

PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE

¹Nova Tri Romadloni, ²Imam Santoso, ³Sularso Budilaksono

^{1,2,3}Magister Ilmu Komputer

STMIK Nusa Mandiri

Jakarta, Indonesia

¹novatrir2501@nusamandiri.ac.id, ²imanofani@yahoo.com, ³sularso@upi-yai.ac.id

ABSTRAK

Kondisi lalu lintas di kota Jakarta yang begitu padat seperti keadaan macet di jalan raya semakin meningkat terlebih dengan adanya kendaraan pribadi yang semakin banyak. Permasalahan hal ini membuat para warga yang hendak bekerja dan berpergian kemanapun memerlukan sarana transportasi yang lebih nyaman dan cepat untuk sampai ke tempat tujuan. Oleh karena itu dalam penelitian ini melakukan analisis sentimen terhadap pengguna KRL Commuter Line Jabodetabek pada data sosial media twitter diambil data secara random sebanyak 127 data sehingga diproses melalui beberapa tahap salah satunya menghindari data yang rangkap. Pada penelitian ini menggunakan metode Naive Bayes Classifier, KNN dan Decision Tree dengan beberapa tahapan yaitu *Convert Emoticon, Cleansing, Case Folding, Tokenizing, Stemming*. Adapun hasil uji coba menunjukkan bahwa memiliki perbedaan pada metode Naive Bayes akurasi sebesar 80%, precision 66,67%, sensitivity 100%, specificity 66,67%. Pada metode KNN akurasi sebesar 80%, precision 100%, sensitivity 50%, specificity 100% dan pada metode Decision Tree akurasi sebesar 100%, precision 100%, sensitivity 100%, specificity 100%.

Keywords—Sentimen, KRL Commuter Line, Twitter

ABSTRACT

Traffic conditions in the city of Jakarta, which are so crowded like traffic jams on the highway, are increasing, especially with more and more private vehicles. This problem makes residents who want to work and travel wherever they need transportation facilities that are more comfortable and faster to get to their destination. Therefore, in this study, sentiment analysis of Jabodetabek Commuter Line KRL users on Twitter social media data was taken randomly as many as 127 data so that it was processed through several stages, one of which was to avoid duplicate data. In this study using the Naive Bayes Classifier, KNN and Decision Tree methods with several stages, namely Convert Emoticon, Cleansing, Case Folding, Tokenizing, Stemming. The test results show that the difference between the Naive Bayes method is 80% accuracy, precision 66.67%, sensitivity 100%, specificity 66.67%. In the KNN method, accuracy is 80%, precision 100%, sensitivity 50%, specificity 100% and on the Decision Tree method accuracy is 100%, precision 100%, sensitivity 100%, specificity 100%.

Keywords—Sentimen, KRL Commuter Line, Twitter

1. PENDAHULUAN

Transportasi umum KRL Commuter Line merupakan salah satu pilihan bagi penduduk ibu kota Jakarta yang sangat padat. Terlebih dengan perkembangan teknologi ini dituntut untuk mengatasi masalah kepadatan penduduk yang berakibat kemacetan pada lalu lintas perkotaan. Semakin banyak penduduk berbanding lurus dengan banyaknya kendaraan yang akan digunakan untuk kehidupan sehari-hari. Para

pengguna jalanan umum baik kendaraan pribadi maupun transportasi umum jalan raya perkotaan sebagian memilih untuk naik KRL dan tentunya mengurangi volume kendaraan yang ada. Dengan KRL meminimalisir adanya kemacetan jalan.

Penggunaan transportasi tersebut memiliki berbagai macam kelebihan seperti tarif yang murah dibandingkan dengan transportasi lain dengan perbandingan perjalanan yang jaraknya jauh, adapun kekurangannya ketika jam berangkat dan

pulang kerja serta akhir pekan KRL sangat penuh sehingga harus berebut dan berbagi tempat duduk. Selain itu masih banyak lagi pendapat pro dan kontra dari kalangan masyarakat yang menggunakan transportasi KRL Commuter Line Jabodetabek.

Oleh karena itu perlu diketahui pendapat masyarakat pengguna KRL Commuter Line Jabodetabek tentang apa yang dirasakan saat menggunakan transportasi tersebut. Masyarakat dapat memberikan pendapatnya melalui berbagai macam media. Salah satunya media yang efektif sebagai wadah pendapat penggunaan transportasi KRL Commuter Line Jabodetabek ini adalah dengan media sosial twitter. Twitter termasuk cepat dalam mengabarkan pengalaman yang dirasakan oleh masyarakat sebagai bahan evaluasi untuk pihak transportasi KRL Commuter Line. Selain itu Twitter sendiri merupakan salah satu media sosial yang akrab digunakan oleh masyarakat Indonesia, yang tentunya akan memudahkan untuk pengumpulan pendapat dibandingkan dengan melakukan survey ataupun penyebaran kuisioner. Dengan banyaknya pengguna disertai kemudahan akses Twitter dalam penyampaian opini, maka akan banyak data opini terkumpul yang kemudian bisa menjadi peluang untuk dimanfaatkan sebagai bahan penilaian dan evaluasi atas penggunaan KRL Commuter Line di Jabodetabek.

Dalam penelitian sebelumnya penggunaan twitter, platform microblogging paling populer, untuk tugas analisis sentimen. Peneliti menunjukkan cara mengumpulkan secara otomatis korpus untuk analisis sentimen dan tujuan penambahan pendapat dan membangun suatu sentimen klasifikasi, yang dapat menghasilkan kata positif, negatif dan netral untuk dokumen. Evaluasi eksperimental menunjukkan bahwa teknik yang diusulkan adalah efisien dan berkinerja lebih baik. [10]

Penelitian sebelumnya yang lain mengenai Analisis Sentimen, data yang digunakan berbentuk tweet yang diambil langsung memanfaatkan API twitter. Dalam penelitian ini yang dimaksud dengan fitur adalah kata kunci yang akan menjadi parameter satuan data latih, yaitu dokumen (tweet) untuk diklasifikasikan ke dalam kelas yang telah

ditentukan (positif atau kelas negatif). Dalam kata lain fitur adalah kata yang memiliki nilai sentimen. Setelah dibentuk fitur dengan kemunculannya dari data latih. Selanjutnya menghitung probabilitas dari setiap kelas. Setelah didapat probabilitas dari setiap kelas, selanjutnya menghitung probabilitas setiap fitur pada setiap kelas.[14]

Hasil dari analisis sentimen tersebut dapat membantu dalam penilaian dan evaluasi terhadap penggunaan transportasi umum darat dalam kota. Dengan dilakukannya peningkatan fasilitas dan pelayanan berdasarkan hasil analisis sentimen, maka diharapkan masyarakat akan beralih menggunakan transportasi umum darat dalam kota, yang tentunya akan mengurangi kemacetan. Analisis sentimen dengan metode Support Vector Machine (SVM) dilakukan dengan pengujian terhadap komposisi data yang bervariasi. Dari hasil pengujian untuk kasus pada penelitian ini didapatkan bahwa SVM dapat diimplementasikan dengan nilai akurasi mencapai 78,12%. Variabel yang berpengaruh terhadap akurasi adalah jumlah data, perbandingan jumlah data latih dan uji, serta perbandingan jumlah data positif dan negatif yang digunakan.[15]

2. LANDASAN TEORI

Dalam penelitian ini terdapat berbagai macam istilah yang digunakan sebagai pendukung proses penelitian hingga menghasilkan suatu penilaian.

A. Analisis Sentimen

Sentiment Analysis adalah merupakan perpaduan dari data mining dan text mining, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. Sentiment analysis merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentiment yang terkandung dalam sebuah opini. Pada Sentiment analysis terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan sentimen analysis perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan

atau produk, melalui feedback masyarakat atau pun para ahli [11].

Analisis sentimen adalah studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan dalam teks. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen, kalimat, atau pendapat memiliki aspek positif atau negatif [5].

Analisis sentimen biasa digunakan untuk menilai kesukaan atau ketidaksukaan publik terhadap suatu barang atau jasa. Sentimen tersebut merupakan informasi tekstual yang bersifat

subjektif dan memiliki polaritas positif dan negatif. Nilai polaritas ini dapat digunakan sebagai parameter dalam menentukan suatu keputusan.[9]

B. Twitter

Twitter adalah situs web dimiliki dan dioperasikan oleh Twitter, Inc., yang menawarkan jaringan sosial berupa microblog. Disebut microblog karena situs ini memungkinkan penggunaannya mengirim dan membaca pesan blog seperti pada umumnya namun terbatas hanya sejumlah 140 karakter yang ditampilkan pada halaman profil pengguna.. Twitter memiliki karakteristik dan format penulisan yang unik dengan simbol ataupun aturan khusus.Pesan dalam Twitter dikenal dengan sebutan tweet. [6]

C. Transportasi KRL Commuter Line

KA Commuter Jabodetabek (atau disebut juga KRL Commuter Line, dulu dikenal sebagai KRL Jabotabek) adalah jalur kereta rel listrik yang dioperasikan oleh PT Kereta Api Indonesia (KAI) Commuter Jabodetabek, yang merupakan anak perusahaan dari PT Kereta Api Indonesia (PTKAI). KRL telah beroperasi di wilayah Jakarta sejak tahun 1976, hingga kini melayani rute komuter di wilayah DKI Jakarta, Kota Bogor, Kabupaten Bogor, Kota Bekasi, Kabupaten Lebak, Kota Tangerang, dan Kota Tangerang Selatan. KRL Commuter Line bukan hanya alternatif pilihan transportasi rakyat yang murah, tetapi juga bebas dari kemacetan jalan raya ibu kota. Bahkan, kereta api bukan hanya pilihan bagi mereka yang berekonomi menengah ke bawah, tetapi juga menjadi

gaya hidup mereka yang "berduit" tetapi mencari kepraktisan serta kenyamanan untuk mencapai tempat tujuan. [11]

Kereta api adalah salah satu moda transportasi publik yang sangat dibutuhkan masyarakat untuk menjalankan kehidupan sehari-hari. PT KAI Commuter Jabodetabek selaku operator jasa transportasi kereta api di wilayah jabodetabek menciptakan layanan KRL gerbong khusus wanita untuk meningkatkan kualitas layanan bagi penumpang wanita. Dengan adanya pemisahan gerbong antara gerbong umum dengan gerbong khusus wanita, diharapkan kaum perempuan pekerja dan yang berpergian akan merasakan keamanan dan kenyamanan saat menggunakan KRL. [12]

D. Naive Bayes

Algoritma naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam penelitian ini yang menjadi data uji adalah dokumen tweets. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.[13]

Naive Bayes Classifier merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh Bayes adalah Teorema Bayes, yaitu melakukan klasifikasi dengan melakukan perhitungan nilai probabilitas. Klasifikasi dilakukan untuk menentukan kategori dari suatu dokumen. Sebuah keuntungan dari Naive Bayes Classifier adalah bahwa akan hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel-variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya varians dari variabel untuk masing-masing kelas harus ditentukan dan tidak seluruh matriks kovariansi.[14]

E. KNN

K-nearest neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data

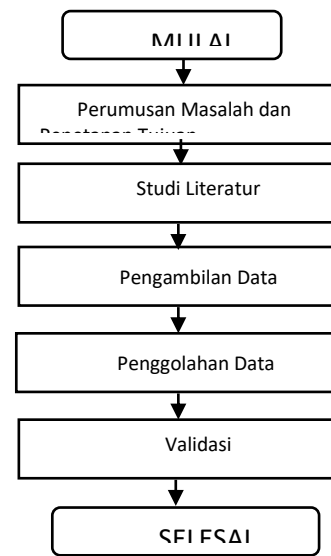
pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data, secara umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antarasetiap klasifikasi menjadi lebih kabur.[17]

Evaluasi digunakan untuk mengukur kinerja metode klasifikasi, dalam penelitian ini digunakan untuk mengukur keakuratan metode klasifikasi yang diukur dengan akurasi, precision dan recall. Recall didefinisikan sebagai persentase antara data kelas data buruk yang dikelaskan dengan benar dan data kelas data buruk yang salah diprediksi ke kelas data baik. Precision adalah persentase dari kelas data buruk yang dikelaskan dengan benar dan kelas yang seharusnya termasuk kelas data baik tetapi dikelaskan sebagai kelas data buruk.

F. Decision Tree

Decision tree merupakan salah satu metode untuk mengklasifikasikan data. Model decision tree merupakan sebuah tree yang terdiri dari root node, internal node dan terminal node. Sementara root node dan internal node merupakan variabel/fitur, terminal node adalah label kelas. Dalam melakukan klasifikasi, sebuah data query akan menelusuri root node dan internal node sampai mencapai terminal node. Pelabelan kelas data query berdasarkan label di internal node. Pada Decision Tree tradisional, data yang digunakan adalah data dengan nilai fitur yang sudah pasti. Pada penelitian ini bertujuan untuk membangun suatu Decision Tree yang dapat menangani data dengan nilai fitur yang tidak pasti yaitu data yang memiliki range nilai. Selama ini penanganan data dengan fitur yang tidak pasti menggunakan metode averaging. Akan tetapi ada penelitian lain yang menyebutkan bahwa ada metode lain yang lebih handal dalam menangani data dengan nilai fitur yang tidak pasti. Metode tersebut adalah metode berbasis distribusi.[18]

3. METODE PENELITIAN



Gambar 1 Metodologi Penelitian

a. Perumusan Masalah dan Penetapan Tujuan

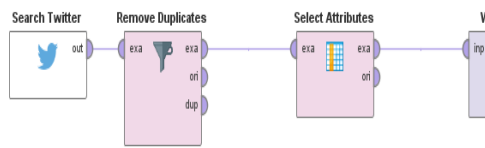
Pada tahapan ini langkah utama sebelum melakukan penelitian ini adalah menentukan masalah yang terjadi dan menetapkan tujuan sehingga agar tepat guna dalam melakukan setiap proses dari awal hingga akhir.

b. Studi Literatur

Sebelum melakukan proses penelitian langsung terhadap kondisi yang terjadi, dalam tahapan ini melihat beberapa *reference* baik buku, paper, media sosial, portal berita dan lain sebagainya yang akan digunakan untuk penyusunan penelitian ini.

c. Pengambilan Data

Pada tahap pengambilan data dalam penelitian ini, data dikumpulkan dari sosial media twitter dengan cara random dan dalam kondisi paling baru (*ter-update*) saat proses pengumpulan data dimulai. Pada proses ini menggunakan *Rapidminer* sebagai alat untuk mengambil data pada sosial media twitter secara random dengan langkah pada gambar berikut.



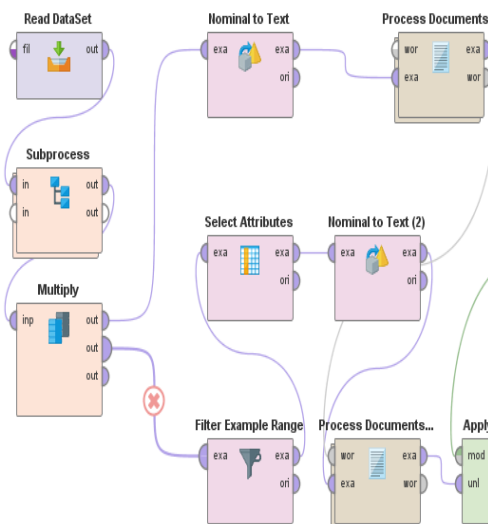
Gambar 2. Proses Pengambilan Data

Langkah pertama pengambilan data pada sosial media twitter secara random sebanyak 150 data. Kemudian untuk mengatasi data yang rangkap dilakukan tahap *Remove Duplikat* pada data yang telah diambil dari sosial media twitter 150 data menjadi 127 data. Setelah dilalui proses tersebut, data yang akan dianalisis sentimen adalah bagian text atau komentar terhadap KRL Commuter Line Jabodetabek. Maka dilakukan proses *Select Attribute* untuk memilih bagian text pada data.

Setelah melalui proses tersebut data diekspor dalam bentuk excel, kemudian setelah itu diberikan label antara sentimen positive maupun negative

d. Pengolahan Data

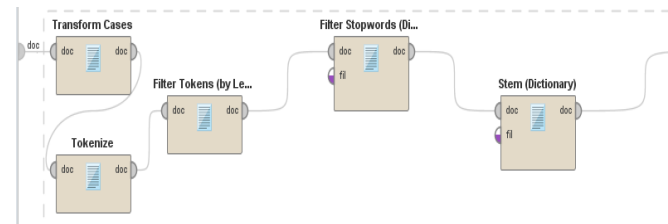
Data yang sudah tersedia dalam bentuk excel akan diolah pada *Rapidminer* dengan beberapa tahapan proses yang akan dilakukan seperti gambar di bawah ini.



Gambar 3. Proses Pengolahan Data

Data training berupa excel yang akan di *convert* menjadi *Nominal to text* untuk mengubah sebuah tipe data yang ada di setiap kalimat, kemudian melalui tahap *Process Document From Data* yang berfungsi untuk mengolah data dari

transform case, Tokenize, Filter Token, Filter Stopword, dan stemming. Berikut gambar proses dokument yang diolah.



4. Gambar *Process Dokument From Data*

Transform Case merupakan tahap mengkonversi semua tulisan yang terdapat pada data training berbentuk huruf kecil semua, kemudian tokenize adalah proses memisahkan kata disetiap kalimat yang ada.

E. Validasi

Naive Bayes Classifier menganggap sebuah dokumen sebagai kumpulan dari kata-kata yang menyusun dokumen tersebut, dan tidak memperhatikan urutan kemunculan kata pada dokumen, sehingga perhitungan probabilitas dapat dianggap sebagai hasil perkalian dari probabilitas kemunculan kata-kata pada dokumen. Perhitungan *Naive Bayes Classifier* dapat dilakukan dengan persamaan di bawah ini.

$$f(W_{kj}C_i) = \frac{(W_{kj}C_i) + 1}{F(C_i) + |W|}$$

Keterangan :

1. $f(W_{kj}C_i)$ = Nilai kemunculan kata pada W_{kj} pada kategori C_i
2. $f(C_i)$ = Jumlah keseluruhan kata pada kategori C_i
3. $|w|$ = Jumlah keseluruhan kata/fitu yang digunakan

$$p(C_i) = \frac{Fd(C_i)}{|D|}$$

Keterangan :

1. $Fd(C_i)$ = Jumlah dokumen yang memiliki pada kategori C_i

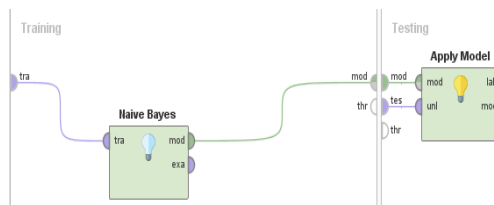
2. $|D|$ = Jumlah keseluruhan tranning dokumen

Untuk penentuan kategori dokumen :

$$C^* = \underset{C \in C}{\text{arg}} \max p_{\mu} p(W_{kj}C_i)$$

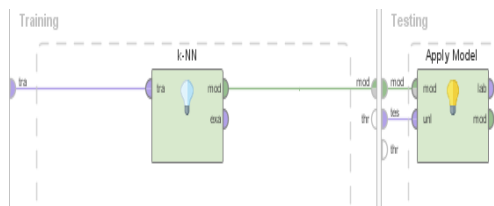
Keterangan :

1. C^* = Kategori yang memiliki nilai probabilitas kemunculan $\underset{C \in C}{\text{arg}} \max p_{\mu} p(W_{kj}C_i)$
2. $\underset{C \in C}{\text{arg}} \max p_{\mu} p(W_{kj}C_i)$ = Nilai maksimal dari seluruh data training.



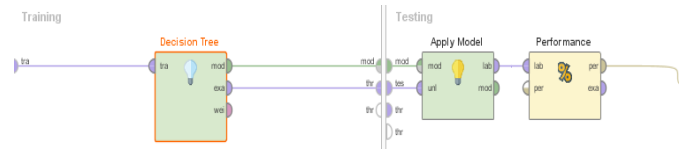
Gambar 5. Proses Cross Validation Naive bayes

Pada tahapan tersebut menggunakan metode *naive bayes* dilakukan dari data training terhadap data testing dimana diambil 5 Record dari keseluruhan data yang ada. Sehingga menghasilkan akurasi dari pengolahan data.



Gambar 6. Proses Cross Validation KNN

Pada metode *KNN* dilakukan dari data training terhadap data testing dimana diambil 5 Record dari keseluruhan data yang ada.



Gambar 7. Proses Cross Validation Decision Tree

Demikian dengan hal yang sama pada metode *Decision Tree* dilakukan dari data training terhadap data testing dimana diambil 5 Record dari keseluruhan data yang ada.

4. HASIL DAN PEMBAHASAN

Dalam hasil dan pembahasan ini penelitian analisis sentimen dari pengguna KRL Commuter line Jabodetabek dengan metode Naive bayes yang diolah dengan menggunakan Rapidminer terdapat gambar di bawah ini.

Sentimen	prediction(S...
Positive	Positive
Positive	Positive
Positive	Negative
Negative	Negative
Negative	Negative

Gambar 8. Hasil Prediction Naive Bayes Hasil prediction pada metode Naive Bayes menunjukkan bahwa dalam 5 record data testing menyatakan 4 hasil yang sama dan satu hal yg berbeda antara sentimen dan *prediction*.

Sentimen	prediction(S...
Positive	Positive
Positive	Positive
Positive	Positive
Negative	Negative
Negative	Positive

Gambar 9. Hasil Prediction KNN

Hasil prediction pada metode KNN menunjukkan bahwa dalam 5 record data testing menyatakan hasil yang sama antara sentimen dan *prediction*.

Sentimen	prediction(S...
Positive	Positive
Positive	Positive
Positive	Positive
Negative	Negative
Negative	Negative

Gambar 10. Hasil Prediction Decision Tree

Hasil prediction pada metode Decision Tree menunjukkan bahwa dalam 5 record data testing menyatakan hasil yang sama antara sentimen dan *prediction*. Terdapat persamaan dengan metode KNN.

accuracy: 80.00%

	true Positive	true Negative
pred. Positive	2	0
pred. Negative	1	2
class recall	66.67%	100.00%

Gambar 11. Hasil Akurasi Naive Bayes

Hasil akurasi pada metode Naive Bayes menyatakan sebesar 80%, diantara lain *predicted class positive* dan *true positive* (TP) menghasilkan 2 sama halnya dengan *predicted class negative* dan *true negative* (TN) menghasilkan 2, kemudian untuk *predicted class negative* dan *true positif* (TP) menghasilkan 1.

accuracy: 80.00%

	true Positive	true Negative
pred. Positive	3	1
pred. Negative	0	1
class recall	100.00%	50.00%

Gambar 12. Hasil Akurasi KNN

Metode KNN menyatakan sebesar 80%, diantara lain *predicted class positive* dan *true positive* (TP) menghasilkan 3 *predicted class positive* dan *true negative* (TN) mehasilkan 1, kemudian untuk *predicted class negative* dan *true negative* (TN) menghasilkan 1.

accuracy: 100.00%

	true Positive	true Negative
pred. Positive	3	0
pred. Negative	0	2
class recall	100.00%	100.00%

Gambar 13. Hasil Akurasi Decision Tree

Hasil akurasi pada metode Decision Tree menyatakan sebesar 100%, diantara lain *predicted class positive* dan *true positive* (TP) menghasilkan 3 sedangkan *predicted class negative* dan *true negative* (TN) menghasilkan 2.

precision: 66.67% (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	2	0	100.00%
pred. Negative	1	2	66.67%
class recall	66.67%	100.00%	

Gambar 14. Hasil Precision Naive Bayes

Pada gambar diatas adalah hasil dari pengolahan data sentimen dengan menggunakan Naive Bayes yang menghasilkan precision sebesar 66,67 % dengan *predicted class positive* dan *true positif* (TP) sebanyak 2, sama dengan *class prediction negative* dan *true negative* (TN), sedangkan *class prediction negative* dan *true positive*(TP) sebanyak 1.

precision: 100.00% (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	3	1	75.00%
pred. Negative	0	1	100.00%
class recall	100.00%	50.00%	

Gambar 15. Hasil Precision KNN

Pada gambar diatas adalah hasil dari pengolahan data sentimen dengan menggunakan KNN yang menghasilkan precision sebesar 100 % dengan *predicted class positive* dan *true positif* (TP) sebanyak 3, *class prediction positive* dan *true negative* (TN) sebanyak 1 sama halnya dengan *class prediction negative* dan *true negative* (TN).

precision: 100.00% (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	3	0	100.00%
pred. Negative	0	2	100.00%
class recall	100.00%	100.00%	

Gambar 16. Hasil Precision Decision Tree

Pada metode Decision Tree yang menghasilkan precision sebesar 100 % dengan *predicted class positive* dan *true positif* (TP) sebanyak 3 sedangkan *class*

prediction negative dan true negative (TN) sebanyak 2.

sensitivity: 100.00% (positive class: Negative)

	true Positive	true Negative
pred. Positive	2	0
pred. Negative	1	2
class recall	66.67%	100.00%

Gambar 17. Hasil Sensitivity Naive Bayes

Dari hasil diatas dapat diketahui bahwa sensitivity pada metode Naive Bayes sebesar 100% pada *positive class:negative*.

sensitivity: 50.00% (positive class: Negative)

	true Positive	true Negative
pred. Positive	3	1
pred. Negative	0	1
class recall	100.00%	50.00%

Gambar 18. Hasil Sensitivity KNN

Dari hasil diatas dapat diketahui bahwa sensitivity pada metode KNN sebesar 50% pada *positive class:negative*.

sensitivity: 100.00% (positive class: Negative)

	true Positive	true Negative
pred. Positive	3	0
pred. Negative	0	2
class recall	100.00%	100.00%

Gambar 19. Hasil Sensitivity Decision Tree

Dari hasil diatas dapat diketahui bahwa sensitivity pada metode Decision Tree sebesar 100% pada *positive class:negative*.

specificity: 66.67% (positive class: Negative)

	true Positive	true Negative
pred. Positive	2	0
pred. Negative	1	2
class recall	66.67%	100.00%

Gambar 20. Hasil Specificity Naive Bayes

Diperoleh data specificity dari metode Naive Bayes sebesar 66,67% yang menghasilkan *class precision positive* sebesar 100 % dan *class precision negative* 66,67%.

specificity: 100.00% (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	3	1	75.00%
pred. Negative	0	1	100.00%
class recall	100.00%	50.00%	

Gambar 21. Hasil Specificity KNN

Diperoleh data specificity dari metode KNN sebesar 100% yang menghasilkan *class precision positive* sebesar 75,00 % dan *class precision negative* 100%.

specificity: 100.00% (positive class: Negative)

	true Positive	true Negative	class precision
pred. Positive	3	0	100.00%
pred. Negative	0	2	100.00%
class recall	100.00%	100.00%	

Gambar 22. Hasil Specificity Decision Tree

Diperoleh data specificity sebesar 100% yang menghasilkan *class precision positive* sebesar 100 % dan *class precision negative* 100%.

5. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada bahasan sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut

Analisis sentimen terhadap data Twitter mengenai penggunaan transportasi KRL Commuter Line Jabodetabek dapat dilakukan dengan metode Naive Bayes, dengan akurasi mencapai 80,00% , precision 66,67 % , sensitifity 100% dan Specificity 66,67 % . Pada metode KNN akurasi sebesar 80%, precision 100%, sensitivity 50%, specificity 100% dan pada metode Decision Tree akurasi sebesar 100%, precision 100%, sensitivity 100%, specificity 100%.

Saran yang dikemukakan dalam penelitian ini diharapkan dapat lebih meningkatkan hasil yang telah didapatkan. Berikut beberapa saran yang disampaikan sebagai pengembangan dalam penelitian selanjutnya.

- a. Jumlah data yang diolah diharapkan lebih banyak dari yang sebelumnya, agar dapat memberikan informasi pendukung yang mampu membantu dalam menangani kendala yang ada dalam pelayanan transportasi KRL Commuter Line Jabdetabek.
- b. Perbandingan metode dapat dilakukan dengan berbagai macam metode yang lainnya.

DAFTAR PUSTAKA

- A. Novantirani, M.K Sabariah, dan V. Effendy, "Analisis Sentimen Pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," vol. 2 no. 1, pp. 1177-1183, April 2015
- A. Pak, dan P. Paroukbek, "Twitter As a Corpus for Sentiment Analysis and Opinion Mining", LIMSI-CNRS. France, pp. 1320-1326.
- A. Roshdi, dan A. Roohparvar, "Review : Information Retrieval Techniques and Applications", IJCNCs. Iran, vol. 3 no 9, pp. 373-377, September 2015.
- D. Rustina, dan N. Rahayu, "Analisis Sentimen Pasar Otomotif Mobil : Tweet Twitter Menggunakan Naive Bayes", SIMETRIS. Indonesia, vol. 8 no 1, pp. 113-120, April 2017.
- F. Zulkifli, dan Syahputra, "Analisis Faktor Kualitas Pelayanan KRL Commuter Line Jabodetabek Gerbong Khusus Wanita (Studi Pada PT KAI Commuter Jabodetabek Rute Bogor-Jakarta Kota)". Indonesia, vol 3 no 2, pp. 2007-2009, Agustus 2016.
- G. Vinodhini, dan RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining", IJARCSSE. India, vol. 2, pp. 282-292, Juni 2012.
- Indriati, dan A. Ridok, "Sentiment Analysis For Review Mobile Application Using Neighbor Method Weighted K-Nearest Neighbor(NWKNN)", JEEEST. Indonesia : Malang, vol. 3 no 01, pp. 23-32, Juli 2016.
- K. Zuva, dan T. Zuva, "Evaluation Of Information Retrieval Systems", IJCSIT. Kanada, vol. 4 no 3, pp. 35-43, Juni 2012.
- Liu, B. "Sentiment Analysis and Opinion Mining. San Rafael". Morgan & Claypool Publishers. 2012.
- L. Zhang et al., "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis," HPL-2011-89, vol. 89, Jun. 2011.
- Nazwirman, dan Hulmansyah, "Karakteristik Penumpang Pengguna KRL Commuter Line Jabodetabek", JEBA. Indonesia, vol. 2 no 1 , pp.26-35, Juli 2017.
- P. Gamallo, dan M. Garcia, "A Naive Bayes Strategy for Sentiment Analysis on English Tweets", SemEval. Ireland : Dublin, pp. 171-175, Agustus 2014.
- Tan Songbo et al, "Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis". ECIR, China, 2009, pp. 337-349
- V. Narayanan, I. Arora, and A. Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model", IDEAL. Varanasi: India, 2013, pp.194-201.
- W. Witanti, H. Rahmanto dan F. Renaldi, "Pembangunan Sistem Temu Balik Informasi (Information Retrieval) Dalam Pemilihan Pemain Sepak Bola Berkualitas Di Indonesia Berbasis Analisis Sentimen", SENTIKA. Indonesia : Yogyakarta, pp. 484-491, Maret 2016.