

SENTIMEN ANALISIS KEBIJAKAN GANJIL GENAP DI TOL BEKASI MENGUNAKAN ALGORITMA NAIVE BAYES DENGAN OPTIMALISASI INFORMATION GAIN

Heru Sukma Utama¹, Didi Rosiyadi², Dedi Aridarma³, Bobby Suryo Prakoso⁴

^{1,2,3,4} Program Studi Magister Ilmu Komputer
STMIK Nusa Mandiri
www.nusamandiri.ac.id

14002126@nusamandiri.ac.id; didi.rosiyadi@gmail.com; 14002154@nusamandiri.ac.id;
14002107@nusamandiri.ac.id

Abstract—*Sentiment analysis of odd even numbered systems in Bekasi toll using the Naive Bayes Algorithm, is a process of understanding, extracting, and processing textual data automatically to obtain a sentiment information contained in an opinion sentence. After doing the weighting process, the data goes through the classification stage using the Naive Bayes method. Naive Bayes is a technique for making predictions, both in the case of classification and regression. In addition there are stages of text mining that are broadly defined as intensive knowledge processes that allow users to interact with a collection of documents from time to time using various kinds of analysis. The research method used in this study is to do text mining on comments related to posts about even odd effectiveness on Bekasi toll on Twitter, Instagram, Youtube and Facebook. In doing text mining the following steps will be carried out: selection, preprocessing, transformation, datamining and evaluation. The text mining tool that will be used is using Rapidminer version 9.0 for the modeling front that is used. Then the analytical sentiment is carried out from tweets, comments and post submitted by the public. After the research was done, the results showed that the result of the tests carried out through the Naive Bayes model resulted in Confusion Matrix, namely accuracy of 79.55%, Precision of 80,37.03%, and Sensitivity or Recall of 80.51%.*

Keywords : *Sentiment, Analysis, Text Mining, Naive Bayes*

Intisari—*Sentimen analisis sistem ganjil genap di tol Bekasi menggunakan Algoritma Naive Bayes, merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Setelah dilakukannya proses pembobotan, data tersebut melalui tahap pengklasifikasian dengan menggunakan metode Naive Bayes. Naive Bayes adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Selain itu*

terdapat tahapan text mining yang didefinisikan secara luas sebagai proses pengetahuan intensif yang memungkinkan pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu menggunakan berbagai macam analisis. Metode penelitian yang dilakukan dalam penelitian ini adalah melakukan text mining pada komentar-komentar terkait postingan mengenai efektifitas ganjil genap di tol bekasi pada Twitter, Instagram, Youtube dan Facebook. Dalam melakukan text mining akan dilakukan tahap-tahap sebagai berikut : selection, preprocessing, transformation, datamining dan evaluation. Text mining tool yang akan digunakan adalah menggunakan Rapidminer versi 9.0 untuk penerepan permodelan yang digunakan. Lalu dilakukan sentimen analisis dari tweet, komentar dan postingan yang disampaikan oleh masyarakat luas. Setelah dilakukan penelitian dengan menggunakan model Naive Bayes maka didapatkan hasil Confusion Matrix, yaitu accuracy sebesar 79,55%, Precision sebesar 80,37%, dan Sensitivity atau Recall sebesar 80,51%.

Kata Kunci : *Sentimen Analisis, Text Mining, Naive Bayes*

PENDAHULUAN

Analisis sentimen atau *opinion mining* merupakan salah satu solusi mengatasi masalah untuk mengelompokkan opini atau *review* menjadi opini positif atau negatif secara otomatis (Indrayuni, 2016). Dengan analisis sentimen ini, maka pendapat seseorang dapat beri label positif maupun negatif. Analisis sentimen dibutuhkan untuk mendapatkan informasi yang dapat digunakan sebagai manajemen reputasi dan sarana evaluasi bagi perusahaan (Oktasari, Chrisnanto, & Yuniarti, 2016).

Salah satu media yang dapat digunakan dalam melakukan analisis sentimen adalah media sosial (Oktasari et al., 2016). Media Sosial menjadi sarana dalam menunjang perusahaan atau

organisasi untuk kegiatan penyampaian informasi kepada masyarakat seperti yang dapat dilihat pada perusahaan penyedia layanan asuransi (Oktasari et al., 2016). Media sosial mempunyai pengaruh yang besar dalam kehidupan seseorang. Seseorang yang awalnya kecil bisa menjadi besar dengan media sosial, atau sebaliknya (Putri, Nurwati, & S., 2016).

Jalan tol di Indonesia disebut juga sebagai jalan bebas hambatan yaitu jalan yang dikhususkan untuk kendaraan bersumbu dua atau lebih (mobil, bus, truk) dan bertujuan untuk mempersingkat jarak dan waktu tempuh dari satu tempat ke tempat lain. Sistem ganjil genap adalah satu konsep pembatasan kendaraan yang mengacu pada dua nomor terakhir pelat nomor kendaraan. Dengan begitu, nantinya setiap kendaraan yang melintas akan bergantian sesuai hari pemberlakuan dua digit angka terakhir pelat nomornya.

Metode penelitian yang dilakukan dalam penelitian ini adalah melakukan *text mining* pada komentar-komentar terkait postingan mengenai efektifitas ganjil genap di tol Bekasi pada *Twitter*, *instagram*, *youtube* dan *facebook*. Lalu dilakukan sentimen analisis dari tweet, komentar dan postingan yang disampaikan oleh masyarakat luas. Text mining adalah salah satu teknik penambangan data yang berupa teks (Oktasari et al., 2016). Dengan text mining tentunya kita bisa menggali data yang berasal dari pernyataan-pernyataan. Pendapat yang lain juga menyebutkan *Text mining* merupakan salah satu teknik yang digunakan untuk menggali kumpulan dokumen text sehingga dapat diambil intisariannya (Budi, 2017).

Salah satu sumber informasi yang dibutuhkan oleh Pemerintah untuk dapat meningkatkan kinerjanya adalah umpan balik dari masyarakat (Faradhillah, Kusumawardani, & Hafidz, 2016). Oleh karena itu penelitian ini sangat diperlukan untuk membantu pemerintah dalam mengevaluasi program yang dijalankan. Tanpa adanya evaluasi dari masyarakat, maka pemerintah tidak akan mengetahui sejauh mana keberhasilan program yang dijalankan. Setelah dilakukannya proses pembobotan, data tersebut akan melalui tahap pengklasifikasian dengan metode *Naive Bayes* agar dapat melakukan analisis dari sekumpulan contoh dokumen yang telah diklasifikasikan sebelumnya.

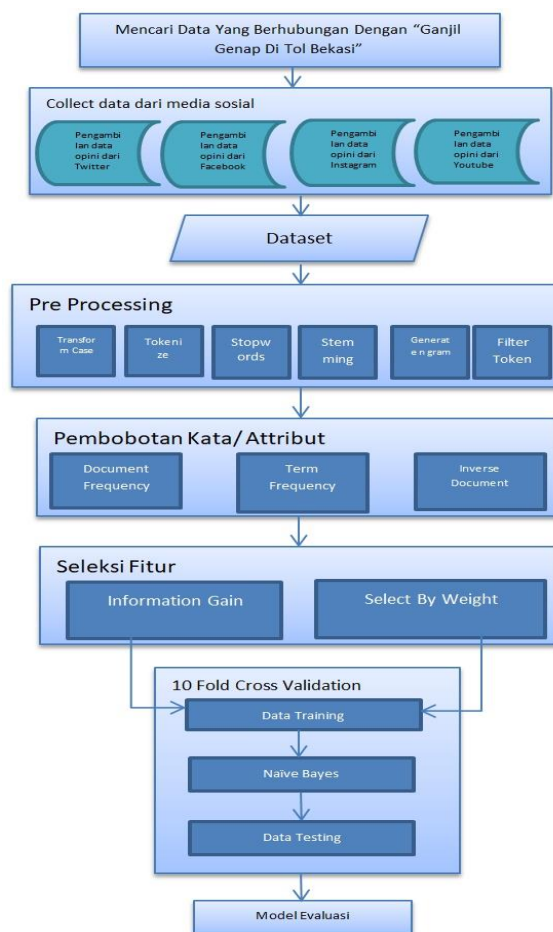
Tujuan Penelitian ini adalah mengetahui tingkat *accuracy*, *recall* dan *precision* dari *opinion mining* yang dihasilkan menggunakan algoritma *Naive Bayes*, memberikan informasi sentimen masyarakat terhadap efektifitas sistem ganjil genap tol bekasi di media sosial.

Penelitian ini pernah dilakukan sebelumnya oleh Nanang Ruhyana, 2019 menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil

genap, dalam penelitian ini menghasilkan *accuracy* 86,67%, *precision* 71,43% dan *recall* 80,00% (Ruhyana, 2019) dan Dwi Suci Ariska Yanti, Indriati, Putra Pandu Adikara, 2019 berdasarkan hasil pengujian, sistem ini memiliki nilai *F-Measure* tertinggi sebesar 66,1% dan nilai akurasi sebesar 66,5% (Suci, Yanti, & Adikara, 2019). Untuk mendapatkan solusi dari permasalahan di atas, penelitian difokuskan untuk menjawab pertanyaan riset sebagai berikut : Bagaimana cara memanfaatkan *Twitter*, *instagram*, *youtube*, *facebook* untuk menganalisis sentiment efektifitas sistem ganjil genap tol bekasi menggunakan seleksi fitur *Information Gain* dengan *Naive Bayes*? dan Bagaimana mengukur tingkat *accuracy*, *recall* dan *precision* dari *opinion mining* yang dihasilkan menggunakan seleksi fitur *Naive Bayes* Dan melihat sentiment masyarakat terhadap efektifitas sistem ganjil genap tol bekasi ?

BAHAN DAN METODE

Rancangan penelitian yang akan dilakukan disajikan dalam bentuk kerangka berpikir yang disajikan pada Gambar 1 berikut :



Sumber: (Utama, Rosiyadi, Prakoso, & Ariadarma, 2019)

Gambar 1. Kerangka Berpikir Penelitian

Dalam melakukan *text mining* akan dilakukan tahap-tahap sebagai berikut : *selection, preprocessing, transformation, datamining* dan *evaluation*. *Text mining tool* yang akan digunakan adalah menggunakan *Rapidminer* versi 9.0 untuk penerepan permodelan yang digunakan. Penggunaan *Rapidminer* sendiri memungkinkan untuk menerapkan algoritma yang telah disebutkan pada bagian tahap *modelling*, lalu penerapan algoritma *Naïve Bayes*. Penelitian ini memiliki hipotesis bahwa penggunaan algoritma *Naïve Bayes* dapat menentukan pola terbaik dengan mempelajari sentiment analisis berupa text yang ada saat ini dan mempercepat dalam proses perhitungannya. Penelitian ini juga diharapkan dapat memberikan kontribusi dalam bidang *text mining* berbahasa indonesia. Objek penelitian ini berkaitan dengan data mining dan *text mining*, yang bersumber datanya berupa data teks yang diambil dari media *social*, yang bertepatan ganjil genap Bekasi. Media *social* tersebut terdiri dari *facebook, twitter, instagram* dan *youtube*. Keempat media *social* tersebut termasuk dalam paling banyak terakses oleh para pengguna internet yang bisa memberikan gambaran penggunaan dalam dataset yang digunakan.

Tahap *selection* merupakan tahap pertama yaitu tahap mencari data. Data yang digunakan sebagai bahan mentah penelitian. Data tersebut kemudian diseleksi dengan fitur dan mengujinya dengan algoritma *naïve bayes*. Kemudian pemahaman tersebut diubah menjadi sebuah rencana awal *data mining* yang dirancang untuk mencapai tujuan. Sedangkan tahap *Preprocessing* terdiri dari beberapa tahap yaitu *cleansing, tokenizing, stopword removal, dan stemming* (Hidayat, 2015). *Cleansing data* yaitu mengurangi *noise* pada data *tweet, transform case* tahapan ini merupakan proses merubah bentuk huruf menjadi huruf kecil (*lower case*) atau dapat disebut juga penyeragaman bentuk huruf. *Stopword Removal*, merupakan proses menghilangkan daftar kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan seperti “yang”, “di”, “ke”, “itu” dan lain sebagainya, *tokenizing* atau *parsing* adalah tahap pemotongan *string* input berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses *tokenizing* adalah pemenggalan kalimat menjadi kata, *filter tokens* adalah tahap dimana menghilangkan kata yang dikonfigurasi untuk dihilangkan berdasarkan jumlah hurufnya. Sebagai contoh *filter tokens* dengan minimal 3 huruf maka kata “ya”, “yg”, dan “kk” akan hilang pada kalimat tersebut. Tahap ini juga sangat penting karena akan mempengaruhi akurasi penelitian. Banyak kata-kata yang tidak baku, karena dimedsos biasanya orang menuliskan kalimat dan kata

dengan bahasa tidak resmi. Demikian juga dalam penggunaan tanda bacanya.

Pengurutan data atau *sorting* merupakan suatu proses dimana suatu susunan data yang semula dalam kondisi acak dapat menjadi urut, baik dari data terkecil sampai dengan data yang terbesar, atau sebaliknya dari data terbesar sampai dengan data terkecil (Wahyudi, 2009). Tahapan ini sangat penting, karena memudahkan peneliti dalam menganalisis data.

Algoritma klasifikasi yang dapat melakukan teks mining diantaranya *Support Vector Machine* (SVM), *Naïve Bayesian classification* (NBC) dan *K-Nearest Neighbor* (K-NN) (Ipmawati, Kusri, & Taufiq Luthfi, 2017). Metode *Naïve Bayes* adalah suatu metode yang digunakan untuk memprediksi berbasis probabilitas (Wasiati & Wijayanti, 2014). *Naïve bayes* merupakan algoritma klasifikasi data mining yang menganggap masing-masing atribut tidak saling berhubungan (Wahyuni, Sutojo, & Luthfiarta, 2004). Algoritma *Naïve Bayes* bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan data transaksi penjualan dalam memprediksi perilaku pola pembeli (Kom, Dewi, Mulyani, Kom, & Nurhasanah, 2015). *Naïve Bayes* merupakan teknik prediksi berbasis probabilitik sederhana yang berdasar pada penerapan teorema atau aturan bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Yuda Septian, 2009).

Information Gain merupakan metode seleksi fitur paling sederhana dengan melakukan perangkikan atribut dan banyak digunakan dalam aplikasi kategorisasi teks, analisis data *microarray* dan analisis data citra (Divya & Vijaya Babu, 2016). *Information Gain* dapat membantu mengurangi *noise* yang disebabkan oleh fitur-fitur yang tidak relevan. *Information Gain* mendeteksi fitur-fitur yang paling banyak memiliki informasi berdasarkan kelas tertentu.

Dalam penelitian ini, data yang akan menjadi bahan penelitian adalah tentang opini masyarakat berkaitan dengan penerapan ganjil genap di tol Bekasi. Opini-opini tersebut diambil dari media sosial. Dilihat dari permasalahan yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran maupun keluhan yang diterima oleh Jasa Marga selaku operator jalan tol Bekasi sehingga dapat diketahui informasi sentimen mengenai efektifitas ganjil genap di tol bekasi. Adapun permasalahan pengklasifikasian sebuah kalimat sentimen ke dalam kelas – kelas tertentu dapat diselesaikan dengan *Naïve Bayes*. Rumus untuk menghitung *entropy* ditunjukkan pada persamaan berikut. Setelah mendapatkan nilai

entropy, maka perhitungan *Information Gain* dapat dilakukan dengan menggunakan persamaan (Aini, Sari, & Arwan, 2018)

$$Entropy(s) = \sum_i^c - P_i \log_2 P_i \dots \dots \dots (1)$$

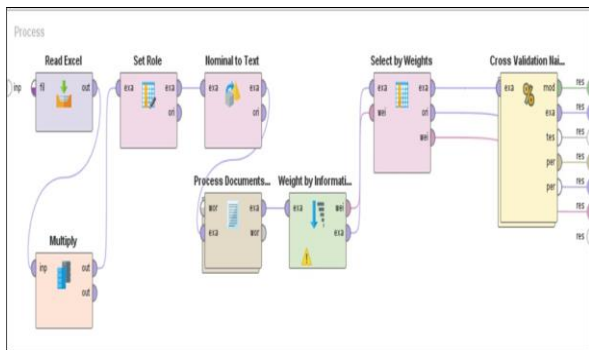
Dengan *c* adalah jumlah nilai yang ada pada kelas klasifikasi dan *P_i* merupakan jumlah sampel untuk kelas *i*.

$$Gain(S, A) = Entropy(s) - \sum_{Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \dots \dots (2)$$

Dengan *A* merupakan atribut, *v* adalah nilai yang mungkin untuk atribut *A*, *Values(A)* adalah himpunan nilai-nilai yang mungkin untuk *A*, *|S_v|* adalah jumlah sampel untuk nilai *v*, *|S|* merupakan jumlah seluruh sampel data dan *Entropy(S_v)* adalah *entropy* untuk sampel-sampel yang memiliki nilai *v*.

HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk mengetahui tingkat *accuracy*, *recall* dan *precision* dari *opinion mining* yang dihasilkan menggunakan algoritma *Naïve Bayes* memberikan informasi sentiment masyarakat terhadap efektifitas sistem ganjil genap tol bekasi di media sosial : *Twitter*, *instagram*, *youtube*, *facebook*. Proses pemodelan menggunakan metode *Naïve Bayes* seperti tampak pada Gambar 2 berikut :



Sumber: (Utama et al., 2019)
Gambar 2. Proses Pemodelan Menggunakan *Naïve Bayes*

Gambar 2 menunjukkan langkah-langkah penggunaan Algoritma dari awal sampai akhir. Sedangkan dataset yang digunakan pada penelitian ini berjumlah 440 dataset, terbagi menjadi dua yang terdiri dari 220 set bernilai positif dan 220 set bernilai negatif seperti pada Tabel 1 berikut adalah contoh beberapa data yang diambil dari media sosial, yang kemudian di olah menjadi data utama penelitian. Berikut disajikan Tabel 1.

Tabel 1. Dataset Penelitian

No	Nilai	Ulasan
1	positif	Pak, kl mobil dari Bandung menuju Jakarta bagaimana pengenaaan ganjil genapny
2	positif	Semoga solusi pemerintah bisa berjalan seperti apa yg diharapkan.. terlalu banyak pengguna roda 4 mungkin ya di ibu kota
3	positif	Nah begini negara bisa maju kayanya eheheh
4	positif	Bagus pak polisi lanjutkan
5	positif	ayo galakkan naik krl dan busway walau penuh tp itulah yg namanya nyari duit di ibukota
6	negatif	Ganjil genap kayak gimana???
7	negatif	Puyeng puyang dah yang kaya gini . Kasian yang gak paham dan tbatba di tilang ... ahhsudahlah!
8	negatif	Demem bgt ribetttttttt
9	negatif	Polisi nya jd maen gala asin gtu...
10	negatif	Tambah ruwet ju

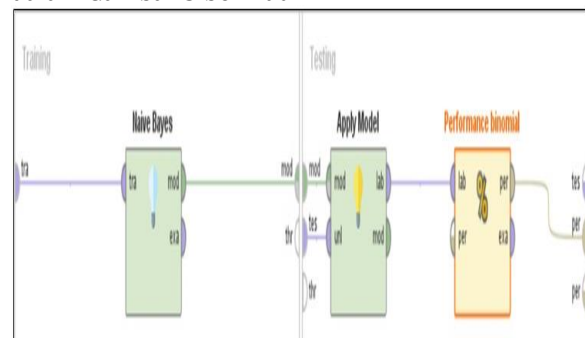
Sumber: (Utama et al., 2019)

Data pada Tabel 1 tersebut diolah dengan melalui tahapan yang terdapat pada Gambar 3 berikut:



Sumber: (Utama et al., 2019)
Gambar 3. Tahap *Preprocessing*

Hasil pengolahan data yang dilakukan dengan menggunakan algoritma *Naïve Bayes* disajikan dalam Gambar 3 berikut :



Sumber: (Utama et al., 2019)
Gambar 4. Cross Validation *Naïve Bayes*

Evaluation

Tahap evaluasi merupakan tahapan dimana dilakukan interpretasi terhadap hasil *text mining* yang telah dihasilkan pada tahapan sebelumnya. Evaluasi yang dilakukan, akan dilakukan secara mendalam dengan tujuan agar hasil pada tahapan sebelumnya sudah sesuai dengan tujuan yang *text data mining* yang dilakukan, telah sesuai dengan keinginan perusahaan. Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC* serta ditampilkan dalam bentuk kurva *ROC*

Evaluate Results

Pada tahapan ini dilakukan evaluasi terhadap keluaran yang dihasilkan oleh proses *text mining*. Hasil tersebut dibandingkan dengan tujuan yang terdapat pada tahap *business understanding*, kemudian diketahui sejauh mana hasil dari proses *text mining* ini memenuhi tujuan yang ingin dicapai. Hasil pengujian yang dilakukan melalui model SVM menghasilkan *Confusion Matrix*, yaitu *accuracy* sebesar 79,55%, *Precision* sebesar 80,37%, dan *Sensitivity* atau *Recall* sebesar 80,91% seperti terlihat pada Gambar 5 dibawah ini :

accuracy: 79.55% +/- 4.01% (micro average: 79.55%)			
	true Positif	true Negatif	class precision
pred. Positif	172	42	80.37%
pred. Negatif	48	178	78.76%
class recall	78.18%	80.91%	

Sumber: (Utama et al., 2019)

Gambar 5. Accuracy dari Model Naïve Bayes

Perhitungan manualnya adalah sebagai berikut :

1. Nilai Akurasi

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{172 + 178}{172 + 178 + 42 + 48}$$

$$= 0.795454$$

$$= 79,55\%$$

2. Class Recall Positif

$$Class\ Recall\ Positif = \frac{TP}{TP + FN}$$

$$= \frac{172}{172 + 48}$$

$$= 0.781818$$

$$= 78.18\%$$

3. Class Recall Negatif

$$Class\ Recall\ Negatif = \frac{TN}{TN + FP}$$

$$= \frac{178}{178 + 42}$$

$$= 0.809090$$

$$= 80.91\%$$

4. Class Precision Positif

$$Class\ Precision\ Positif = \frac{TP}{TP + FP}$$

$$= \frac{172}{172 + 42}$$

$$= 0.803738$$

$$= 80.37\%$$

5. Class Precision Negatif

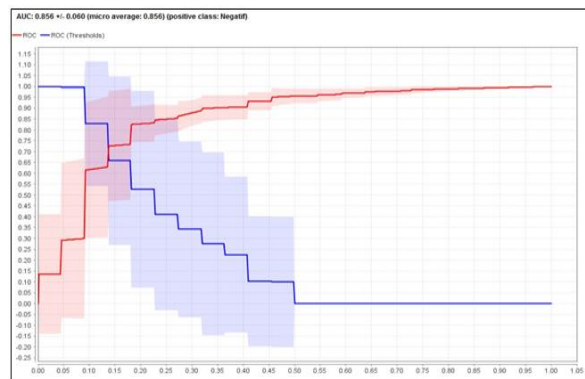
$$Class\ Precision\ Negatif = \frac{TN}{TN + FN}$$

$$= \frac{178}{178 + 48}$$

$$= 0.787610$$

$$= 78.76\%$$

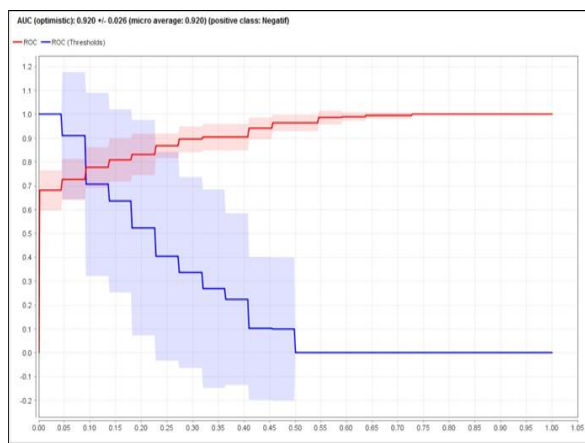
Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC* serta akan ditampilkan dalam bentuk kurva *ROC*. Berikut ini adalah alur penelitian yang akan dilakukan *Area Under Curve* dari model Algoritma *Naïve Bayes* dapat dilihat pada Gambar 6 dibawah ini :



Sumber: (Utama et al., 2019)

Gambar 6. AUC dari model Naïve Bayes

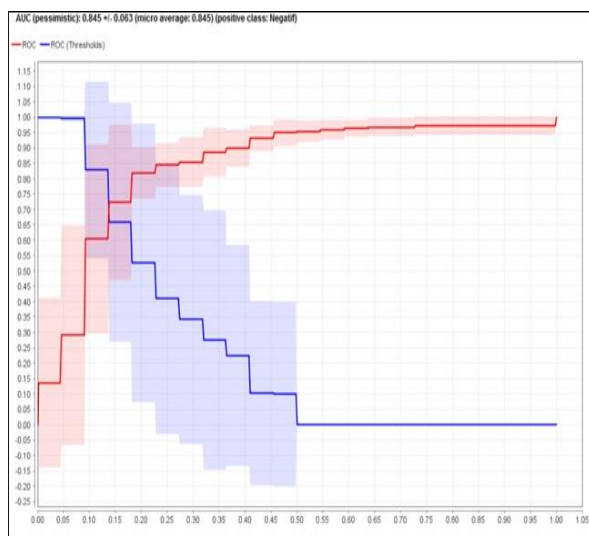
curve optimis juga terdapat *Naïve Bayes* yang hasilnya disajikan pada Gambar 7 berikut :



Sumber: (Utama et al., 2019)

Gambar 7. AUC Optimis dari model Naïve Bayes

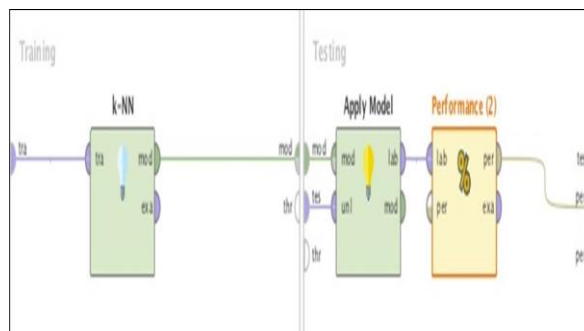
Sedangkan Area Under Curve pesimistis dari algoritma *Naïve Bayes* dapat dilihat pada Gambar 8 dibawah ini :



Sumber: (Utama et al., 2019)
Gambar 8. AUC Pesimistis dari model Naive Bayes

Review Process

Setelah dilakukan pemeriksaan terhadap keseluruhan proses, dimana pemeriksaan dilakukan dengan kembali ke tahapan awal. Dipastikan bahwa tidak ada faktor atau parameter penting yang terlewatkan dari proses *text mining*. Selain itu peneliti juga membandingkan hasil penelitian *Naive Bayes* ini dengan metode *K-Nearest Neighbor* (K-NN). Hasilnya ternyata *Naive Bayes* lebih baik dibandingkan dengan K-NN. Tingkat akurasi *Naive Bayes* lebih tinggi dibandingkan dengan K-NN. Algoritma *k-Nearest Neighbor* (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN (Krisandi, Prihandono, & Helmi, 2013). Sedangkan pendapat yang lain menyebutkan bahwa K-NN merupakan metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan (Hermaduanty & Kusumadewi, 2008), dan ada juga pendapat yang menyebutkan bahwa metode K-Nearest Neighbor (K-NN), yaitu metode yang memperhitungkan kemiripan jumlah kemunculan kata antara satu dokumen dengan dokumen lain (Darujati, 2010). Pada penelitian ini tingkat *Accuracy Naive Bayes* adalah sebesar 79.55% +/- 4.01% (micro average: 79.55%) sedangkan akurasi K-NN 57.05% +/- 6.54% (micro average: 57.05%). Berikut adalah gambar praprocessing dengan menggunakan K-NN.



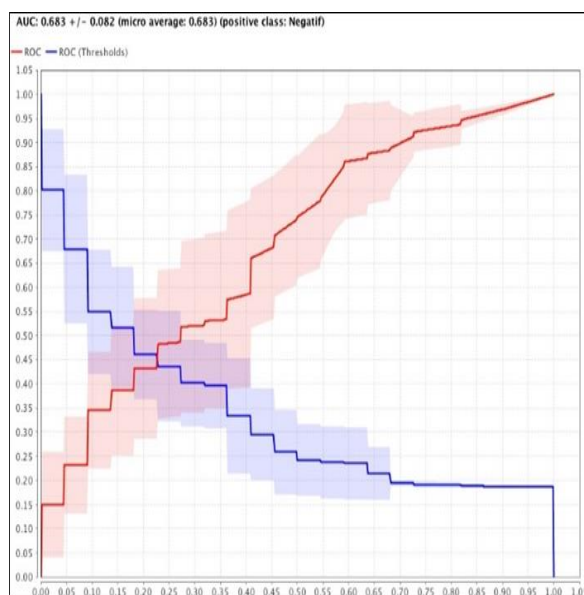
Sumber: (Utama et al., 2019)
Gambar 9. Tahap praprocessing dengan menggunakan metode K-NN
Tingkat akurasi digambarkan pada Gambar 10 sebagai berikut :

accuracy: 57.05% +/- 6.54% (micro average: 57.05%)

	true Positif	true Negatif	class precision
pred. Positif	183	152	54.63%
pred. Negatif	37	68	64.76%
class recall	83.18%	30.91%	

Sumber: (Utama et al., 2019)
Gambar 10. Tingkat akurasi menggunakan K-NN

Berikut ini adalah alur penelitian yang akan dilakukan *Area Under Curve* dari model Algoritma K-NN :



Sumber: (Utama et al., 2019)
Gambar 11. AUC dari model K-NN

KESIMPULAN

Setelah dilakukan penelitian maka dapat disimpulkan bahwa tingkat Hasil pengujian yang dilakukan melalui model *Naïve Bayes* menghasilkan *Confusion Matrix*, yaitu *accuracy* sebesar **79,55%**, *Precision* sebesar **80,37%**, dan *Sensitivity* atau *Recall* sebesar **80,91%**. Penelitian selanjutnya yang akan dilakukan adalah membandingkan metode algoritma antara *Naïve Bayes* dengan menambahkan seleksi fitur *Information Gain* Untuk mencari hasil akurasi mana algoritma terbaik diantara keduanya.

REFERENSI

- Budi, S. (2017). Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means. *Techno.Com*. <https://doi.org/10.33633/tc.v16i1.1263>
- Darujati, C. (2010). PERBANDINGAN KLASIFIKASI DOKUMEN TEKS MENGGUNAKAN METODE NAÏVE BAYES DENGAN K-NEAREST NEIGHBOR Abstrak. *Universitas Stuttgart*.
- Faradhillah, N. Y. A., Kusumawardani, R. P., & Hafidz, I. (2016). Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin. *Prosiding Seminar Nasional Sistem Informasi Indonesia 2016*.
- Hermaduanty, N., & Kusumadewi, S. (2008). Sistem Pendukung Keputusan Berbasis Sms Untuk Menentukan Status Gizi Dengan Metode K-Nearest Neighbor. *Seminar Nasional Aplikasi Teknologi Informasi ISSN 1907-5022*.
- Hidayat, andi nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik)*.
- Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Evolusi Volume 4 Nomor 2 - 2016*.
- Ipmawati, J., Kusrini, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. In *Indonesian Journal on Networking and Security* (Vol. 6).
- Kom, S., Dewi, E., Mulyani, S., Kom, S., & Nurhasanah, I. R. (2015). Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes. *Konferensi Nasional Sistem Dan Informatika (KNS&I)*. <https://doi.org/10.1007/s12237-007-9014-7>
- Krisandi, N., Prihandono, B., & Helmi. (2013). Algoritma K - Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. MINAMAS Kecamatan Parindu. *Buletin Ilmiah Math.Stat.Dan Terapannya(Bimaster)*.
- Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier. *Prosiding SNST*.
- Putri, W. S. R., Nurwati, N., & S., M. B. (2016). PENGARUH MEDIA SOSIAL TERHADAP PERILAKU REMAJA. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*. <https://doi.org/10.24198/jppm.v3i1.13625>
- Ruhyana, N. (2019). Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode Naive Bayes. 3(1), 94–99.
- Suci, D., Yanti, A., & Adikara, P. P. (2019). Analisis Sentimen Tentang Kebijakan Ganjil Genap Kendaraan Bermotor di DKI Jakarta Pada Twitter Menggunakan BM25 dan K-Nearest Neighbor. 3(3), 2626–2631.
- Utama, H. S., Rosiyadi, D., Prakoso, B. S., & Ariadarma, D. (2019). JURNAL RESTI Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan. 1(10), 2–8.
- Wahyudi, E. N. (2009). Algoritma Sederhana dalam Memahami Proses Pengurutan Data. *Fakultas Teknologi Informasi, Universitas Stikubank Semarang*.
- Wahyuni, D. T., Sutojo, T., & Luthfiarta, A. (2004). Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naive Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi. *UDINUS*.
- Wasiati, H., & Wijayanti, D. (2014). Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes. *Jurnal Ilmiah NERO*.

Yuda Septian, N. (2009). Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian

Nuswantoro. *Jurnal Semantik* 2013.