

# ANALISIS SENTIMEN APLIKASI TRANSPORTASI ONLINE KRL ACCESS MENGGUNAKAN METODE NAIVE BAYES

Siti Nurwahyuni  
Universitas Bina Sarana Informatika  
Email: [sitinunu29@gmail.com](mailto:sitinunu29@gmail.com)

## ABSTRAK

Angkutan umum dianggap sebagai solusi untuk persoalan yang terkait dengan mobilitas di daerah perkotaan. Terkait dengan transportasi, jabodetabek memiliki masalah seperti kota-kota besar lainnya di dunia, kemacetan lalu lintas. Kemacetan lalu lintas terjadi di jam sibuk di pagi dan sore hari, untuk itu dibutuhkan aplikasi yang dapat digunakan untuk mengakses alat transportasi dengan mudah dan cepat, salah satunya adalah aplikasi transportasi secara online yang saat ini banyak digunakan oleh pengguna smartphone mobile. Analisis Sentimen dapat dianggap sebagai proses klasifikasi yang memiliki 3 tingkat klasifikasi utama yaitu tingkat pada dokumen, tingkat pada kalimat dan tingkat pada aspek. Pada tingkat dokumen analisis sentimen bertujuan untuk mengklasifikasikan dokumen opini sebagai pernyataan pendapat atau sentimen positif dan negatif. Metode yang digunakan dalam penelitian kali ini adalah yang banyak digunakan dalam mengolah data dalam bentuk text mining dengan tingkat akurasi yang cukup tinggi yaitu algoritma Naive Bayes Classifier (NBC). Dengan algoritma Naive Bayes Classifier (NBC) didapatkan hasil sebesar 84.00%.

**Kata Kunci:** Analisis Sentimen, Aplikasi Transportasi, Naive Bayes classifier algorithm (NBC).

## ABSTRACT

*Public transport is considered as a solution to problems related to urban mobility. Associated with the transport, jabodetabek have problems like other big cities in the world, traffic congestion. Traffic congestion occurs at peak hours in the morning and evening, for it needed an application that can be used to access transport easily and quickly one of which is transport applications online that is currently widely used by mobile smartphone. Sentiment analysis can be considered as a classification process that has three major classification level is the level documents, at the sentence level and at the level of aspects. At the level of sentiment analysis document aims to classify a document as a statement of opinion or the opinion of the positive and negative sentiment. In addition the use of method used in the present study was methods that are widely used in data processing in the form of text mining with a fairly high degree of accuracy that the algorithm Naive Bayes Classifier (NBC). while The Naive Bayes classifier algorithm ( NBC ) is obtained by 84.00%.*

**Key Words:** *sentiment analysis,Transportation applications, Naive Bayes classifier algorithm (NBC).*

## 1. PENDAHULUAN

Angkutam umum dianggap sebagai solusi untuk persoalan yang terkait dengan mobilitas di daerah perkotaan, sistem transportasi umum yang ramah pengguna (*user-friendly*) harus mempertimbangkan aksesibilitas ke stasiun transportasi umum, mobilitas sistem transportasi publik, dan konektivitas tanpa batas dengan berbagai sistem transportasi lainnya (misalnya, kereta api, metro, lokal angkutan kereta ringan, dan bus) (Mishra et al., 2012). Terkait dengan transportasi, jabodetabek memiliki masalah seperti kota-kota besar lainnya di dunia, kemacetan lalu lintas. Kemacetan lalu lintas terjadi di jam sibuk di pagi dan sore hari. kemacetan yang ditimbulkan menyebabkan Rp 5,5 miliar dari keuangan hilang dalam bentuk biaya kendaraan operasional dan waktu perjalanan (JICA Study Team, 2014).

Kereta api merupakan salah satu sarana transportasi tertua di dunia (Wikipedia, 2008), memiliki berbagai keunggulan komparatif dan kompetitif, hemat lahan dan energi, rendah polusi, bersifat masal, dan juga adaptif dengan perubahan teknologi ([www.kereta-api.com](http://www.kereta-api.com)). Kereta api di Indonesia, merupakan sarana transportasi umum yang cukup tua, didirikan tahun 1864 oleh pemerintah Belanda, NV Nederlandsch Indische Spoorweg Mastshappij (NIS), yang merupakan cikal bakal dari perkembangan industri perkeretaapian di Indonesia saat ini.

Industri perkeretaapian saat ini dikelola sepenuhnya oleh negara sebagai sebuah perusahaan layanan publik dengan nama PT. Kereta Api Indonesia yang disingkat dengan PT. KAI (Undang-Undang Reputasi MANAJEMEN PEMASARAN, VOL. 4, NO. 1, April 2009: 23-37 24 blik Indonesia Nomor 13 Tahun 1992 Tentang Perkeretaapian Bab 3 Pasal 4).

Android adalah sistem operasi berbasis Linux yang dirancang untuk perangkat mobile seperti smartphone dan tablet. Aplikasi Android ditulis pada bahasa pemrograman Java, dan *Software Development Kit Android (SDK)* menyediakan perpustakaan API dan alat pengembang yang diperlukan untuk membangun, melakukan tes, dan mengecek kesalahan pada aplikasi. Java adalah bahasa utama di Android. Selain pengembangan aplikasi Java, Android menyediakan paket alat pengembangan dan perpustakaan untuk mengembangkan develop

Native applications, the Native Development Kit (NDK). NDK memungkinkan untuk melaksanakan bagian dari aplikasi Java menggunakan bahasa asli-kode seperti C dan C ++. Menggunakan kode asli tidak mengakibatkan peningkatan kinerja secara otomatis. Kode asli berguna untuk tempat menyimpan kode asli yang ada untuk Android, bukan untuk mempercepat bagian dari sebuah aplikasi Android. Untuk memanfaatkan kemampuan komputasi tinggi pada perangkat Android saat memberikan Renderscript (H. Guihot, 2012).

Google Play adalah layanan konten digital milik Google yang melingkupi toko daring untuk produk-produk seperti musik/lagu, buku, aplikasi, permainan, ataupun pemutar media berbasis awan. Layanan ini dapat diakses baik melalui web, aplikasi android (Play Store), dan Google TV. Konten yang telah dibeli tersedia pada seluruh platform-platform di atas Google Play mulai dikenalkan pada bulan Maret 2012 sebagai pengganti dari Android Market dan Layanan Musik Google (Wikipedia, 2015).

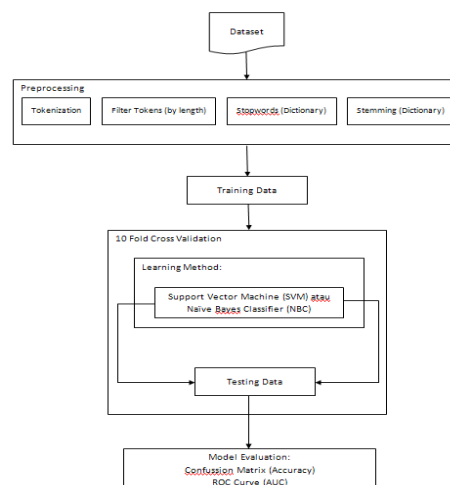
Salah satu metode klasifikasi yang dapat digunakan adalah metode Naïve Bayes yang sering disebut dengan *Naïve Bayes Classifier* (NBC). Kelebihan NBC adalah sederhana tetapi memiliki akurasi yang tinggi. Berdasarkan hasil eksperimen, NBC terbukti dapat digunakan secara efektif untuk mengklasifikasikan berita secara otomatis dengan akurasi mencapai 90.23%. Algoritma NBC yang sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi (Yudi Wibisono, 2008). Sesuai dengan penelitian yang sebelumnya telah dilakukan oleh Amir Hamzah dalam (Klasifikasi Teks dengan Naives Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Academis, 2012). Bahwa penggunaan Algoritma NBC memiliki kinerja yang cukup tinggi untuk klasifikasi dokumen teks, baik dokumen berita maupun dokumen akademik. Pada klasifikasi dokumen berita didapatkan akurasi yang lebih tinggi (maksimal 91%) dibandingkan dengan dokumen akademik (maksimal 82%). Baik pada dokumen berita maupun dokumen akademik, penggunaan 50% dokumen sebagai dokumen pelatihan memberikan kinerja akurasi diatas 75%.

Yudi Wibisono dalam Klasifikasi Berita Berbahasa Indonesia menggunakan *Naïve Bayes Classifier*, 2006 telah melakukan klasifikasi pada 582 dokumen berbahasa Indonesia menggunakan metode NBC dan memperoleh hasil eksperimen nilai akurasi NBC tinggi, terutama jika dokumen contoh yang digunakan besar ( $\geq 400$  dokumen). Hal yang menarik adalah akurasi tidak menunjukkan peningkatan yang signifikan walaupun dokumen contoh telah meningkat banyak dari 70% menjadi 90% serta akurasi masih relative tinggi walaupun dokumen contoh secara ekstrim dikurangi hanya 58 dokumen (10%).

Fatimah Wulandini dan Anto Satriyo Nugroho dalam (*Text Classification Using Support Vector Machine for Webmining Based Spation Temporal Analysis of the Spread of Tropical Diseases*, 2009) mendapatkan hasil bahwa metode SVM menunjukkan hasil paling baik pada kategorisasi teks berbahasa Indonesia. Eksperimen dilakukan pada 3713 *features* dan 360 *instances*. Data tersebut dibagi menjadi 120 *instances* sebagai data uji dan 240 *instances* sebagai data latih. hasil dari algoritma tersebut menunjukkan performansi yang tidak berbeda jauh antara metode SVM dan NBC walaupun metode NBC adalah metode yang lebih konvensional dan lebih sederhana.

Fabrice Colas dan Pavel Brazdil dalam penelitiannya berjudul *Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks* mendapatkan bahwa metode NBC memiliki performansi yang lebih baik dibandingkan KNN dan SVM untuk menyelesaikan *binary classification* pada dokumen berbahasa Inggris. Hasil penelitiannya juga menyebutkan waktu komputasi yang jauh lebih pendek oleh metode NBC dan KNN. Waktu komputasi SVM berkembang secara kuadrat seiring dengan perkembangan jumlah data latih.

## 2. METODE PENELITIAN



Gambar 1. Kerangka Pemikiran Penelitian

Tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan Data  
Data diambil dari sebuah website pengunduh aplikasi untuk smartphone android, kemudian diseleksi agar sesuai dengan kebutuhan penelitian.
2. Pengolahan Awal Data  
Pengolahan awal data bertujuan untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga dapat berupa konversi nilai-nilai redundan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model.
3. Metode Yang Diusulkan  
Metode yang digunakan dalam penelitian ini adalah 2 metode algoritma yang umum digunakan dalam mengolah data dengan metode text mining dalam analisis sentimen yaitu algoritma algoritma *Naive Bayes Classifier* (NBC).
4. Eksperimen dan Pengujian Metode  
Dalam eksperimen dan metode yang dilakukan penulis membandingkan 2 metode algoritma tersebut untuk mengolah data sehingga menghasilkan nilai akurasi yang tinggi. eksperimen dilakukan secara terpisah agar didapatkan hasil secara maksimal dan dapat ditentukan metode algoritma mana yang dapat menghasilkan akurasi yang lebih tinggi.

5. Rancangan Klasifikasi Teks dengan *Naive Bayes Classifier* (NBC) dari aturan bayes

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$

maka dapat dibuat sebuah asumsi tentang bagaimana kita menghitung probabilitas dari kemunculan dokumen yang setara dengan perkalian (*product*) dari probabilitas kemunculan tiap kata di dalamnya. Hal ini menyebabkan tidak ada hubungan antara kata yang satu dengan kata yang lainnya. Asumsi independen ini tidak sepenuhnya benar, banyak kata-kata lebih sering muncul bersamaan daripada muncul secara individual, tapi hal ini menyederhanakan proses klasifikasi.

6. Rancangan Klasifikasi Teks dengan *Support Vector Machine* (SVM)

Proses klasifikasi menggunakan SVM dimulai dengan mengubah text menjadi data vector. Vector dalam penelitian ini memiliki dua komponen yaitu dimensi (*word id*) dan bobot. Bobot ini sering dikombinasikan ke dalam sebuah nilai *tf-idf*, secara sederhana dengan mengalikan mereka bersama-sama. Ada banyak variasi pada gagasan dasar *tf-idf*,

7. Evaluasi dan Validasi Hasil

Evaluasi berfungsi untuk mengetahui akurasi dari kedua model algoritma yang diusulkan. Akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan akan ditampilkan dalam bentuk *Curve ROC* (*Receiver Operating Characteristic*). Teknik validasi yang digunakan adalah *Cross Validation*.

### 3. HASIL PENELITIAN

#### A. Pengumpulan Data

Review aplikasi transportasi darat dikumpulkan melalui situs download aplikasi smartphone android:

<https://play.google.com/store/apps/>

Tabel 1. Contoh Review Data Negatif

Nama Dokumen	Isi Dokumen
negatif1.txt	Aplikasi sudah kadaluwarsa, jadwal waktu yang tidak realtime,
negatif2.txt	akhir akhir ini suka gak bisa kebuka,

	terus lemot ujung ujungnya masalah di server
negatif3.txt	Error dan error Yg suka latih kesabaran boleh di download
negatif4.txt	error akun sering expired .reset password gabisa terus
negatif5.txt	Loading utk penarikan datany lama bngt
negatif6.txt	Idiiiihhh lemot banget sekarang gojek susah banget mau order ,
negatif7.txt	Aplikask lemot bikin HP lemot, kena korban 1 HP rusak setelah download aplikasi ini.
negatif8.txt	Lambat Respon nya lama banget klo top up
negatif9.txt	Kecewa bgt saya,
negatif10.txt	Aplikasi tdk bisa pilih stasiun, Coba re install, malah ga bisa open aplikasi
negatif11.txt	Pdhl ngandelin ni aplikasi tapi kesel stiap di refresh atau d buka suka gagal server tolong di perbaiki trima kasih
negatif12.txt	Tolong dong di perbaiki aplilasi suka lemot kalau buka suka ada tulisan penarikan data bisa setengah jam

Tabel 2. Contoh Review Data Positif

Nama Dokumen	Isi Dokumen
positif1.txt	Sangat membantu saya di saat gk di jemput suami heheee
positif2.txt	Sangat membantu dalam memperkirakan jadwal kereta api akan tiba di stasiun tujuan akhir.

positif3.txt	Good Bagus ah... sedikit lemot tp gak apa2
positif4.txt	aplikasi ini sgt membantu
positif5.txt	Info ini sangat membantu saya. Saya dari rawabuntu station-tanah abang.dari tanah abang -Jatinegara
positif6.txt	Bagus bgt cukup membantu
positif7.txt	Wow Membantu banget.. buat gw yg suka ngubek2 kota..
positif8.txt	Bagus. sangat membantu sekali untuk pengetahui posisi kereta tujuan jadi bisa memprediksikan waktu berangkat dr rumah.
positif9.txt	Cukup membantu tp lbh bagus lg tmbh vending mesin d setiap stasiun spy tdk panjang antrian
positif10.txt	Aplikasi ini sangat membantu Saya yg stiap hari berpergian mnggunakan KRL jadi tau jadwal dan keterlambatan nya
positif11.txt	Cukup bagus dan membantu,sayangnya aplikasi lama yg terpasang gk bisa diupdate.
positif12.txt	Good App Membantu sekali. Thanks

**Tabel 3. Perbandingan teks sebelum dan sesudah dilakukan proses *tokenization***

Teks data <i>Review</i> sebelum dilakukan proses <i>tokenization</i>	Tidak memuaskan!!! Untuk log on susah, mengaktifkan lupa kata sandi juga tidak bisa. Gimama nih?!?!??
Teks data <i>Review</i> setelah dilakukan proses <i>tokenization</i>	Tidak memuaskan Untuk log on susah mengaktifkan lupa kata sandi juga tidak bisa Gimama nih

**Tabel 4. Perbandingan teks sebelum dan sesudah dilakukan proses *filter tokens (by lenght)***

Teks data <i>Review</i> sebelum dilakukan proses <i>filter tokens (by lenght)</i>	Aplikasi tdk bisa pilih stasiun, Coba re install, malah ga bisa open aplikasi
Teks data <i>Review</i> setelah dilakukan proses <i>filter tokens (by lenght)</i>	Aplikasi bisa pilih stasiun Coba re install malah ga bisa open aplikasi

**Tabel 5. Perbandingan teks sebelum dan sesudah dilakukan proses *Stopwords***

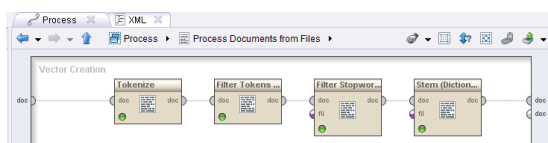
Teks data <i>Review</i> sebelum dilakukan proses <i>stopwords</i>	Tidak memuaskan!!! Untuk log on susah, mengaktifkan lupa kata sandi juga tidak bisa. Gimama nih?!?!??
Teks data <i>Review</i> setelah dilakukan proses <i>stopwords</i>	Tidak memuaskan Untuk susah mengaktifkan lupa kata sandi juga tidak bisa Gimana

**Tabel 6. Perbandingan teks sebelum dan sesudah dilakukan proses *Stemming***

Teks data <i>Review</i> sebelum dilakukan proses <i>stemming</i>	Payah Diupdate malah jd keliatan banget bug nya.. masa tempat duduknya gk kedetect? Payah!!!! Tolong diperbaiki secepatnya..!!!
Teks data <i>Review</i> setelah dilakukan proses <i>stemming</i>	payah diupdate malah keliatan banget masa tempat duduknya kedetect payah tolong diperbaiki secepatnya

**B. Pengolahan Data Awal (*Preprocessing*)**

Dalam *preprocessing* ada beberapa tahapan yang dilakukan, diantaranya adalah *tokenize*, *filter tokens (by lenght)*, *Stopwords* dan *stemming*.



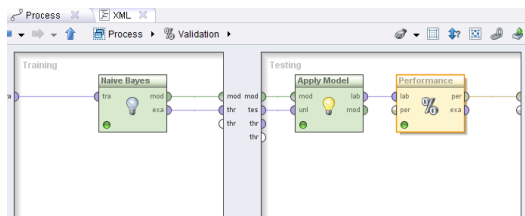
**Gambar 2. Desain Tahap Preprocessing pada Rapid Miner**

**C. Klasifikasi Teks Menggunakan Naive Bayes Classifier (NBC)**

Setelah melalui tahap preprocessing pada rapid miner, kemudian digunakan algoritma NBC untuk mengetahui tingkat akurasi dari data yang diolah. Proses klasifikasi dokumen pada penelitian ini adalah untuk menentukan sebuah kalimat sebagai anggota kelas positif atau kelas negatif berdasarkan nilai perhitungan pembobotan.

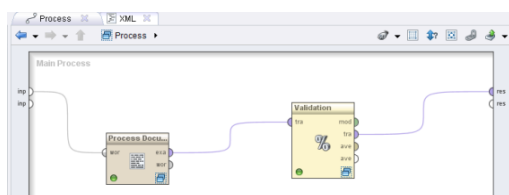
**Tabel 7. Tabel Dokumen yang dihitung nilai bobotnya**

Dokumen	Atribut						Kelas
	Bagus	Membantu	Aplikasi	Gojek	Error	Kecewa	
positif34.txt	0,182	0,138	0,164	0,182	0	0	Positif
negatif7.txt	0	0	0,200	0,233	0,203	0,354	Negatif



**Gambar 3. Desain Validasi Model NBC dengan menggunakan Rapid Miner**

Proses validasi dengan menggunakan 10-fold cross validation. Nilai k yang diambil adalah 10 karena beberapa penelitian sebelumnya menunjukkan bahwa nilai tersebut menunjukkan hasil terbaik untuk proses validasi.



**Gambar 4. Desain Model NBC dengan menggunakan Rapid Miner**

Proses terakhir adalah penerapan algoritma Naive Bayes Classifier sebagai metode yang akan digunakan pada kasus analisis sentiment review aplikasi transportasi darat. Pada penelitian ini data yang digunakan berbentuk teks proses klasifikasi tidak dapat diasumsikan secara linear, oleh karena itu akan digunakan kernel sebagai parameter yang dapat melakukan klasifikasi terhadap

data yang tidak bersifat internal. Apply model digunakan untuk dapat membaca data yang akan diestimasi berdasarkan data yang sudah dipelajari sebelumnya.

**D. Hasil Klasifikasi Teks antara algoritma Naive Bayes Classifier (NBC)**

**1. Evaluasi Metode Naive Bayes Classifier**

Data Review Positif disatukan dalam satu folder dengan nama positif dan data Review Negatif disatukan dalam satu folder dengan nama negatif. setelah proses klasifikasi, kemudian melalui tahapan preprocessing. Hasil yang didapatkan melalui pengolahan menggunakan Rapidminer sebagai berikut:

**Tabel 9. Contoh hasil proses vector creation menggunakan Rapidminer**

No.	Kata	Statistic	Range
1	abis	avg = 0.003 +/- 0.033	[0.000 ; 0.460]
2	access	avg = 0.005 +/- 0.040	[0.000 ; 0.382]
3	adanya	avg = 0.004 +/- 0.037	[0.000 ; 0.410]
4	admin	avg = 0.001 +/- 0.021	[0.000 ; 0.360]
5	agak	avg = 0.003 +/- 0.036	[0.000 ; 0.548]
6	agar	avg = 0.001 +/- 0.019	[0.000 ; 0.321]
7	ahhhhh	avg = 0.001 +/- 0.022	[0.000 ; 0.382]

**Tabel 10. Confusion Matrix Model Algoritma Naive Bayes Classifier (NBC) dengan Rapid Miner**

accuracy: 84.00% +/- 7.12% (mikro: 84.00%)

	true negatif	true positif	class precision
pred. negatif	133	31	81.10%
pred. positif	17	119	87.50%
class recall	88.67%	79.33%	

precision: 88.70% +/- 9.26% (mikro: 87.50%) (positive class: positif)

	true negatif	true positif	class precision
pred. negatif	133	31	81.10%
pred. positif	17	119	87.50%



class recall	88.67%	79.33%	
--------------	--------	--------	--

recall: 79.33% +/- 12.09% (mikro: 79.33%) (positive class: positif)

	true negatif	true positif	class precision
pred. negatif	133	31	81.10%
pred. positif	17	119	87.50%
class recall	88.67%	79.33%	



**Gambar 7. Kurva ROC Model Naive Bayes Classifier (NBC)**

Kurva ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah pada sumbu X dan positif benar pada sumbu Y. Hasil perhitungan pada kurva ROC, menggambarkan kurva ROC untuk algoritma *Naive Bayes Classification* (NBC). Dapat disimpulkan bahwa satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Kurva ROC *Naive Bayes Classification* (NBC) memiliki **AUC** sebesar **0.976** yang termasuk ke dalam kategori *excellent classification*.

```

PerformanceVector
PerformanceVector:
accuracy: 84.00% +/- 7.12% (mikro: 84.00%)
ConfusionMatrix:
True:  negatif positif
negatif:  133   31
positif:   17  119
precision: 88.70% +/- 9.26% (mikro: 87.50%) (positive class: positif)
ConfusionMatrix:
True:  negatif positif
negatif:  133   31
positif:   17  119
recall: 79.33% +/- 12.09% (mikro: 79.33%) (positive class: positif)
ConfusionMatrix:
True:  negatif positif
negatif:  133   31
positif:   17  119
AUC (optimistic): 0.976 +/- 0.027 (mikro: 0.976) (positive class: positif)
AUC: 0.581 +/- 0.126 (mikro: 0.581) (positive class: positif)
AUC (pessimistic): 0.707 +/- 0.120 (mikro: 0.707) (positive class: positif)
    
```

**Gambar 8. Performance Vector Model Naive Bayes Classification (NBC)**

Akurasi yang diperoleh yaitu 84.00% dari 150 data review negatif dan 150 data review positif. Data review yang dihasilkan RapidMiner dengan model NBC menunjukkan bahwa klasifikasi untuk review positif yang sesuai prediksi yaitu **133**. Data review negatif yang termasuk ke dalam prediksi positif yaitu **17**. Data review positif yang termasuk ke dalam prediksi negatif yaitu **31** dan data review negatif yang termasuk ke dalam prediksi negatif yaitu **119**.

Acc (Accuracy)

$$\frac{TP+TN}{TP+TN+FP+FN} = \frac{133 + 119}{133+17+31+119} = \frac{275}{300} = 0.84$$

Sensitivity

$$\frac{TP}{TP + FP} = \frac{133}{133+31} = \frac{133}{164} = 0.811$$

Specificity

$$\frac{TN}{TN + FN} = \frac{119}{17 + 119} = \frac{119}{136} = 0.875$$

PPV

$$\frac{TP}{TP + FN} = \frac{133}{133 + 17} = \frac{133}{150} = 0.887$$

NPV

$$\frac{TN}{FP + FN} = \frac{119}{17 + 119} = \frac{119}{136} = 0.875$$

#### 4. KESIMPULAN

- Kesimpulan dari penelitian ini adalah:
1. Metode NBC memberikan hasil dengan akurasi hingga 84.00% untuk data uji review opini positif berbahasa Indonesia pada pemilihan aplikasi transportasi darat pada smartphone.
  2. Metode NBC memberikan hasil yang tepat dalam mengklasifikasikan opini dalam bentuk paragraf yang terdiri dari beberapa kalimat.
  3. Diketahui bahwa dokumen teks memiliki struktur yang tidak teratur. Maka tahap *preprocessing* sangat dibutuhkan untuk pemrosesan data awal. Karena terbukti dapat mengurangi dimensi dokumen dan membantu membuang kata yang tidak perlu. Sehingga kinerja tahap seleksi fitur dan klasifikasi menjadi lebih mudah dengan hasil yang lebih akurat.
  4. Pada algoritma NBC dapat ditambahkan metode tambahan agar akurasi dapat menyamai algoritma SVM seperti *sample boosting* dengan metode tambahan tersebut akurasi dari metode NBC dapat meningkat hingga hampir 10% sehingga

dapat menyamai akurasi dari algoritma SVM.

*applications: A survey. Ain Shams Engineering Journal*  
doi:10.1016/j.asej.2014.04.011

## REFERENSI

A.D. Reid, K. Flautner, E. Grimley-Evans, Y. Lin, SoC-C: efficient programming abstractions for heterogeneous multicore systems on chip, in: E.R. Altman (Ed.), (2008). *Proceedings of the 2008 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems, CASES'08, ACM, Atlanta, GA, USA*, pp. 95–104.

Mishra, S., Welch, T.F., Jha, M.K., (2012). *Performance indicators for public transit connectivity in multi-modal transportation networks. Transport. Res. Part A: Policy Practice* 46 (7), 1066–1085.

Tan, P. N., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. Boston : Pearson Addison Wesley.

Turban, E.; et.al. (2011). *Decision Support and Business Intelligence Systems* (9 ed.). New Jersey: Pearson Education, Inc.

Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). *Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. Expert Systems with Applications*, 41(8), 3955–3964.  
doi:10.1016/j.eswa.2013.12.035

Wibisono, Y. 2005. *Klasifikasi Berita Berbahasa Indonesia menggunakan Naïve Bayes Classifier*. [Online]. Tersedia di: [http://fpmipa.upi.edu/staff/yudi/yudi\\_0805.pdf](http://fpmipa.upi.edu/staff/yudi/yudi_0805.pdf) [diunduh: 1 Nov 2010].

Hamzah, A. (2012). *Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstrak Akademis. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, Yogyakarta*, 269-277.

H. Guihot, Renderscript, (2012). in: *Pro Android Apps Performance Optimization*, Springer, pp. 231–263.

<http://www.suaramerdeka.com/harian/0403/29/kot2.htm>

<http://www.suaramerdeka.com/harian/0403/29/kot2.htm>

JICA Study Team. (2004). *The Study on Integrated Transportasi Master Plan for Jabodetabek (Phase II), Find Report, Main Report, I: Masterplan Study, in Coordination with National Development Planning Agency (BAPPENAS) Republic Indonesia*

Manning, C. D., Ragnavan, P., & Schutze, H. (2008). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and*