

**ANALISIS SENTIMEN TRANSPORTASI *ONLINE*
MENGUNAKAN ALGORITMA *NAIVE BAYES*,
SUPPORT VECTOR MACHINE DAN *KNN* DENGAN
*PARTICLE SWARM OPTIMIZATION***



TESIS

BENI RAHMATULLAH
14002117

PROGRAM STUDI ILMU KOMPUTER (S2)
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2019

PERSETUJUAN DAN PENGESAHAN TESIS

Tesis ini diajukan oleh :

Nama : Beni Rahmatullah
NIM : 14002117
Program Studi : Ilmu Komputer
Jenjang : Program Strata Dua (S2)
Konsentrasi : *Data Mining*
Judul Tesis : Analisis Sentimen Transportasi Online Menggunakan
Algoritma Naive Bayes, Support Vector Machine dan
KNN Dengan Particle Swarm Optimization

Telah dipertahankan pada periode 2019-2 dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Magister Ilmu Komputer (M.Kom) pada Program Strata Dua (S2) Program Studi Ilmu Komputer di Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 22 Januari 2020

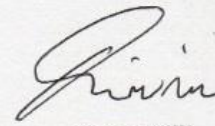
PEMBIMBING TESIS

Dosen Pembimbing : Dr. Didi Rosiyadi, M.Kom



DEWAN PENGUJI

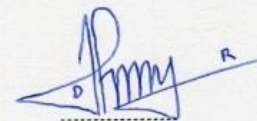
Penguji I : Dr. Dwiza Riana, S.Si, MM, M.Kom



Penguji II : Dr. Hilman Ferdinandus Pardede, ST, MEICT



Penguji III/
Dosen Pembimbing : Dr. Didi Rosiyadi, M.Kom



ABSTRAK

Nama : Beni Rahmatullah
NIM : 14002117
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining
Judul Tesis : “Analisis Sentimen Transportasi *Online* Menggunakan Algoritma *Naive Bayes*, *Support Vector Machine* dan *KNN* Dengan *Particle Swarm Optimization*”.

Transportasi merupakan kebutuhan yang paling digunakan dalam aktifitas sehari-hari dalam bekerja, berbelanja dan kegiatan yang lainnya. Transportasi online menjadi pilihan masyarakat. Aplikasi seperti Gojek dan Grab saat ini yang paling banyak diminati. Saran dan kritik dari pelanggan guna memperbaiki sistem dan pelayanan yang diberikan. Dalam hal ini penulis memperoleh data dari komentar di twitter dan menggunakan algoritma *Naive Bayes*, *Support Vector Machine* dan *K-Nearest Neighbor* sebagai perbandingan akurasi yang dihasilkan. Ketiga algoritma tersebut menggunakan salah satu fitur seleksi *Particle Swarm Optimization*. Hasil akurasi yang didapat membuktikan dengan fitur seleksi *PSO* lebih tinggi dan akurasi yang paling tinggi oleh Algoritma *Support Vector Machine (PSO)* dengan Akurasi 91.75 % dan AUC 0.957.

Kata kunci:

Analisis Sentimen, *Text Mining*, Klasifikasi, *Naive Bayes*, *Support Vector Machine*, *KNN*, *Particle Swarm Optimization*

ABSTRACT

Name : Beni Rahmatullah
NIM : 14002117
Study of Program : Ilmu Komputer
Levels : Strata Dua (S2)
Concentration : *Data Mining*
Title : “*Sentiment Analysis Online Transportation Using Naive Bayes, Support Vector Machine, and KNN Algorithm With Particle Swarm Optimization*”.

Transportation is the most used requirement in daily activities in work, shopping and other activities. Online transportation is the choice of the people. Applications like Gojek and Grab are currently the most popular. Suggestions and criticisms from customers to improve the system and services provided. In this case the authors obtain data from comments on Twitter and use the Naive Bayes algorithm, Support Vector Machine and K-Nearest Neighbor as a comparison of the resulting accuracy. The three algorithms use one of the selection features *Particle Swarm Optimization*. Accuracy results obtained prove that the PSO selection feature is higher and the highest accuracy is by the Support Vector Machine (PSO) dengan Akurasi 91.75 % dan AUC 0.957.

Keywords:

Sentiment Analysis, Text Mining, Klasifikasi, Naive Bayes, Support Vector Machine, KNN, Particle Swarm Optimization

DAFTAR ISI

	Halaman
Lembar Judul Tesis	i
Lembar Pernyataan Orisinalitas	iii
Lembar Halaman Pengesahan	iv
Kata Pengantar	v
Lembar Pernyataan Persetujuan Publikasi Karya Ilmiah	vii
Abstraksi	viii
Daftar Isi.....	x
Daftar Tabel	xii
Daftar Gambar	xiii
Daftar Lampiran	xvi

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah.....	1
1.2. Identifikasi Masalah	3
1.3. Tujuan Penelitian	4
1.4. Ruang Lingkup	4
1.5. Hipotesis	5
1.5.Sistematika Penulisan	5

BAB II LANDASAN/KERANGKA PEMIKIRAN

2.1.Tinjauan Pustaka	7
2.1.1. <i>Data Mining</i>	7
2.1.2. Pengelompokan <i>Data Mining</i>	8
2.1.3. Tahapan Proses <i>Data Mining</i>	9
2.1.4. <i>Text Mining</i>	9
2.1.5. Ekstraksi Dokumen.....	10
2.1.6. <i>Case Folding</i>	11
2.1.7. Analisis Sentimen	12
2.1.8. <i>Filtering</i>	12
2.1.9. <i>Tokenizing</i>	13
2.1.10. <i>Naïve Bayes Classifier</i>	13
2.1.11. <i>Support Vector Machine</i>	14
2.1.12. <i>K-Nearest Neighbor</i>	15
2.1.13. <i>Particle Swarm Optimization</i>	16
2.1.14. Model CRISP-DM	16
2.1.15. Evaluasi dan Model Validasi	18
2.2. Tinjauan Studi	20
2.2.1 Penelitian Terkait	20
2.3. Tinjauan Objek Penelitian	23

2.4. Kerangka Pemikiran.....	24
BAB III METODE PENELITIAN	
3.1. Pengumpulan Data	26
3.2. Pengolahan Data Awal	26
3.3. Model Yang Diusulkan	27
3.4. Eksperimen dan Hasil Pengujian	28
3.5. Evaluasi dan Validasi Hasil	28
BAB IV HASIL DAN PEMBAHASAN	
4.1. <i>Business Understanding</i>	29
4.2. <i>Data Understanding</i>	29
4.3. <i>Data Preparation</i>	29
4.3.1. <i>Tokenize</i>	30
4.3.2. <i>Filter Token (By Length)</i>	31
4.3.3. <i>Stopword Removal</i>	32
4.3.4. <i>Transform Cases</i>	34
4.4. <i>Pemodelan (Modelling)</i>	36
4.4.1. <i>Desain Model Naïve Bayes, Support Vector Machine</i> <i>dan KNN</i>	36
4.5. <i>Fase Evaluasi (Evaluation Phase)</i>	37
4.5.1. <i>Model Support Vector Machine</i>	37
4.5.2. <i>Model Naïve Bayes</i>	38
4.5.3. <i>Model KNN</i>	38
4.6. <i>Deployment</i>	49
BAB V PENUTUP	
5.1. Kesimpulan	51
5.2. Saran	51
DAFTAR REFERENSI	
DAFTAR RIWAYAT HIDUP	
LEMBAR KONSULTASI BIMBINGAN	
LAMPIRAN	

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Informasi mengenai kejadian aktual yang terjadi setiap hari, atau yang terjadi setiap menit yang saat ini bisa dengan mudah didapatkan seperti facebook, instagram dan twitter sifatnya umum memuat berbagai informasi teraktual, misal komentar tentang kesehatan, politik, ekonomi, pendidikan, hiburan, olahraga dan lain sebagainya. Hal tersebut bisa didapatkan dengan membuka berbagai media *online* yang saat ini sangat beragam jenisnya (Zain,2015). Perkembangan peradaban manusia pada dasarnya merupakan pengaruh perkembangan teknologi. Dengan kata lain, perkembangan teknologi yang mendorong kemajuan peradaban umat manusia. Tentu saja selain faktor teknologi masih ada banyak faktor yang mendorong kemajuan peradaban, misalnya kemajuan di bidang ekonomi, kedokteran, kesusasteraan dan sebagainya. Akan tetapi, dari sekian banyak faktor tersebut, yang menjadi faktor dominan adalah kemajuan teknologi.

Teknologi informasi telah mengalami perubahan secara cepat dan dinamis. Komputer merupakan salah satu bagian penting dalam peningkatan teknologi informasi. Kemampuan komputer dalam mengingat dan menyimpan informasi dapat dimanfaatkan tanpa harus bergantung kepada hambatan-hambatan seperti yang dimiliki oleh manusia. Dengan menyimpan informasi dan aturan memungkinkan komputer untuk memberikan informasi mengenai isu-isu yang sedang ramai dibicarakan, dan memberikan kesimpulan atau mengambil keputusan terhadap isu-isu tersebut.

Sosial media telah menjadi informasi yang dibutuhkan seiring dengan berkembangnya teknologi dan internet. Sosial media yang bermunculan pada dasarnya memiliki isi informasi yang penting seperti media cetak. Perbedaan yang ada antara keduanya hanya pada penggunaan medianya saja dan pemikiran yang dikarenakan adanya hasrat untuk memberikan sebuah informasi secepat mungkin tanpa menunggu harus dicetak (Wira, 2017).

Kebutuhan masyarakat akan akses yang mudah dan cepat menjadi salah satu tumbuh kembangnya industri online di Indonesia. Bermunculannya toko *online* seperti tokopedia, bukalapak, shopee dan lain-lain membuat peluang industri transportasi berbasis online tumbuh pesat di Indonesia. Dengan adanya transportasi online memudahkan masyarakat dalam pengiriman barang dan alat transportasi. Meskipun transportasi online banyak diminati sebagian masyarakat tentu tidak semuanya memberikan komentar positif sebagian masyarakat mungkin ada yang memberikan komentar negative, maka dari itu dibutuhkan suatu teknik dan metode untuk mengetahui presentase klasifikasi komentar-komentar yang diberikan masyarakat pada transportasi online.

Berdasarkan penjelasan tersebut maka diperlukan sebuah cara agar dapat mengklasifikasikan suatu isi dari komentar menjadi sebuah pengetahuan baru yaitu berupa kesimpulan negatif atau positif mengenai isi dari komentar yang ada di dalam suatu situs komentar. Hal tersebut dimungkinkan dengan menggunakan analisis sentimen. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat (Buntoro,2017).

Text mining merupakan penerapan konsep dari teknik *data mining* untuk mencari pola dalam teks, bertujuan untuk mencari informasi yang bermanfaat dengan tujuan tertentu (Anjani,2015). Keduanya memiliki tujuan yang sama yaitu untuk memperoleh informasi dan pengetahuan dari sekumpulan data yang sangat besar. Data tersebut bisa berbentuk sebuah database. Namun keduanya memiliki perbedaan jenis data. *Data mining* memiliki input data dari data yang sudah terstruktur sedangkan *text mining* dimulai dengan data yang tidak terstruktur. *Text mining* dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk *summarization*, pencarian dokumen teks dan sentimen analisis (Maulana et.al, 2016).

Data yang sudah didapat dilabeli positif dan negatif kemudian dikoreksi oleh ahli bahasa. Setelah itu dilakukan *preprocessing* baik itu mengubah kata tidak baku menjadi baku atau biasa disebut normalisasi menggunakan kamus dan

mencari akar kata yaitu *stemming* dengan bantuan aplikasi Sastrawi Jawa. Selanjutnya dilakukan juga tokenisasi dan filter terhadap kalimat, kemudian menghilangkan kata-kata yang umum digunakan dan tidak mempunyai Informasi yang berharga pada suatu konteks atau biasa disebut *stopword removal*. Algoritma klasifikasi yang digunakan yaitu *Naive Bayes Classifier*, *Support Vector Machine* dan KNN.

Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap transportasi online. Dimulai dari tahap *preprocessing* sampai tahap analisis sentimen dengan *Naive Bayes Classifier*, *Support Vector Machine* dan KNN berbasis *Particle Swarm Optimization* serta bagaimana mengukur kualitas hasil analisis menggunakan dari masing-masing algoritma klasifikasi. *Particle Swarm Optimization (PSO)* merupakan teknik optimasi yang digunakan untuk menerapkan dan memodifikasi beberapa parameter dan meningkatkan bobot atribut.

1.2. Identifikasi Masalah

Identifikasi masalah dalam penelitian ini adalah bagaimana penggunaan transportasi online pada saat ini karena banyak beredar pengemudi yang tidak sopan terhadap konsumen selain itu juga dalam penelitian ini melihat pengaruh dari seleksi fitur *Particle Swarm Optimization (PSO)* dalam menganalisis sentimen sebuah komentar pada media *online* tentang transportasi dengan menggunakan algoritma klasifikasi *Naive Bayes Classifier*, *Support Vector Machine* dan KNN, serta manakah algoritma klasifikasi yang memiliki akurasi lebih tinggi antara *Naive Bayes Classifier*, *Support Vector Machine* dan KNN berbasis *Particle Swarm Optimization (PSO)*.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan di dalam tesis ini sebagai berikut :

1. Sebagai bahan perbaikan sistem gojek maupun grab terhadap pelayanan konsumen.

2. Agar penerimaan pengemudi dapat dilakukan lebih baik lagi karena ada beberapa pengemudi yang tidak sopan terhadap konsumen.
3. Mengklasifikasikan sentimen pada sebuah komentar dengan *preprocessing* dan menggunakan metode *Naive Bayes Classifier*, *Support Vector Machine*, *KNN* dan *PSO* sehingga bisa mempercepat proses klasifikasi dan mendapatkan kategori sentimen yang sesuai.
4. Membantu pengembangan dan penerapan teori dalam membantu para peneliti yang berkaitan dengan analisis.

Sedangkan tujuan penulisan tesis ini adalah untuk mendapatkan *classifier* yang terbaik dalam menentukan akurasi untuk mengklasifikasi analisis sentimen dari media sosial mengenai transportasi online dan melihat pengaruh dari penerapan fitur seleksi *Particle Swarm Optimization*.

1.4. Ruang Lingkup

Dalam hal ini penulis membatasi pembahasan atau ruang lingkup yang ada di dalam penulisan tesis. Ruang lingkup yang penulis sajikan di dalam membahas mengenai analisis sentimen transportasi online grab dan gojek terdiri dari penggunaan algoritma *Naive Bayes Classifier*, *Support Vector Machine* dan *KNN*. Selain itu menerapkan metode pemilihan fitur seleksi *Particle Swarm Optimization* untuk meningkatkan nilai akurasi dari masing-masing algoritma yang digunakan serta membandingkan akurasi yang paling tinggi. Data diambil dari twitter yang berhubungan dengan Grab dan Gojek sebanyak 200 data terdiri dari 100 data komentar positif dan 100 data komentar negatif. Begitu juga dengan data yang diambil dari komentar playstore sebanyak 200 data terdiri dari 100 data komentar positif dan 100 data komentar negatif dengan demikian total data yang diambil sebanyak 400 data.

1.5. Hipotesis

Dalam penelitian ini terdapat hipotesis berdasarkan judul dan latar belakang.

1. Nilai akurasi Algoritma *Support Vector Machine* lebih rendah jika tidak menggunakan *Particle Swarm Optimization*.

2. Nilai akurasi Algoritma Support Vector Machine lebih tinggi menggunakan *Particle Swarm Optimization*.
3. Nilai akurasi Algoritma *Naive Bayes* lebih rendah jika tidak menggunakan *Particle Swarm Optimization*.
4. Nilai akurasi Algoritma *Naive Bayes* lebih tinggi menggunakan *Particle Swarm Optimization*.
5. Nilai akurasi Algoritma KNN lebih rendah jika tidak menggunakan *Particle Swarm Optimization*.
6. Nilai akurasi Algoritma KNN lebih tinggi menggunakan *Particle Swarm Optimization*.

1.6 Sistematika Penulisan

Sistematika penulisan tesis mengenai analisis sentimen transportasi online dapat disusun sebagai berikut :

Bab I : PENDAHULUAN

Pada bab ini berisi uraian fakta-fakta berkaitan permasalahan yang akan diteliti beserta rencana penelitian untuk mendapatkan solusi atas permasalahan tersebut,

Bab II : LANDASAN/KERANGKA PEMIKIRAN

Pada bab ini akan dijelaskan tentang landasan-landasan teoritis yang digunakan untuk melaksanakan penelitian dan uraian sistematis dari penelitian-penelitian terkait.

Bab III : METODE PENELITIAN

Pada bab ini membahas metode pengumpulan data dari eksperimen. Menguji algoritma *Naive Bayes Classifier* dan *Support Vector Machine* menggunakan metode pemilihan fitur yaitu *Particle Swarm Optimization* untuk meningkatkan akurasi dalam mengklasifikasi komentar twitter dan playstore.

Bab IV : HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini berisi tentang pembahasan yang dilakukan tentang tata cara *text mining* yang dilakukan, penggunaan algoritma *text mining* yaitu *Naive Bayes Classifier*, *Support Vector Machine* dan KNN, langkah-langkah

perhitungan dari setiap metode yang digunakan, akurasi dari masing-masing metode yang digunakan sebelum dan sesudah penggunaan pemilihan fitur *Particle Swarm Optimization*, membandingkan dari kedua model untuk melihat tingkat akurasi yang paling tinggi.

Bab V : PENUTUP

Pada bab ini berisi tentang kesimpulan dari penjelasan pada bab-bab sebelumnya dan saran untuk penelitian selanjutnya.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Business Understanding*

Pada tahapan *business understanding*, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan menggali informasi melalui komentar mengenai grab dan gojek. Motivasi pada fase ini yaitu komentar yang disajikan biasanya dalam bentuk teks pada media digital yang dikelompokkan berdasarkan isi pembahasan dari masing-masing kategori komentar. Analisis sentimen ini dilakukan untuk mencari metode klasifikasi yang dapat membantu dalam menentukan komentar positif dan negatif.

Pada tahap ini juga dilakukan pemahaman untuk mencari metode klasifikasi yang terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma yang digunakan dan untuk meningkatkan performa dari metode klasifikasi dapat dilakukan dengan menggunakan *feature selection*.

4.2 *Data Understanding*

Pada tahap *data understanding*, dilakukan proses pengambilan data mentah sesuai dengan atribut yang dibutuhkan. Data diperoleh dari twitter dan playstore mengenai grab dan gojek. Dengan menggunakan sumber data yang diperoleh, dibuat dataset dengan atribut yaitu data komentar yang berisi semua data-data komentar tentang gojek dan grab. Semua data komentar tersebut dikelompokkan menjadi satu baik itu komentar positif atau komentar negatif dan disimpan dalam bentuk ekstensi .xlsx

4.3 *Data Preparation*

Tahap *data preparation* merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Dalam *text mining* tahapan awal yang akan dilakukan adalah tahap *preprocessing*. Berikut merupakan tahapan yang dilakukan dalam *preprocessing*:

4.3.1 Tokenize

Dalam proses tokenisasi ini, semua kata yang ada di dalam tiap dokumen akan dikumpulkan dan dihilangkan tanda baca, serta dihilangkan juga simbol, karakter khusus ataupun bukan huruf.

Tabel 4.1 Perbandingan Teks Sebelum dan Sesudah Proses *Tokenize*

No	Sebelum	Sesudah
1	Oknum Driver Gojek Diduga Tipu Pemesan Go Food hingga Rp 9 Juta https://t.co/K04NYfuyO0 #LINETODAY - tiati gaes kal... https://t.co/13y7n6aFy5	Oknum Driver Gojek Diduga Tipu Pemesan Go Food hingga Rp Juta https://t.co/K04NYfuyO0 LINETODAY tiati gaes kal https://t.co/13y7n6aFy5
2	Kronologi Pelanggan Tertipu Oknum Driver hingga Rp 9 Juta, Gojek Merespons, Tak Ada Ganti Rugi https://t.co/OX2Cz39E1Q	Kronologi Pelanggan Tertipu Oknum Driver hingga Rp 9 Juta Gojek Merespons Tak Ada Ganti Rugi https://t.co/OX2Cz39E1Q
3	Gojek sudah menonaktifkan akun driver dengan inisial AY sebagai respon aduan dari Agnes. https://t.co/ZYbdHbWj5F	Gojek sudah menonaktifkan akun driver dengan inisial AY sebagai respon aduan dari Agnes https://t.co/ZYbdHbWj5F
4	@kompascom Marak pembajakan akun, keamanan akun driver gojek cuma mengandalkan OTP beda dengan akun driver grab mengandalkan VERIFIKASI MUKA	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan OTP beda dengan akun driver grab mengandalkan VERIFIKASI MUKA
5	Pelanggannya Tertipu Oknum Driver hingga Rp 9 Juta, Ini Respons Gojek https://t.co/E7iL1KSQfo	Pelanggannya Tertipu Oknum Driver hingga Rp 9 Juta Ini Respons Gojek https://t.co/E7iL1KSQfo
6	Judulnya misleading Harusnya: "Akun Seorang Driver Gojek Diredas, Lalu Digunakan Untuk Menipu Pelanggan Rp 28 Juta" https://t.co/kuukDdwcd4	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diredas Lalu Digunakan Untuk Menipu Pelanggan Rp 28 Juta https://t.co/kuukDdwcd4

7	Dear Driver gojek&grab, kalau punya helm baiknya di cuci jangan di semprot parfum aja. Saya pribadi kalau ada helm... https://t.co/kVLNJxSkT4	Dear Driver gojek grab kalau punya helm baiknya di cuci jangan di semprot parfum aja Saya pribadi kalau ada helm https://t.co/kVLNJxSkT4
8	Pelanggan Gojek Tertipu Oknum Driver, Rp 9 Juta Raib https://t.co/Me4SJeO7RX	Pelanggan Gojek Tertipu Oknum Driver Rp Juta Raib https://t.co/Me4SJeO7RX
9	@awaliasista Mari budayakan malas pergi buat beli makan — tulisan di warung ayam geprek yg punyaanya driver gojek	awaliasista Mari budayakan malas pergi buat beli makan tulisan di warung ayam geprek yg punyaanya driver gojek
10	@mbaknefi Coba order ke gojek, orderanmu pasti jadi prioritas driver	mbaknefi Coba order ke gojek orderanmu pasti jadi prioritas driver

4.3.2 Filter Token (By Length)

Dalam proses ini, kata-kata yang memiliki panjang kurang dari 4 atau lebih dari 25 akan dihapus, seperti kata yg, tdk, jd, ga, ane, gan yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

Tabel 4.2 Perbandingan Teks Sebelum dan Sesudah Proses *Filter Token* (By Length)

No	Sebelum	Sesudah
1	Oknum Driver Gojek Diduga Tipu Pemesan Go Food hingga Rp Juta https://t.co/KNYfuyO LINETODAY tiati gaes kal https://t.co/yNaFy	Oknum Driver Gojek Diduga Tipu Pemesan Food hingga Rp Juta https://t.co/KNYfuyO LINETODAY tiati gaes https://t.co/yNaFy
2	Kronologi Pelanggan Tertipu Oknum Driver hingga Rp Juta Gojek Merespons Tak Ada Ganti Rugi https://t.co/OXCzEQ	Kronologi Pelanggan Tertipu Oknum Driver hingga Rp Juta Gojek Merespons Ganti Rugi https://t.co/OXCzEQ

3	Gojek sudah menonaktifkan akun driver dengan inisial AY sebagai respon aduan dari Agnes https t co ZYbdHbWj F	Gojek sudah menonaktifkan akun driver dengan inisial sebagai respon aduan dari Agnes https ZYbdHbWj
4	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan OTP beda dengan akun driver grab mengandalkan VERIFIKASI MUKA	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan beda dengan akun driver grab mengandalkan VERIFIKASI MUKA
5	Pelanggannya Tertipu Oknum Driver hingga Rp Juta Ini Respons Gojek https t co E iL KSQfo	Pelanggannya Tertipu Oknum Driver hingga Juta Respons Gojek https KSQfo
6	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diredas Lalu Digunakan Untuk Menipu Pelanggan Rp Juta https t co kuukDdwcd	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diredas Lalu Digunakan Untuk Menipu Pelanggan Juta https kuukDdwcd
7	Dear Driver gojek grab kalau punya helm baiknya di cuci jangan di semprot parfum aja Saya pribadi kalau ada helm https t co kVLNJxSkT	Dear Driver gojek grab kalau punya helm baiknya cuci jangan semprot parfum Saya pribadi kalau helm https kVLNJxSkT
8	Pelanggan Gojek Tertipu Oknum Driver Rp Juta Raib https t co Me SJeO RX	Pelanggan Gojek Tertipu Oknum Driver Juta Raib https SJeO
9	awaliasista Mari budayakan malas pergi buat beli makan tulisan di warung ayam geprek yg punyaanya driver gojek	awaliasista Mari budayakan malas pergi buat beli makan tulisan warung ayam geprek punyaanya driver gojek
10	mbaknefi Coba order ke gojek orderanmu pasti jadi prioritas driver	mbaknefi Coba order gojek orderanmu pasti jadi prioritas driver

4.3.3 Stopword Removal

Pada tahap ini, operator yang digunakan adalah *filter stopwords (dictionary)* karena dataset yang digunakan berbahasa Indonesia. Pada proses ini terlebih dahulu dibuat daftar kata-kata yang termasuk *stopwords* kemudian *file* nya akan diupload ke dalam operator *filter stopwords (dictionary)*. Dalam tahap ini, kata-kata yang tidak relevan akan dihapus seperti kata tetapi, untuk, dengan, yang

merupakan kata-kata yang tidak memiliki makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan sentimen.

Tabel 4.3 Perbandingan Teks Sebelum dan Sesudah Proses *Stopwords Removal*

No	Sebelum	Sesudah
1	Oknum Driver Gojek Diduga Tipu Pemesan Food hingga Juta https NYfuyO LINETODAY tiati gaes https	Oknum Driver Gojek Diduga Tipu Pemesan Food hingga Juta LINETODAY tiati gaes
2	Kronologi Pelanggan Tertipu Oknum Driver hingga Juta Gojek Merespons Ganti Rugi https	Kronologi Pelanggan Tertipu Oknum Driver hingga Juta Gojek Merespons Ganti Rugi
3	Gojek sudah menonaktifkan akun driver dengan inisial sebagai respon aduan dari Agnes https ZYbdHbWj	Gojek sudah menonaktifkan akun driver dengan inisial sebagai respon aduan dari Agnes
4	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan beda dengan akun driver grab mengandalkan VERIFIKASI MUKA	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan beda dengan akun driver grab mengandalkan VERIFIKASI MUKA
5	Pelanggannya Tertipu Oknum Driver hingga Juta Respons Gojek https KSQfo	Pelanggannya Tertipu Oknum Driver hingga Juta Respons Gojek
6	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diredas Lalu Digunakan Untuk Menipu Pelanggan Juta https kuukDdwcd	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diredas Lalu Digunakan Untuk Menipu Pelanggan Juta
7	Dear Driver gojek grab kalau punya helm baiknya cuci jangan semprot parfum Saya pribadi kalau helm https kVLNJxSkT	Dear Driver gojek grab kalau punya helm baiknya cuci jangan semprot parfum Saya pribadi kalau helm
8	Pelanggan Gojek Tertipu Oknum Driver Juta Raib https SJeO	Pelanggan Gojek Tertipu Oknum Driver Juta Raib

9	awaliasista Mari budayakan malas pergi buat beli makan tulisan warung ayam geprek punyaanya driver gojek	awaliasista Mari budayakan malas pergi buat beli makan tulisan warung ayam geprek punyaanya driver gojek
10	mbaknefi Coba order gojek orderanmu pasti jadi prioritas driver	mbaknefi Coba order gojek orderanmu pasti jadi prioritas driver

4.3.4 Transform Cases

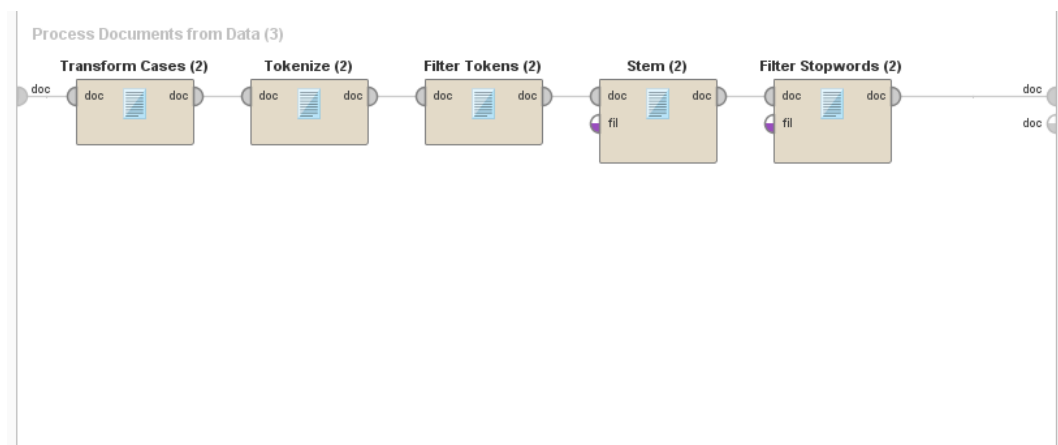
Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat berhubungan dengan sentimen.

Tabel 4.4 Perbandingan Teks Sebelum dan Sesudah Proses Transform Cases

No	Sebelum	Sesudah
1	Oknum Driver Gojek Diduga Tipu Pemesan Food hingga Juta LINETODAY tiati gaes	oknum driver gojek diduga tipu pemesan food hingga juta linetoday tiati gaes
2	Kronologi Pelanggan Tertipu Oknum Driver hingga Juta Gojek Merespons Ganti Rugi	kronologi pelanggan tertipu oknum driver hingga juta gojek merespons ganti rugi
3	Gojek sudah menonaktifkan akun driver dengan inisial sebagai respon aduan dari Agnes	gojek sudah menonaktifkan akun driver dengan inisial sebagai respon aduan dari agnes
4	kompascom Marak pembajakan akun keamanan akun driver gojek cuma mengandalkan beda dengan akun driver grab mengandalkan VERIFIKASI MUKA	kompascom marak pembajakan akun keamanan akun driver gojek cuma mengandalkan beda dengan akun driver grab mengandalkan verifikasi muka
5	Pelanggannya Tertipu Oknum Driver hingga Juta Respons Gojek	pelanggannya tertipu oknum driver hingga juta respons gojek

6	Judulnya misleading Harusnya Akun Seorang Driver Gojek Diretas Lalu Digunakan Untuk Menipu Pelanggan Juta	judulnya misleading harusnya akun seorang driver gojek diretas lalu digunakan untuk menipu pelanggan juta
7	Dear Driver gojek grab kalau punya helm baiknya cuci jangan semprot parfum Saya pribadi kalau helm	dear driver gojek grab kalau punya helm baiknya cuci jangan semprot parfum saya pribadi kalau helm
8	Pelanggan Gojek Tertipu Oknum Driver Juta Raib	pelanggan gojek tertipu oknum driver juta raib
9	awaliasista Mari budayakan malas pergi buat beli makan tulisan warung ayam geprek punyanya driver gojek	awaliasista mari budayakan malas pergi buat beli makan tulisan warung ayam geprek punyanya driver gojek
10	mbaknefi Coba order gojek orderanmu pasti jadi prioritas driver	mbaknefi coba order gojek orderanmu pasti jadi prioritas driver

Berikut merupakan tahap desain model *preprocessing* yang digunakan untuk tahapan awal pengolahan data :



Gambar 4.1 Desain *Preprocessing*

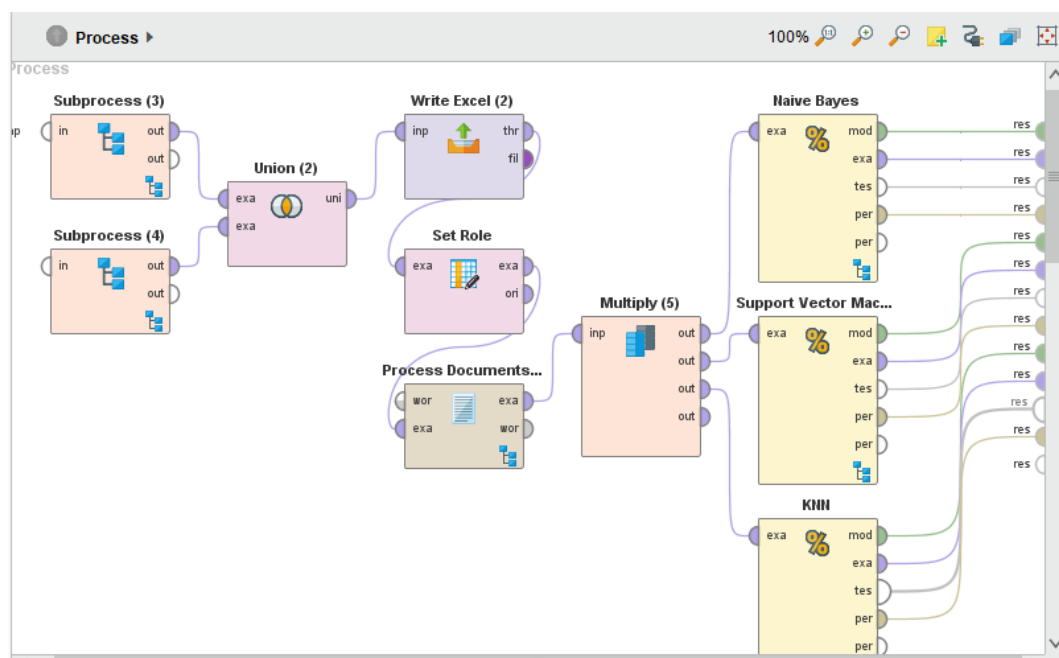
Gambar 4.1 Menjelaskan tentang tahapan *preprocessing*, dimana tahapan tersebut terdapat di dalam operator *Process Document From Files*. Tahapan

preprocessing yang digunakan yaitu *tokenize*, *filter tokens (by length)*, *filter stopwords (dictionary)* dan *transform cases*.

4.4 Pemodelan (*Modelling*)

Merupakan fase pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. *Tool* yang digunakan adalah *RapidMiner* versi 7.3. Hasil pengujian model yang dilakukan adalah mengklasifikasikan komentar positif dan komentar negatif menggunakan algoritma *Naive Bayes Classifier* dan *Naive Bayes Classifier* berbasis *Particle Swarm Optimization (PSO)* serta *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization (PSO)* untuk mendapatkan nilai akurasi terbaik. Berikut adalah desain model *Rapidminer* yang digunakan yaitu :

4.4.1 Desain Model *Naive Bayes Classifier*, *Support Vector Machine* dan *KNN*



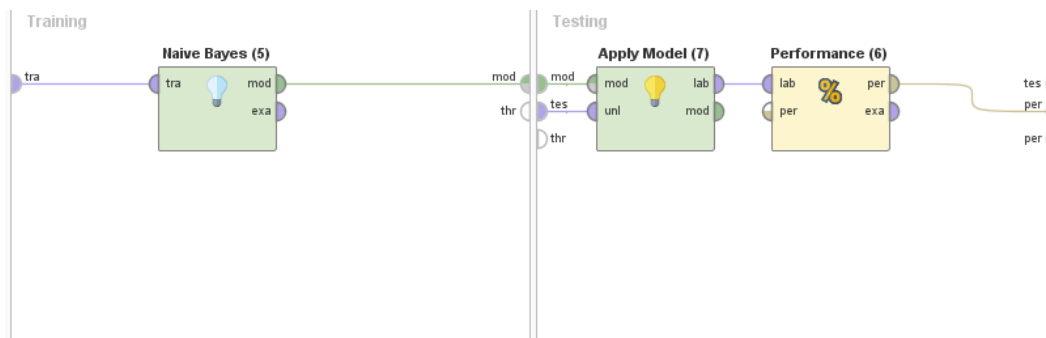
Gambar 4.2 Desain Algoritma (NB, SVM dan KNN)

Berikut adalah desain model *Rapidminer* untuk tahapan algoritma *Naive Bayes Classifier* berbasis *Particle Swarm Optimization (PSO)* dan *Support Vector Machine* berbasis *Particle Swarm Optimization (PSO)* :

Gambar 4.4 Menjelaskan desain proses di dalam operator *cross validation* SVM . Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator *Read Excel* , hal ini dilakukan karena dataset disimpan dalam bentuk Excel (.xlsx). *Process documents from files* untuk mengkonversi *files* menjadi dokumen. Proses validasi terdiri dari *data training* dan *data testing*.

4.5.2 Model Naive Bayes Classifier

Berikut ini desain proses pengujian model metode *Naive Bayes Classifier* yang digunakan yaitu :

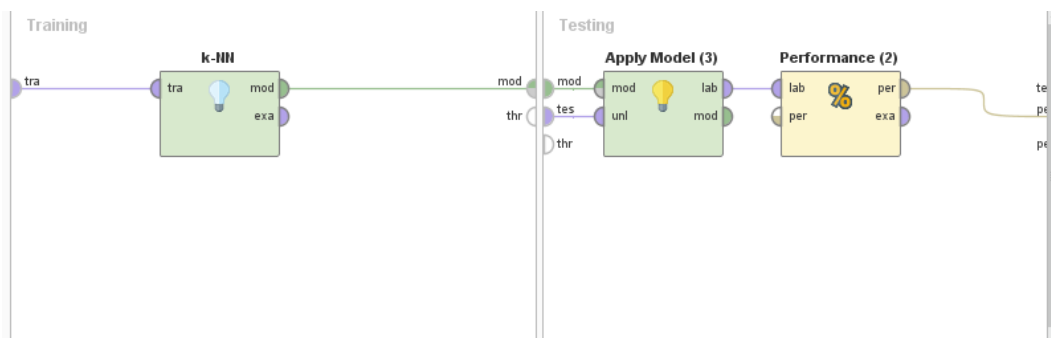


Gambar 4.5 Desain Proses 10-Fold Cross Validation untuk NB

Gambar 4.5 Menjelaskan desain proses di dalam operator *cross validation* NB. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Proses validasi terdiri dari *data training* dan *data testing*. Subproses *training* digunakan untuk melatih model. Model yang dilatih kemudian diterapkan pada subproses *testing* .

4.5.3 Model KNN

Berikut ini desain proses pengujian model metode *K-Nearest Neighbor* yang digunakan yaitu :



Gambar 4.6 Desain Proses 10-Fold Cross Validation untuk KNN

Gambar 4.5 Menjelaskan desain proses di dalam operator *cross validation* KNN. Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Proses validasi terdiri dari *data training* dan *data testing*. Subproses *training* digunakan untuk melatih model. Model yang dilatih kemudian diterapkan pada subproses *testing*.

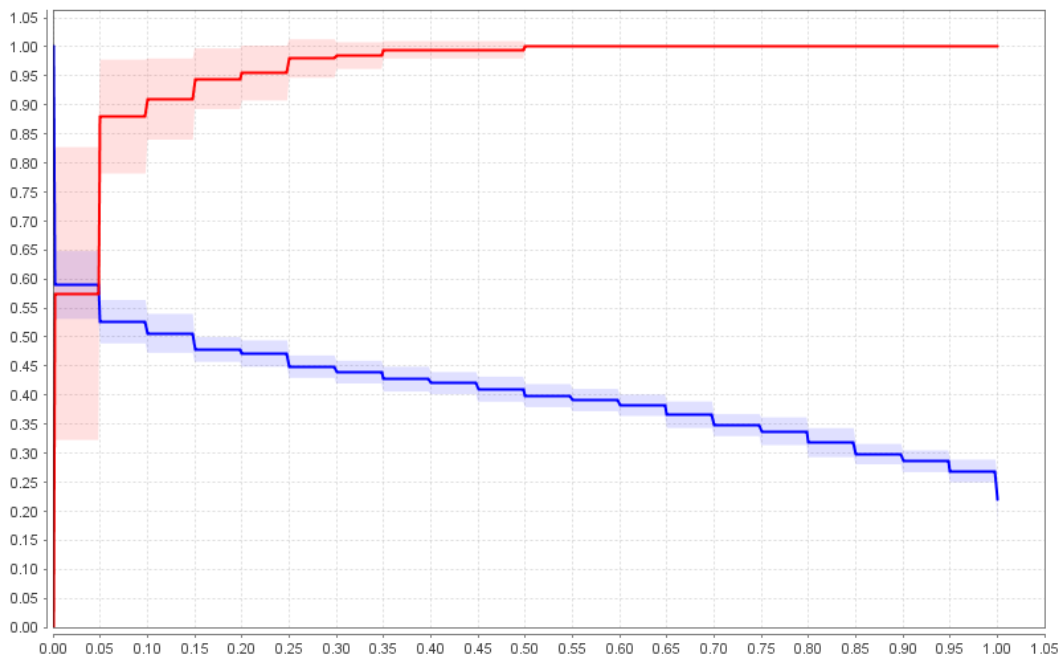
Pada pengujian kali ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator Read Excel, hal ini dilakukan karena dataset disimpan dalam bentuk Excel. *Process documents from files* untuk mengkonversi *files* menjadi dokumen. Proses validasi terdiri dari *data training* dan *data testing*. Pada tahap ini juga menggunakan *Set Role* yang berfungsi untuk menentukan *field* pada kelas kemudian menggunakan *Particle Swarm Optimization (PSO)* agar akurasi yang dihasilkan lebih tinggi. Dari hasil pemodelan yang telah dilakukan sebelumnya Berikut ini akan dijelaskan AUC dan *Accuracy* dari masing-masing algoritma :

1. AUC Support Vector Machine

Berikut ini akan dijelaskan AUC dan *Accuracy* dari algoritma *Support Vector Machine*:

AUC: 0.961 +/- 0.020 (mikro: 0.961) (positive class: negative)

— ROC — ROC (Thresholds)



Gambar 4.7 AUC SVM

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC SVM dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.7 diatas sebesar 0.961 dimana diagnosa hasilnya adalah *excellent classification*.

2. Accuracy

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 89.00% +/- 3.00% (mikro: 89.00%)

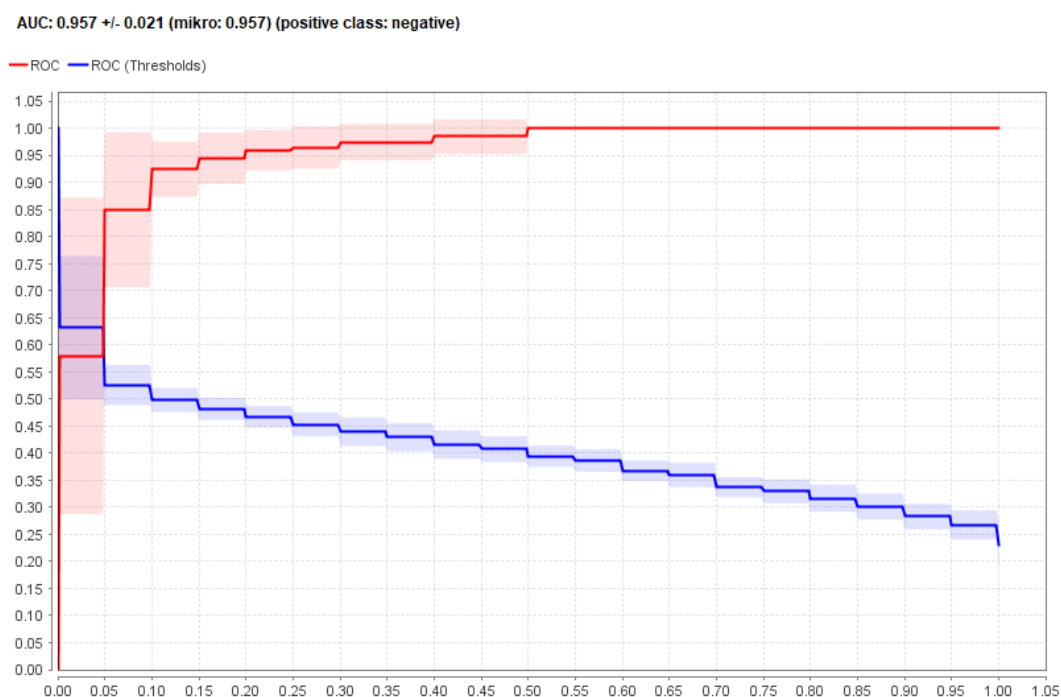
	true positive	true negative	class precision
pred. positive	179	23	88.61%
pred. negative	21	177	89.39%
class recall	89.50%	88.50%	

Gambar 4.8. Accuracy SVM

Akurasi yang diperoleh yaitu 89 % dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab. Data komentar positif yang sesuai prediksi yaitu 179 data. Data komentar positif yang tidak sesuai prediksi yaitu 21 data. Data komentar negative yang sesuai prediksi yaitu 177 data. Data komentar negative yang tidak sesuai prediksi yaitu 23 data.

1. AUC Support Vector Machine (PSO)

Berikut ini akan dijelaskan AUC dan Accuracy dari algoritma *Support Vector Machine (PSO)*:



Gambar 4.9. AUC SVM berbasis Particle Swarm Optimization (PSO)

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*

2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.9 diatas sebesar 0.957 dimana diagnosa hasilnya adalah *excellent classification*.

2. Accuracy SVM PSO

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 91.75% +/- 4.04% (mikro: 91.75%)

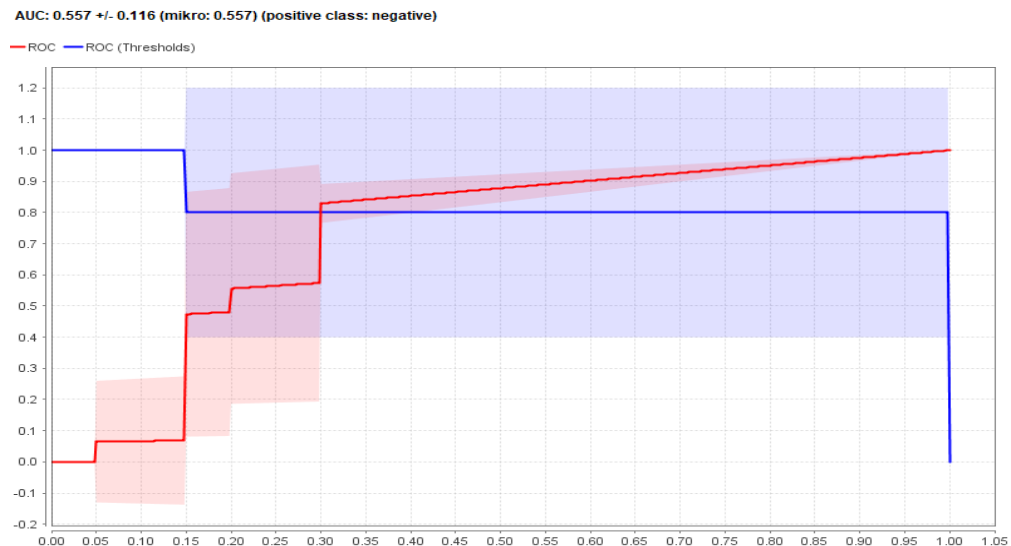
	true positive	true negative	class precision
pred. positive	183	16	91.96%
pred. negative	17	184	91.54%
class recall	91.50%	92.00%	

Gambar 4.10. Accuracy SVM (PSO)

Akurasi yang diperoleh yaitu 91.75 % dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab. Data komentar positif yang sesuai prediksi yaitu 183 data. Data komentar positif yang tidak sesuai prediksi yaitu 17 data. Data komentar negative yang sesuai prediksi yaitu 184 data. Data komentar negative yang tidak sesuai prediksi yaitu 16 data.

2. AUC Naive Bayer Classifier

Berikut ini akan dijelaskan AUC dan *Accuracy* dari algoritma *Naive Bayer Classifier*:



Gambar 4.11. AUC NB

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC NB dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.11 diatas sebesar 0.557 dimana diagnosa hasilnya adalah *failure*

2. Accuracy (Accuracy)

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 80.00% +/- 3.54% (mikro: 80.00%)

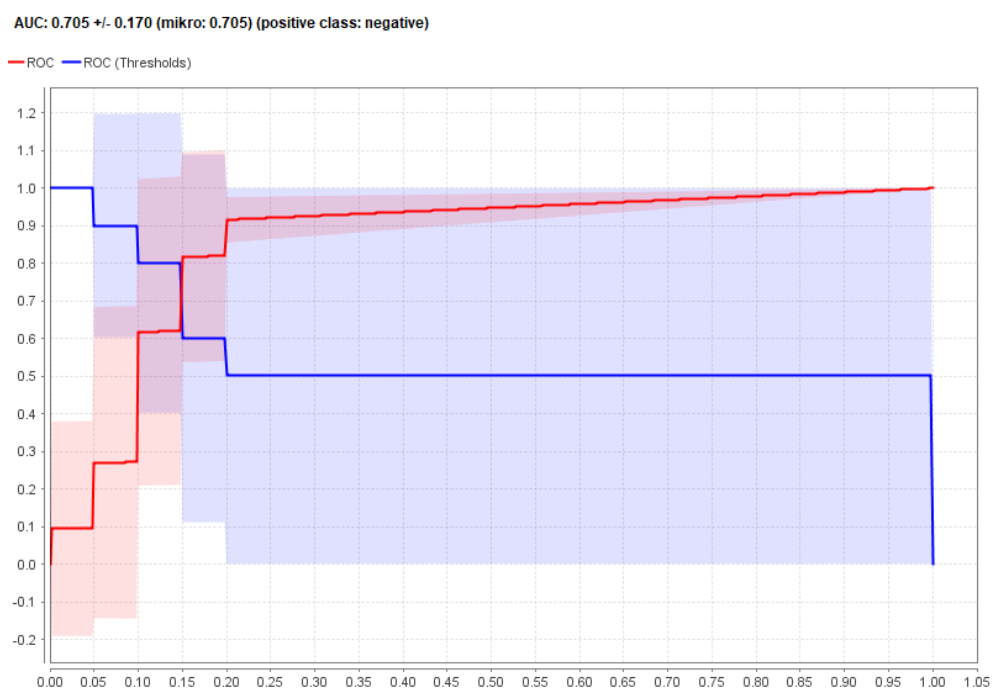
	true positive	true negative	class precision
pred. positive	162	42	79.41%
pred. negative	38	158	80.61%
class recall	81.00%	79.00%	

Gambar 4.12. Accuracy NB

Akurasi yang diperoleh yaitu 80 % dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab. Data komentar positif yang sesuai prediksi yaitu 162 data. Data komentar negatif yang termasuk ke dalam prediksi positif yaitu 38 data. Data komentar positif yang termasuk kedalam prediksi negatif yaitu 158 data dan data komentar negatif yang sesuai prediksi yaitu 42 data.

3. AUC Naive Bayes Classifier (PSO)

Berikut ini akan dijelaskan AUC dan Accuracy dari algoritma Naive Bayes Classifier (PSO):



Gambar 4.13. AUC Naive Bayes Classifier berbasis Particle Swarm Optimization (PSO)

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC *Naive Bayes Classifier* berbasis *Particle Swarm Optimization (PSO)* dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.13 diatas sebesar 0.705 dimana diagnosa hasilnya adalah *classification*.

4. Accuracy (Accuracy)

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 89.00% +/- 4.06% (mikro: 89.00%)

	true positive	true negative	class precision
pred. positive	180	24	88.24%
pred. negative	20	176	89.80%
class recall	90.00%	88.00%	

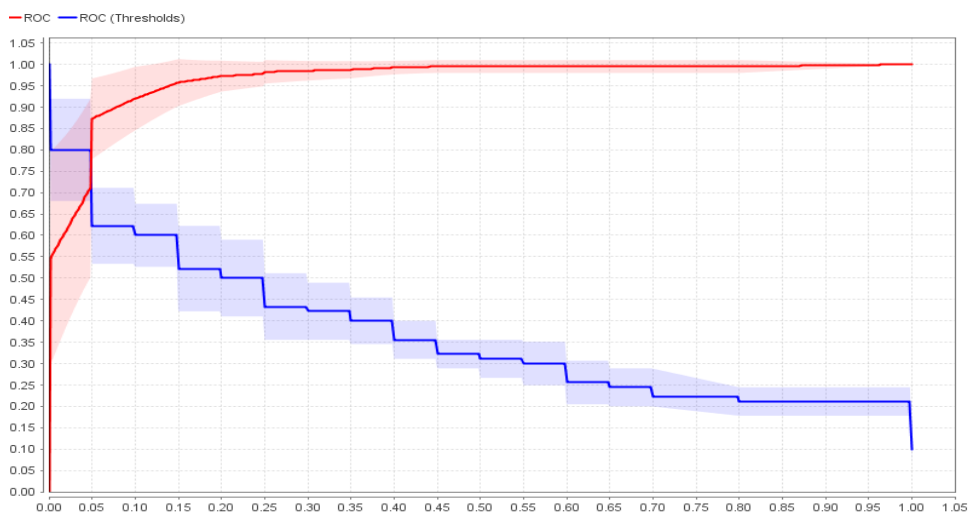
Gambar 4.14. Accuracy NB (PSO)

Akurasi yang diperoleh yaitu 89% dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab. Data komentar positif yang sesuai prediksi yaitu 180 data. Data komentar negatif yang termasuk ke dalam prediksi positif yaitu 20 data. Data komentar neagtive yang termasuk kedalam prediksi postive yaitu 24 data dan data komentar negatif yang sesuai prediksi yaitu 176 data.

3. AUC K-Nearest Neighbor

Berikut ini akan dijelaskan AUC dan *Accuracy* dari algoritma *K-Nearest Neighbor*:

AUC: 0.966 +/- 0.024 (mikro: 0.966) (positive class: negative)



Gambar 4.15. AUC KNN

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC KNN dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.15 diatas sebesar 0.966 dimana diagnosa hasilnya adalah *excellent classification*

2. Accuracy (Accuracy)

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 91.00% +/- 3.91% (mikro: 91.00%)

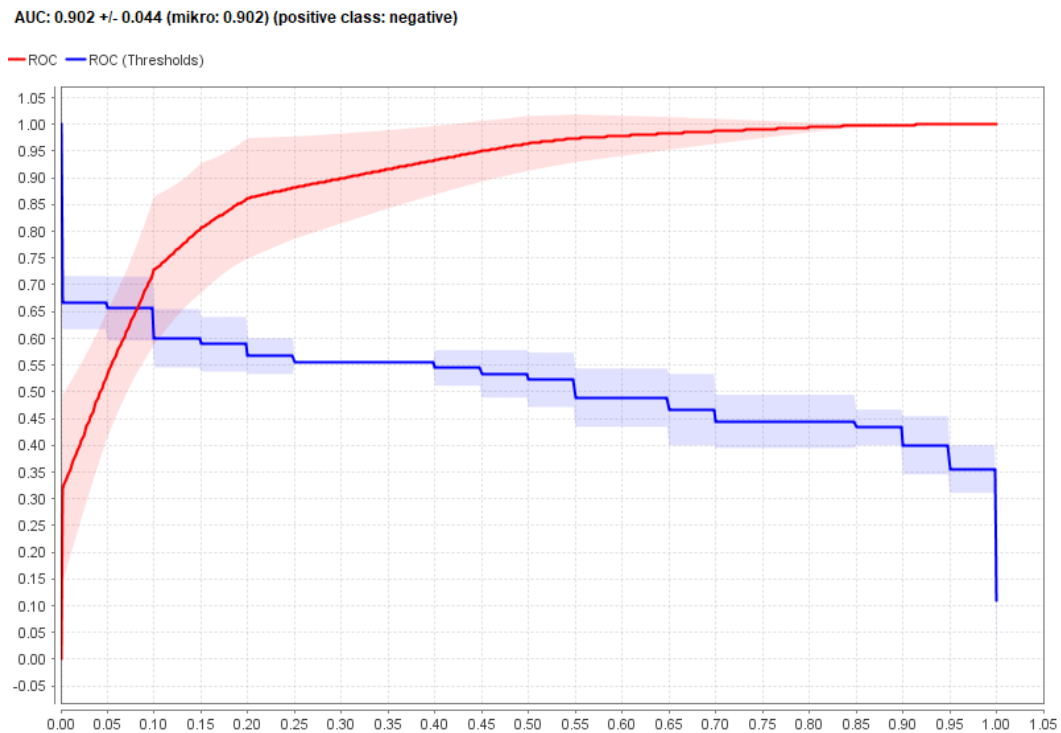
	true positive	true negative	class precision
pred. positif	177	13	93.16%
pred. negatif	23	187	89.05%
class recall	88.50%	93.50%	

Tabel 4.16. Accuracy KNN

Akurasi yang diperoleh yaitu 91% dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab dengan nilai K-9. Data komentar positif yang sesuai prediksi yaitu 177 data. Data komentar negatif yang termasuk ke dalam prediksi positif yaitu 23 data. Data komentar positif yang termasuk kedalam prediksi negatif yaitu 13 data dan data komentar negatif yang sesuai prediksi yaitu 187 data.

5. AUC K-Nearest Neighbor (PSO)

Berikut ini akan dijelaskan AUC dan *Accuracy* dari algoritma *K-Nearest Neighbor (PSO)*:



Gambar 4.17. AUC *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* (PSO)

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

AUC *K-Nearest Neighbor* berbasis *Particle Swarm Optimization* (PSO) dengan nilai AUC (*Area Under Curve*) yang dihasilkan dari gambar 4.17 diatas sebesar 0.902 dimana diagnosa hasilnya adalah *excellent classification*.

6. Accuracy (Accuracy)

Accuracy memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Accuracy* berisi informasi aktual dan prediksi pada sistem klasifikasi.

accuracy: 85.25% +/- 6.47% (mikro: 85.25%)

	true positive	true negative	class precision
pred. positive	175	34	83.73%
pred. negative	25	166	86.91%
class recall	87.50%	83.00%	

Gambar 4.18. Accuracy K-Nearest Neighbor (PSO)

Akurasi yang diperoleh yaitu 85.25% dari 200 data komentar positif dan 200 komentar negatif tentang Gojek dan Grab. Data komentar positif yang sesuai prediksi yaitu 175 data. Data komentar negatif yang termasuk ke dalam prediksi positif yaitu 25 data. Data komentar neagtive yang termasuk kedalam prediksi postive yaitu 34 data dan data komentar negatif yang sesuai prediksi yaitu 166 data.

Adapun perbandingan hasil komparasi akurasi dan AUC Algoritma telah digunakan sebagai berikut :

Tabel 4.5. Akurasi dan AUC

Algoritma	Accuracy	AUC
SVM	89%	0.961
SVM + PSO	91.75 %	0.957
NB	80 %	0.557
NB + PSO	89 %	0.705
KNN	91%	0.966
KNN + PSO	85.25 %	0.902

Dalam penelitian ini, hasil perhitungan metode SVM mendapatkan nilai akurasi 89% sedangkan hasil perhitungan SVM (PSO) mendapatkan nilai akurasi 91.75%. Hasil perhitungan metode NB menghasilkan nilai akurasi 80%, sedangkan hasil perhitungan NB (PSO) mendapatkan nilai akurasi 89%. Hasil perhitungan metode KNN menghasilkan nilai akurasi 91%, sedangkan hasil

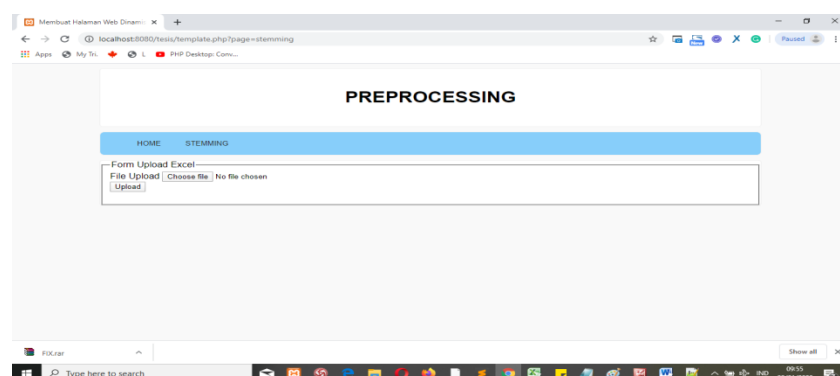
perhitungan KNN (PSO) mendapatkan nilai akurasi 85.25%. Akurasi SVM, NB memiliki perbedaan nilai sekitar 9 %, Akurasi SVM dan KNN memiliki perbedaan nilai sekitar 2%, Akurasi SVM (PSO) dan NB (PSO) memiliki perbedaan nilai sekitar 2.75 %, Akurasi SVM (PSO) dan KNN (PSO) memiliki perbedaan nilai sekitar 5.5%. Berdasarkan Tabel 4.5, dapat disimpulkan bahwa akurasi SVM berbasis PSO mendapatkan akurasi yang lebih tinggi dibandingkan dengan SVM, NB, NB (PSO), KNN dan KNN (PSO) dapat meningkatkan nilai akurasi untuk metode SVM, NB dan KNN. Model klasifikasi teks yang digunakan dapat memudahkan untuk mengetahui komentar positif dan komentar negatif. Berdasarkan data komentar yang diolah menggunakan *tool Rapidminer*, data komentar akan terpisah menjadi kata-kata yang memiliki bobot pada setiap kata-katanya. Kata-kata tersebut akan digunakan untuk melihat kata –kata yang berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi dan dapat digunakan untuk mengetahui komentar positif dan komentar negatif.

4.6. Deployment

deployment yang dibuat yaitu proses stemming dan menggunakan bahasa program php dengan menggunakan library sastrawi jawa. Berikut merupakan tampilan dari halaman menu dalam analisis sentimen terhadap komentar Gojek dan Grab antara lain :

1. Halaman Menu Utama

Berikut ini tampilan menu utama Gojek dan Grab sebagai berikut :

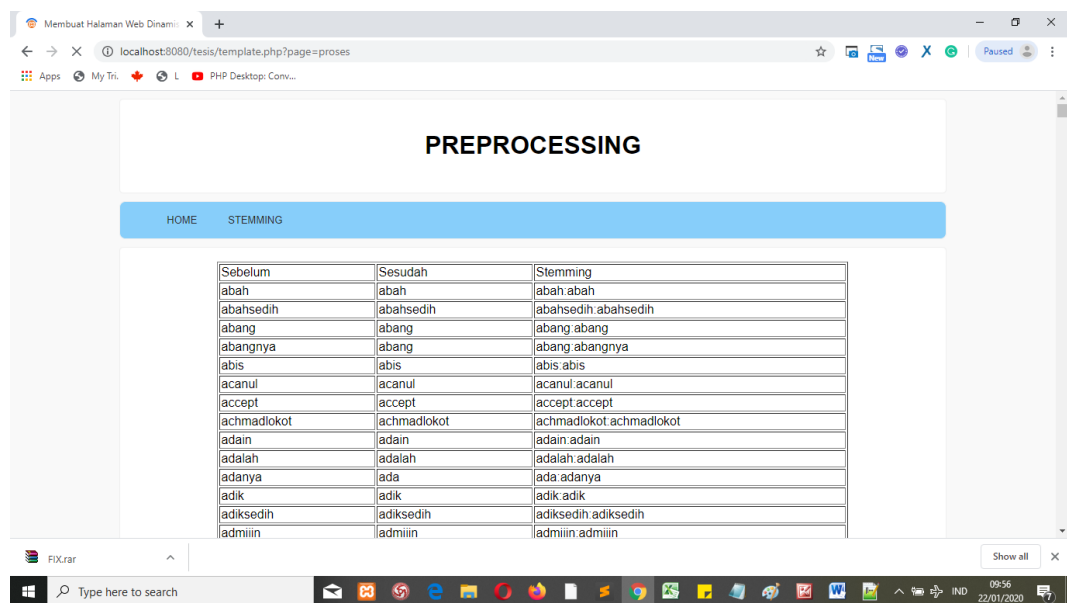


Gambar 4.19. Tampilan Halaman Utama

Gambar 4.21 Menjelaskan tentang halaman menu utama. Pada halaman ini proses pengujian terhadap teks tentang Gojek dan Grab dilakukan. Teks komentar dapat di masukkan ke dalam konten. Setelah teks sudah dimasukkan maka klik tombol proses.

2. Halaman Hasil *Text Mining* (*Tokenizing*, *Stemming*, *Stopword*)

Berikut ini tampilan menu Hasil *Text Mining Tokenizing* dalam analisis sentimen terhadap Gojek dan Grab sebagai berikut :



Sebelum	Sesudah	Stemming
abang	abang	abang abang
abahnya	abahnya	abahnya abahnya
abangnya	abangnya	abang abangnya
abis	abis	abis abis
acanul	acanul	acanul acanul
accept	accept	accept accept
achmadlokot	achmadlokot	achmadlokot achmadlokot
adain	adain	adain adain
adalah	adalah	adalah adalah
adanya	ada	ada adanya
adik	adik	adik adik
adiksedih	adiksedih	adiksedih adiksedih
admiin	admiin	admiin admiin

Gambar 4.20. Tampilan Proses Stemming

Gambar 4.22. Menjelaskan tentang halaman menu hasil *text mining* (*Tokenizing*, *Stemming*, *Stopword*). Pada halaman ini menjelaskan hasil dari *tokenizing*, *stemming* dan *stopword* teks dimana semua kata yang ada dalam dokumen akan dikumpulkan dan dihilangkan tanda baca, simbol, karakter khusus ataupun bukan huruf dan menghilangkan kata tidak baku.

DAFTAR REFERENSI

- Al-Amrani, Y., Lazaar, M., El Kadiri, E.K., 2018. Sentiment Analysis Using Hybrid Method of Support Vector Machine And Decision Tree. *Journal of Theoretical and Applied Information Technology* 96(7), pp.1886-1895.
- Liu, B., 2015. *Sentiment Analysis Opinions, Sentiment, and Emotion in Text*. Chicago, USA: Cambridge University Press.
- Han, Jiawei, Micheline Kamber dan Jian Pei.2015. *Data mining Concepts And Techniques Third Edition*. Waltham, MA:Morgan Kaufmann.
- Rokach, Lior dan Oded Maimon.2015. *Data mining With Decision Trees Theory and Applications 2nd Edition*. America : World Scientific Publishing.
- Zgheib, Wassim A dan Aziz M. Barbar.2017. *A Study using Support Vector Machines to Classify the Sentiments of Tweets*.Lebanon : International Journal of Computer Applications (0975 – 8887) Volume 170 – No.2, July 2017.
- Amrullah, Ahmad Afief, Ahmad Tantoni, Nahrowi Hamdani, Rahmat Taufik R.L.Bau, Muhammad Rafiqudin Ahsan dan Ema Utami.2016. Review Atas Analisis Sentimen Pada Twitter Sebagai Representasi Opini Publik Terhadap Bakal Calon Pemimpin.
- Anjani, D. 2015. Bab II Landasan Teori *Text Mining*. <http://repository.widyatama.ac.id/xmlui/bitstream/handle/123456789/5867/Bab%202.pdf?sequence=9>. Yogyakarta : PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U) KE-2 Tahun 2016 Kajian Multi Disiplin Ilmu dalam Pengembangan IPTEKS untuk Mewujudkan Pembangunan Nasional Semesta Berencana (PNSB) sebagai Upaya Meningkatkan Daya Saing Global ISBN: 978-979-3649-96-2.
- Buntoro, Asrofi Ghulam. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di *Twitter*. Jakarta : Integer Journal Vol 1 No 1 Maret 2016:32-41.

- Chandani, Vinita, Romi Satria Wahono, dan Purwanto. 2015. Komparasi Algoritma Klasifikasi *Machine Learning* Dan *Feature Selection* pada Analisis Sentimen *Review Film*, ISSN 2356-3982. __. *Journal of Intelligent Systems*, Vol. 1, No. 1, February 2015
- Dhande, Lina L, dan Dr. Girish Patnaik. 2014. *Review of Sentiment Analysis using Naive Bayes and Neural Network Classifier*, ISSN :2319-8885. India : *International Journal of Scientific Engineering and Technology Research* Vol. 03 Issue.07 Mei 2014 Page 1110-1113.
- Hadna, Nurrin Muchammad Shiddieqy, Paulus Insap Santosa dan Wing Wahyu Winarno. 2016. Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di *Twitter*, ISSN :2089-9815. Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi (Sentika 2016).
- Ibrahim, Danny. 2017. Analisis Hubungan antar Faktor dan Komparasi Algoritma Klasifikasi pada Penentuan Penundaan Penerbangan, ISSN: 2579-9045 ISBN: 978-602-74355-1-3. Tegal : Seminar Nasional IPTEK Terapan (SENIT) 2017.
- Intiyaz, Muhammad Zain, Muhammad Nasrun, dan Umar Ali Ahmad. 2015. Analisis dan Implementasi *Framework* CRISP-DM Untuk Mengetahui Perilaku Data Transaksi Pelanggan, ISSN : 2355-9365. Bandung : *e-Proceeding of Engineering* : Vol.2, No.1 April 2015 | Page 596.
- Indrayuni, Elly. 2016. Analisa Sentimen *Review Hotel* Menggunakan Algoritma *Support Vector Machine* Berbasis *Particle Swarm Optimization*, ISSN: 2338 – 8161. Pontianak : Jurnal Evolusi Volume 4 Nomor 2 – 2016.
- Ipmawati, Joang, Kusrini dan Emha Taufiq Luthfi. 2016. Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen, ISSN :2461-0690. Jakarta : *IJSE – Indonesian Journal on Software Engineering* Vol. 2 No 2.
- Maulana, Rizky dan Sri Redjeki. 2016. Analisis Sentimen Pengguna *Twitter* Menggunakan Metode *Support Vector Machine* Berbasis *Cloud Computing*. Yogyakarta : Jurnal TAM (*Technology Acceptance Model*) Volume 6, Juli 2016.

- Mukminin, Amirul, Dwiza Riana. 2017. Komparasi Algoritma C4.5, *Naïve Bayes* Dan *Neural Network* Untuk Klasifikasi Tanah, ISSN: 2355-6579 E-ISSN: 2528-2247. Bandung : JURNAL INFORMATIKA, Vol.4 No.1 April 2017, pp. 21-31.
- Nurhuda, Faishol, Sari Widya Sihwi dan Afrizal Doewes. 2015. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari *Twitter* Menggunakan Metode *Naive Bayes Classifier*, ISSN :2301-7201. Jakarta : Jurnal ITSMART Vol. 2 No. 2 Desember 2015.
- Saputra, Nurirwan, Teguh Bharata Adji, dan Adhistya Erna Permanasari.2015. Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan SVM.Yogyakarta : Jurnal Dinamika Informatika Volume 5, Nomor 1, November 2015.
- Suyanto. 2017. *Data mining* Untuk Klasifikasi Dan Klasterisasi Data.Bandung:Informatika Bandung.
- Utami, Lilyani Asri. 2017. Analisis Sentimen Opini Berita Kebakaran Hutan Melalui Komparasi Algoritma *Support Vector Machine* Dan *K-Nearest Neighbor* Berbasis *Particle Swarm Optimization*.Jakarta : Jurnal Pilar Nusa Mandiri Volume 13 No.1, Maret 2017.
- Vulandari, Retno Tri.2017. *Data mining* Teori dan Aplikasi Rapidminer. Yogyakarta : Gava Media.
- Widodo, Pangestu, Januar Adi Putra, Suwanto Afiadi, Agus Zainal Arifin, dan Darlis Herumurti. 2016. Klasifikasi Kategori Dokumen Berita Berbahasa Indonesia Dengan Metode Kategorisasi Multi-Label Berbasis *Domain Specific Ontology*, ISSN : 2407 – 391.Surabaya : Jurnal Ilmiah Teknologi Informasi Terapan Volume II, No 2, 15 April 2016.
- Wira, Shantana. (2017, 22 September). Perkembangan Media *Online* di Dunia dan Indonesia. Diperoleh dari : <https://www.kompasiana.com/shantanawp/59c5223778b6851318117972/perkembangan-media-online-di-dunia-dan-indonesia>. (5 November 2017).

Zain , Noorma Fitriana M. 2015. Manfaat Berita *Online*. Diambil dari :
<http://www.noormafitrianamzain.com/2015/06/manfaat-berita-online.html>.
(5 November 2017).