

Analisis Sentimen Terhadap Layanan Internet Di Indonesia Menggunakan Metode K-Nearest Neighbor

Ade Septiansyah¹, Agistia Yuliawati², Imam Santoso³

^{1,2}Sekolah Tinggi Ilmu Cipta Karya Informatika, ³Dosen Universitas Teknologi Muhammadiyah Jakarta
adetians@stikomcki.ac.id¹, agistia@stikomcki.ac.id², imam.santoso@utmj.ac.id³

ABSTRACT

Sejarah internet Indonesia dimulai pada awal tahun 1990-an. Saat itu jaringan internet di Indonesia lebih dikenal sebagai paguyuban *network*, di mana semangat kerjasama, kekeluargaan dan gotong royong sangat hangat dan terasa di antara para pelakunya. Agak berbeda dengan suasana Internet Indonesia pada perkembangannya kemudian yang terasa lebih komersial dan individual di sebagian aktivitasnya, terutama yang melibatkan perdagangan Internet. Sejak 1988, ada pengguna awal Internet di Indonesia yang memanfaatkan dan untuk mengakses internet. Analisis sentimen terhadap lamabatnya internet diindonesia ini sangat penting untuk masyarakat indonesia. Tanggapan klasifikasi menjadi sentien positif dan negatif menggunakan metode *K-Nearest Neighbor*. Pelabelan data dapat dilakukan dengan manual yang di data pada microsoft excel dengan status positif dan negatif yang di proses oleh rapidminer. Dari 378 opini pengguna youtube yang berhasil dikumpulkan, didapat 121 opini positif dan 253 opini negatif. Kemudian, data tersebut diklasifikasikan menggunakan algoritma *K-Nearrest Neighbor*, selanjutnya dilakukan optimasi dengan menggunakan Rapidminer. Sehingga penelitian yang telah dilakukan mendapatkan hasil akurasi sebesar 75.42% menggunakan *K-Nearrest Neighbor* sedangkan menggunakan Naïve Bayes 60,17%. Ada peningkatan hasil akurasi yang signifikan, membuktikan bahwa penggunaan seleksi *K-Nearrest Neighbor* mempengaruhi untuk meningkatkan hasil akurasi dibandingkan *Naïve bayes*. Oleh karena itu saya membuat jurnal ini menggunakan metode *K-Nearrest Neighbor*

Kata Kunci: Analisis Sentimen, *K-Nearrest Neighbor*, *Rapidminer*.

I. PENDAHULUAN

Teknologi internet merupakan jaringan data yang mampu menghubungkan seluruh dunia, menyamakan batas ruang dan waktu, dan memungkinkan berbagai aktivitas elektronik seperti e-commerce maupun layanan data publik. Dengan beragam kemudahan dan fasilitas yang ditawarkan, internet menjadi elemen yang penting dalam dunia komunikasi yang modern. Dengan semakin luasnya pemanfaatan internet oleh masyarakat, trafik backbone menjadi padat dan kualitas koneksi menjadi tantangan. Penyelenggara akses internet baik penyelenggara jaringan (*network operator*) maupun penyelenggara jasa (*Internet Service Provider*) secara kompetitif menyediakan layanan dengan ragam *Quality of Service (QoS)* untuk trafik jaringan.

Namun sayangnya, akses internet di Indonesia dinilai masih tergolong lambat dan mahal. Hasil pengukuran yang dilakukan Google terhadap

kecepatan membuka web pada komputer desktop dan perangkat mobile di 50 negara pada bulan April 2012,

Indonesia adalah negara dengan kecepatan internet paling lambat jika mengakses lewat desktop dengan rata-rata kecepatan membuka halaman web 20,3 detik,

dibandingkan dengan yang tercepat Slowakia dengan rata-rata kecepatan 3,3 detik. Jika menggunakan perangkat mobile, peringkat Indonesia lebih baik dari 50 negara yang lain dimana di Indonesia rata-rata dibutuhkan 12,9 detik untuk membuka halaman web lewat perangkat mobile, dibandingkan yang tercepat Korea Selatan dengan rata-rata 4,8 detik dan yang terlambat yaitu Uni Emirat Arab dengan rata-rata 26,7. Hasil pengukuran lainnya dilakukan oleh Akamai, perusahaan jaringan delivery content global, yang menyebutkan bahwa kecepatan koneksi internet rata-rata di Indonesia adalah 772 kbps, cukup rendah dibandingkan angka rata-rata kecepatan koneksi

internet global yang mencapai 2,3 Mbps. Penetrasi internet broadband di Indonesia juga terhitung masih sedikit, baru 5% populasi menikmati kecepatan internet sekitar 2 Mbps dan 18% populasi mendapat kecepatan koneksi dikisaran 256 Kbps.

Beberapa justifikasi penyebab akses lambat tersebut muncul seperti jumlah penduduk yang sangat besar lebih dari 237 juta jiwa, kondisi geografis yang sangat luas, maupun kurangnya server didalam negeri. Selain itu, beberapa

penyebab teknis lainnya turut mempengaruhi perbedaan kualitas akses internet yang dirasakan oleh pelanggan antara lain pemilihan providers dan paket yang dilanggan, lokasi akses maupun perangkat akses yang digunakan.

Pada penelitian analisis sentimen sebelumnya, yaitu analisis sentimen mengenai komentar pengguna aplikasi *youtube* dengan menggunakan *K-Nearest Neighbor*, dimana pada penelitian tersebut bertujuan mengukur kualitas hasil analisis menggunakan tanpa *feature selection* dan dengan *feature selection*. Hasil yang didapat dari penelitian tersebut *Cross Validation K-Nearest Neighbor* tanpa *feature selection* adalah 75.42 % untuk *Accuracy* dan 0.593 untuk *AUC*.

Pada penelitian ini, peneliti menggunakan *K-Nearest Neighbor*, serta *Rapidminer* sebagai seleksi fitur yang diterapkan untuk meningkatkan performa klasifikasi teks. Melalui penelitian ini diharapkan bisa mendapatkan hasil yang lebih baik dari penelitian yang sudah dilakukan sebelumnya menggunakan Algoritma *K-Nearest Neighbor*

Pada penelitian analisis sentimen ini, dilakukan untuk melihat sebuah opini seseorang atau masyarakat yang ditujukan kepada lambatnya internet di Indonesia, sehingga dihasilkan opini yang dapat masuk ke kategori opini negatif ataupun opini positif. Dengan adanya penilaian maupun pandangan masyarakat yang disampaikan pada media sosial, maka peneliti berinisiatif membuat penelitian ini, peneliti melihat belum ada penelitian tentang *opinion mining* sebelumnya yang mengangkat permasalahan tersebut, sehingga kedepannya diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif, negatif.

II. LITERATUR DAN METODE

A. Analisis Sentimen.

Sentiment Analysis, atau yang biasa disebut juga *Opinion Mining*, adalah bidang studi yang bertujuan untuk menganalisis opini, sentimen, penilaian,

evaluasi, sikap dan emosi publik terhadap suatu entitas dari produk, pelayanan, suatu permasalahan,

organisasi, peristiwa tertentu, topik yang hangat dibicarakan dan atributnya.

Ada beberapa macam metode yang bisa digunakan pada analisis sentimen, yaitu antara lain, NB (*Naive Bayes*), *Decision Tree*, KNN (*K-Nearest Neighbor*),

Neural Networks dan juga *SVM (Support Vector Machines)*.

B. Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama *YALE (Yet Another Learning Environment)*, dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. RapidMiner didistribusikan di bawah lisensi *AGPL (GNU Affero General Public License)* versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai software open source untuk data mining tidak perlu diragukan lagi karena software ini sudah terkemuka di dunia.

RapidMiner menempati peringkat pertama sebagai Software data mining pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011 RapidMiner menyediakan *GUI (Graphic User Interface)* untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file *XML (Extensible Markup Language)* yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis.

Fitur Rapidminer diantara lain :

1. Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
3. Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.

4. Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi
5. loading), data preprocessing, visualisasi, modelling dan evaluasi
6. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI

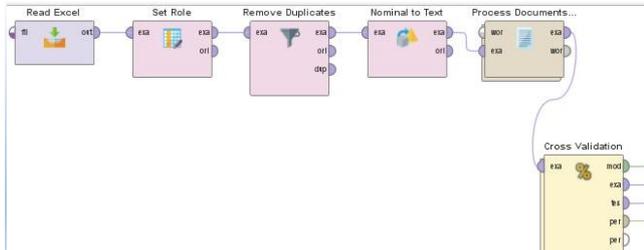
C. Algoritma K-Nearest Neighbor (KNN)

K-nearest neighbors atau knn adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (nearest neighbors). Dengan k merupakan banyaknya tetangga terdekat. *K-nearest neighbors* melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak.

Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik c pada ruang dimensi banyak.

III. METODE

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



Gambar 1. Metode Penelitian



Gambar 2. Proses Dokumentasi From Data

A. Pengumpulan dan Pengolahan Data

Penelitian ini menggunakan data yang didapat dari komentar video *youtube* milik akun yang diposting pada tanggal 18 Desember 2022, berikut *source link*

video

<https://www.youtube.com/watch?v=3yhnmbIVr1u>,

<https://www.youtube.com/watch?v=0yojtbbeeho>,

<https://www.youtube.com/watch?v=usnmyh0Eeme4>

Proses pengambilan data dilakukan secara manual satu

persatu pada tanggal 10 Desember 2022, data yang digunakan dalam penelitian ini sebanyak 378 data yang terdiri dari 121 data *review* positif dan 253 data *review* negatif.

Setelah melakukan pengumpulan data, dilakukan proses pengolahan data atau *preprocessing*, tahapan ini mencakup kegiatan membangun data dan dilanjutkan kegiatan pembersihan data agar dapat dilakukan pengelolaan ketahap selanjutnya. Berikut tahapan *preprocessing* data, antara lain :

1. Remove URL.

Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil pengumpulan data. Biasanya digunakan untuk merujuk pada *source* atau sumber dari suatu berita/informasi.

2. Anotation Removal.

Menghapus tanda Mention (@) beserta teks yang ada dibelakangnya. Pada media sosial biasanya berfungsi untuk memanggil user/pengguna lainnya.

3. Stemming.

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

4. Tokenize.

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

5. Transforms Cases.

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

6. Filter Tokens (By Length).

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

7. Filter Stopword.

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

Preprocessing pertama dilakukan menggunakan dilakukan menggunakan *tools Rapidminer*.

B. Seleksi Fitur dan Algoritma

Peneliti mengusulkan menggunakan metode pemilihan seleksi fitur *rapidminer* yang digunakan untuk meningkatkan akurasi pada pengklasifikasian

algoritma. Algoritma yang akan dijadikan perbandingan atau komparasi pada penelitian ini adalah *K-nearest neighbors* KNN. Peneliti menggunakan K-nearest neighbors sangat baik dalam melakukan proses pengklasifikasian data dan sangat mudah digunakan menggunakan *tools rapidminer*.

C. Evaluasi dan Validasi

Pada tahap pengujian metode, peneliti menggunakan *RapidMiner* dalam melakukan proses eksperimen, data *training* yang digunakan adalah komentar video *youtube* mengenai Lambatnya internet di Indonesia. Kemudian *dataset* dikelompokkan menjadi 2 bagian yaitu opini positif dan negatif.

Peneliti menggunakan *Algoritma KNN* yang hasilnya menjadi performance vector disertai *Confusion Matrix* Kemudian Kurva *ROC* digunakan untuk mengukur nilai *AUC*.

1. Confusion Matrix

Matriks yang menginformasikan hasil prediksi secara keseluruhan dari nilai akurasi dan untuk melihat kinerja pengklasifikasi, yaitu seberapa sering kasus *class X* yang benar diklasifikasikan sebagai *class X* atau kesalahan klasifikasi *class* yang lainnya. Ketika *dataset* hanya memiliki dua kelas, yaitu *class* positif dan *class* negatif, maka dapat dibuatkan seperti gambar dibawah ini :

Tabel 1. *Confusion Matrix*

Correct Classification	Classified As	
	+	-
+	True Positive	False Negative
-	False Positive	True Negative

True positif (TP) merupakan jumlah *record* positif dalam *dataset* yang diklasifikasikan positif.

True negatif (TN) merupakan jumlah *record* negatif dalam *dataset* yang diklasifikasikan negatif. *False* positif.

merupakan jumlah *record* negatif dalam *dataset* yang diklasifikasikan positif. *False* negatif (fn) merupakan jumlah *record* positif dalam *dataset* yang diklasifikasikan negatif.

2. Kurva ROC

Kurva *ROC* menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva *ROC* mengekspresikan *confusion matrix*. *ROC* adalah grafik dua dimensi dengan *false* positif sebagai garis horizontal dan *true* positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan *AUC*.

Kurva *ROC* menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva *ROC* mengekspresikan *confusion matrix*. *ROC* adalah grafik dua dimensi dengan *false* positif sebagai garis horizontal dan *true* positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan *AUC*.

- 0.90 - 1.00 = *Excellent Classification*;
- 0.80 - 0.90 = *Good Classification*;
- 0.70 - 0.80 = *Fair Classification*;
- 0.60 - 0.70 = *Poor Classification*;
- 0.50 - 0.60 = *Failure*

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar video *youtube* mengenai peresmian pabrik mobil esemka, peneliti mengumpulkan 378 data opini pengguna *youtube*. Kemudian data tersebut dikumpulkan dalam format file *xls*, setelah itu data dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif.

Pengelompokan data ini dilakukan oleh beberapa koresponden, sehingga kemudian dihasilkan 121 opini Positif dan 253 opini negatif.

Contoh Tabel Data Excel

Text	status
harus memiliki masukan hadapi kendala yang ada	positif
kami timur merasakan bagaimana lambat jaringan internet untuk kirim foto dengan ukuran kecil perlu waktu lama	negatif
aku rasa akan ada juga dampak cukup besar dalam wisatawan dari lokal pun asing karena itu menjadi salah satu fasilitas semakin kencang jaringan semakin merata internet indonesia aku rasa akan semakin pengunjung datang	positif
indonesia sangat menantang tidak mudah membangun infrastruktur telekomunikasi indonesia tidak seperti brazil afrika selatan	positif

sebagai negara anak benua indonesia memiliki pulau terpencil tidak mudah hubungkan seluruh pulau tantangan alam laut gunung berapi dengan ribuan kelompok bahasa lokal dengan sejuta masalah sulit masih kagum	
semoga indonesia segera ada sumber daya manusia unggul mempunyai teknologi maju secara agar bisa membuat satelit internet sendiri secara mandiri buatan anak bangsa agar tidak lambat bergantung pada negara lain penyebab utama lain adalah perencanaan kota kita indonesia luar biasa buruk	negatif
sehingga para penyewa jasa telekomunikasi juga sulit optimalkan jaringan fiber optik jaringan jaringan menara pemancar negara kita harus akui operator bebani untuk membangun jaringan daerah secara finansial kurang menguntungkan	negatif
indonesia peraturan pelaku penghambat kemajuan maka selalu tertinggal dari negara lain	negatif
semoga dengan isu peningkatan kecepatan internet juga di barengi dengan pemerataan akses internet seluruh sudut pelosok negara	positif
jangan indonesia bagian timur sekitaran jakarta saja masih banyak desa belum teratasi handphone internet	negatif
aku masih bersyukur menjadi warga indonesia	positif

re: 82.98% +/- 4.09% (micro average: 82.96%) (positive class: negatif)

	true positif	true negatif	class prec
itif	58	29	66.67%
atif	63	224	78.05%
all	47.93%	88.54%	

Gambar 2. F_Measure hasil pengumpulan data.

B. Pengolahan Data

Setelah pengumpulan data dan membagi data menjadi komentar positif dan komentar negatif, selanjutnya

dilakukan proses pengolahan data atau *preprocessing*. Berikut ini adalah tahapan *preprocessing* :

1. Remove URL.

Yaitu proses menghilangkan *Uniform Resource Locator* yang ada pada dataset hasil pengumpulan data. Berikut contoh perbedaan sebelum dan sesudah proses.

Tabel 2. Hasil *Remove URL*.

Sebelum	Sesudah
aku rasa video ini kurang adegan makan bakso seperti video https://www.youtube.com/watch?v=mtAyXJInSUI	aku rasa video ini kurang adegan makan bakso

2. Anotation Removal.

Menghapus tanda *Mention* (@) beserta teks yang ada dibelakangnya.

Tabel 4. Hasil *Anotation Removal*.

Sebelum	Sesudah
dari dulu di tonton di ulas di review @indihome tetapi jaringan tetap lambat	dari dulu di tonton di ulas di review tetapi jaringan tetap lambat

3. Stemming.

Tahapan ini berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan, baik awalan ataupun akhiran.

Tabel 5. Hasil *Stemming*.

Sebelum	Sesudah
internet di indonesia paling lambat, tapi kalau korupsi paling cepat	internet di indonesia paling lambat tapi kalau korupsi paling cepat

4. Tokenize.

Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.

5. Transforms Cases.

Mengubah huruf kapital yang masih ada di dataset menjadi huruf - huruf kecil. Hal ini bertujuan agar terjadi keseragaman text pada model klasifikasi dan tidak terjadi kesalahan pada proses *tokenize*.

Tabel 6. Hasil *Transforms Cases*.

Sebelum	Sesudah
luas negara ikut	luas negara ikut
mempengaruhi cepat lambat	mempengaruhi cepat lambat
pemerataan internet. Hal yang seperti ini harus ditangani	pemerataan internet

6. *Filter Tokens (By Length)*.

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

7. *Filter Stopword*.

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

Tabel 7. Hasil *Filter Stopword*.

Sebelum	Sesudah
Semua provider internet indonesia paling lambat. tetapi paling mahal	memang internet Indonesia paling lambat tetapi paling mahal

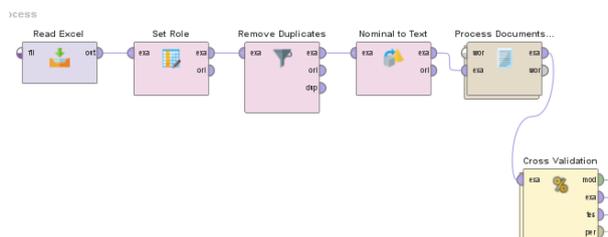
Berikut gambar proses pengolahan data pada *tools rapidminer*.



Gambar 3. Proses Pengolahan Data Rapidminer.

C. *Fitur seleksi Particle Swarm Optimization*.

Peneliti mengusulkan menggunakan *Particle Swarm Optimization* untuk meningkatkan akurasi dari pengklasifikasi Algoritma *K-Nearest Neighbor*. Penelitian ini nantinya menghasilkan akurasi dan nilai *AUC* dengan menggunakan aplikasi *RapidMiner* untuk hasil evaluasi. Pada Read Excel kita masukan format data yang kita sudah buat secara manual.

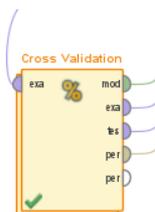


Gambar 5. Proses Pengholaan Data *RapidMiner*.

D. *Proses Validasi Algoritma*.

Selanjutnya adalah penggunaan algoritma *K-Nearest Neighbor* untuk melakukan klasifikasi data yang dihubungkan dengan pengujian *cross validation*

dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak untuk data *testing*.



Gambar 6. Proses *Cross Validation* dengan *Rapidminer*.

Tahapan selanjutnya adalah proses Evaluasi, ialah mengevaluasi kinerja terhadap permodelan dan perhitungan yang telah digunakan dengan menerapkan algoritma *K-Nearest Neighbor* agar dihasilkan peningkatan akurasi pada penghitungan. Untuk mengestimasi performa dari model algoritma yang telah dipilih digunakan *Cross Validation* sehingga dapat menghasilkan nilai akurasi (Proses *Cross Validation* dapat dilihat pada Gambar 6). Dibawah ini adalah model Proses Document Dari data, dan jangan lupa pada filter stopwords kita masukan data dan stem dictionary



Gambar 7. Model Proses Data *RapidMiner*.

E. *Evaluasi*

Penelitian ini menggunakan algoritma *K-Nearest Neighbor* untuk melakukan analisis sentimen. Kemudian dalam menentukan evaluasinya peneliti menggunakan *Accuracy* dan *AUC (Area Under Curve)*.

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan *dataset* sebanyak 378 komentar opini pengguna *Youtube* mengenai video Lambatnya internet di Indonesia, maka hasil Akurasi Algoritma *K-Nearest Neighbor* tanpa menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 75.42 %, sedangkan untuk nilai *AUC* sebesar 0.743.

Berikut ini adalah Tabel *Confusion Matrix* Algoritma *Naive Bayes* tanpa fitur seleksi.

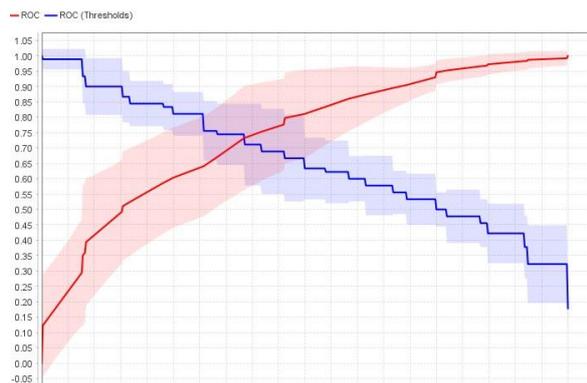
Tabel 8. *Confusion Matrix* Algoritma *NB*.

	True Positive	True Negative	
Pred Positive	58	29	66.67%
Pred Negative	63	224	78.05%
	47.93%	88.54%	

Pada tabel 8 *Confusion Matrix* dapat dilihat, sebanyak 58 data diprediksi *class* positif ternyata sesuai, yaitu masuk kedalam *class* positif, sebanyak 63 data yang diprediksi *class* negatif ternyata termasuk kedalam prediksi *class* positif.

Dan sebanyak 29 data yang diprediksi *class* positif ternyata masuk dalam *class* negatif, kemudian 224 data di prediksi *class* negatif sesuai yaitu termasuk kedalam prediksi *class* negatif.

AUC: 0.743 +/- 0.092 (micro average: 0.743) (positive class: negatif)



Gambar 9. Grafik Area Under Curve (AUC) NB.

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 378 komentar opini pengguna *Youtube* mengenai video lambatnya internet di Indonesia, maka hasil Akurasi Algoritma algoritma *K-Nearest Neighbor*

REFERENSI

1. Kiki Setiawan.(2019). Komparasi metode naive bayes dan support vector machine
2. Listyani Hapsari. (2022). Kecepatan Internet Indonesia Paling Lambat di Asia Tenggara, Apa Penyebabnya.

Diakses pada tanggal 25 Desember 2022 melalui <https://kumparan.com/listyanihapsari171/kecepatan-internet-indonesia-paling-lambat-di-asia-tenggara-apa-penyebabnya-1yRrujdWhgi/1>

dengan menggunakan Fitur Seleksi *Particle Swarm Optimization* yaitu sebesar 75.42 %, sedangkan untuk nilai *AUC* sebesar 0.743.

V. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan, dapat disimpulkan bahwa hasil akurasi dan *AUC* dengan algoritma *K-Nearest Neighbor* bahkan dengan algoritma pengklasifikasian *AUC* nilai keakuratannya dapat dikategorikan sebagai *Excellent Classification*. Begitu juga pengaruh penggunaan fitur seleksi *Particle Swarm Optimization* terhadap algoritma tersebut juga sangat besar, dapat dilihat dari hasil pengujian diatas, terjadi peningkatan nilai akurasi dan *AUC* sangat signifikan. Dan pada penelitian ini juga, hasil akurasi dan *AUC* dengan menggunakan algoritma *K-Nearest Neighbor* dengan dan tanpa menggunakan fitur seleksi *Particle Swarm Optimization*, memiliki hasil yang lebih baik dari penelitian *Naive Bayes* yang saya uji coba sebelumnya.

Untuk mendukung penelitian ini, peneliti akan mengembangkan aplikasi opini dari komentar youtube mengenai lambatnya internet di indonesia untuk mengklasifikasikan opini negatif dan positif menggunakan *Tools Rapidminer*.

menggunakan particle swarm optimization untuk analisis sentimen mobil esemka. Diakses pada tanggal 25 Desember 2022 melalui <http://journal.stmikjayakarta.ac.id/index.php/jisamar>

3. Abdulah Umar. (2021). RapidMiner, Definisi dan Fitur-Fiturnya. Diakses pada tanggal 25 Desember 2022 melalui <https://www.abdumar.com/2021/03/rapidminer-definisi-dan-fitur-fiturnya.html?m=1>