidMiner*

D. 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 500 komentar opini pengguna *Tiktok* mengenai video pengambilalihan jalan rusak di Lampung, maka hasil Akurasi Algoritma *KNN* ((*K-Nearest Neighbor*) tanpa menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 84,37%, sedangkan untuk nilai *AUC* sebesar 0.711.

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

Tabel 8. *Confusion Matrix* Algoritma *KNN*.

	True Negative	True Positive	
Pred Negative	12	15	44.44%
Pred Positive	59	387	86.77%
	16.90%	96.27%	

Pada tabel 8 *Confusion Matrix* dapat dilihat, sebanyak 12 data diprediksi *class* negatif ternyata sesuai, yaitu masuk kedalam *class* negatif, sebanyak 59 data yang diprediksi *class* positif ternyata termasuk kedalam prediksi *class* negatif.

Dan sebanyak 15 data yang diprediksi *class* negatif ternyata masuk dalam *class* positif, kemudian 387 data di prediksi *class* positif sesuai yaitu termasuk kedalam prediksi *class* positif.



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

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 500 komentar opini pengguna *Tiktok* mengenai video pengambil alih jalan rusak yang ada di Lampung, maka hasil Akurasi Algoritma *KNN* yaitu sebesar 84,37%, sedangkan untuk nilai *AUC* sebesar 0.711.

4. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa hasil akurasi dan *AUC* dengan algoritma *KNN*, hasil pengujian pengklasifikasian *AUC* nilai keakuratannya dapat dikategorikan sebagai *Excellent Classification*. signifikan.

Lalu juga di penelitian ini banyak sekali yang berkomentar atau beropini positive, ini membuktikan

bahwa tindakan pusat dalam mengambil alih jalan rusak yang ada di lampung adalah keputusan yang baik dikarenakan pemerintahan wilayah sudah tidak mampu untuk menjalankan tugas tersebut.

Untuk mendukung penelitian ini, maka peneliti akan mengembangkan aplikasi opini dari komentar tiktok mengenai mengambil alih jalan rusak yang ada di Lampung untuk mengklasifikasikan opini negatif dan positif menggunakan bahasa pemrograman *php* yang dilakukan oleh *Rapid Miner*.

DAFTAR PUSTAKA

Briliansyah, F. (2020). Sistem klasifikasi kategori berita menggunakan metode k-nearest neighbor (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan analisis sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 3(1), 1-13.

Syarifuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1), 23-28.

Aprilian, R., Habibi, R., & Setyawan, M. Y. H. (2020). Algoritma KNN dalam memprediksi cuaca untuk menentukan tanaman yang cocok sesuai musim. *Kreatif*.

Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining Dan Penerapan*. Deepublish.

Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Deepublish.