

KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES PADA ANALISIS SENTIMEN FORMULA-E JAKARTA TAHUN 2022

Fadida Zanetti Junaedy¹, Rachmat Hidayat Insani², Imam Santoso³

^{1,2}Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, ³UTM Jakarta
fadida@stikomcki.ac.id¹, rachmat@stikomcki.ac.id², imam.santoso@umtj.ac.id³

ABSTRAK

Kota Jakarta berencana untuk mengadakan acara balap Formula E untuk mempromosikan mobil listrik sebagai kendaraan masa depan. Pandemi Covid-19 yang melanda Jakarta membuat rencana tersebut sempat ditunda. Penundaan tersebut sempat menimbulkan polemik di masyarakat di media sosial karena kondisi Jakarta yang sedang dilanda Covid-19 namun pemerintah kota Jakarta tetap ingin menggelar Formula E dengan membayar uang komitmen kepada pihak penyelenggara yang jumlahnya tidak sedikit. Indonesia juga menjadi negara pertama di Asia Tenggara yang telah menyelenggarakan balapan Formula E di Jakarta, sejak awal rencananya gelaran event ini banyak menimbulkan pro dan kontra bahkan setelah selesai gelarannya pun masih ramai dibicarakan di media sosial. Perlu dilakukan penelitian untuk melihat seberapa besar dampak dari terjadi gelaran event tersebut, diharapkan dengan penelitian ini dapat bermanfaat dalam penelitian mengenai opini masyarakat lainnya. Peneliti menggunakan analisis sentimen karena salah satu teknik yang tepat untuk pengelolaan dataset. Terdapat 1051 opini pengguna youtube yang diambil dari 2 video dan juga twitter mendapatkan 538 opini positif dan 513 opini negatif. Data tersebut diklasifikasikan menggunakan algoritma *Support Vector Machine* dan *Naïve Bayes*. Hasil penelitian yang telah dilakukan mendapatkan hasil akurasi *Naïve Bayes* lebih baik daripada *Support Vector Machine* sebesar 70.57%.

Kata kunci : Analisis Sentimen, FIA, Formula E, Jakarta, Support Vector Machine, Naïve Bayes

ABSTRACT

The city of Jakarta plans to hold a Formula E racing event to promote electric cars as the vehicle of the future. The Covid-19 pandemic that hit Jakarta made the plan postponed. The postponement had caused a polemic in the community on social media because the condition of Jakarta was being hit by Covid-19 but the Jakarta city government still wanted to hold Formula E by paying commitment money to the organiser which was not small. Indonesia is also the first country in Southeast Asia to have held a Formula E race in Jakarta, since the beginning of the plan for this event, it has caused many pros and cons, even after the event was completed, it was still being discussed on social media. It is necessary to conduct research to see how much impact the event has had, it is hoped that this research can be useful in research on other public opinions. Researchers use sentiment analysis because it is one of the right techniques for dataset management. There are 1051 youtube user opinions taken from 2 videos and also twitter getting 538 positive opinions and 513 negative opinions. The data is classified using the Support Vector Machine and Naïve Bayes algorithms. The results of the research that has been done get the results of Naïve Bayes accuracy better than Support Vector Machine by 70.57%.

Keyword : Sentiment Analysis, FIA, Formula E, Jakarta, Support Vector Machine, Naïve Bayes

1. PENDAHULUAN

Formula E atau ABB FIA Formula E World Championship adalah sebuah ajang balap mobil kursi tunggal yang menggunakan mobil listrik[1]. Kota Jakarta resmi ditetapkan menjadi salah satu tuan rumah penyelenggaraan ajang balap Formula E yang telah dilaksanakan pada tanggal 4 Juni 2022 dengan slogan Jakarta E-Prix[2]. Namun sayangnya pelaksanaan Jakarta E-Prix sempat tertunda sebagai akibat dari adanya Pandemi Covid-19.

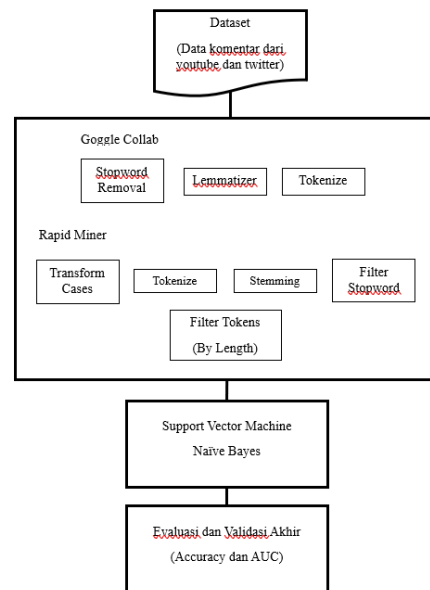
Situs media sosial seperti Youtube dan Twitter, yang sangat populer di Indonesia, telah muncul sebagai tempat bagi masyarakat untuk menyuarakan pendapat mereka tentang kebijakan publik pemerintah. Di Youtube dan Twitter, masyarakat memiliki kecenderungan untuk menyuarakan pendapat mereka tentang kebijakan pemerintah, khususnya kebijakan yang mengatur penyelenggaraan balapan Formula E Jakarta. Platform-platform ini dapat digunakan sebagai sumber analisis sentimen untuk mempelajari pendapat masyarakat umum tentang Formula E Jakarta.

Beberapa penelitian sebelumnya menerapkan Particle Swarm Optimization (PSO) pada metode klasifikasi dan mendapatkan hasil peningkatan akurasi dibandingkan tanpa optimasi dengan PSO[3]. PSO terbukti mampu untuk mengoptimasi akurasi dari metode *Naive Bayes* dan memecahkan masalah pada sentimen analisis.

Particle Swarm Optimization (PSO) adalah algoritma optimasi stokastik berbasis populasi yang dimotivasi oleh kecerdasan perilaku kolektif dari beberapa hewan seperti kawanan burung atau kumpulan ikan[4]. Adapun tujuan dari penelitian ini adalah untuk mengetahui perbedaan nilai akurasi yang dihasilkan oleh *Support Vector Machine* dan *Naive Bayes* tanpa PSO.

2. METODOLOGI

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



Gambar 1. Metode Penelitian

Pengumpulan dan Pengolahan Data

Penelitian ini menggunakan data yang didapat dari komentar 2 video youtube dan unggahan Tweter masyarakat. Video dari akun resmi KOMPASTV Youtube yang diupload pada tanggal 5 Juni 2022, berikut link video tersebut <https://youtu.be/mhAEiHuVFXy> dan tvOneNews yang diupload pada tanggal 2 Juni 2022, berikut link video tersebut https://youtu.be/CTd4xCnSW_I. Proses pengambilan data dari video youtube dilakukan secara manual satu persatu pada tanggal 28 Oktober 2022 sedangkan data yang secara otomatis menggunakan python melalui Google Collab pada tanggal 4 Juni 2023, data yang digunakan dalam penelitian ini sebanyak 1051 data yang terdiri dari 538 data review positif dan 513 data review negative.

Setelah melakukan pengumpulan data, dilakukan proses data atau preprocessing. Selama tahap ini, operasi pembangunan data dan kegiatan pembersihan data yang sedang berlangsung dilakukan untuk mempersiapkan data untuk diproses. Langkah-langkah dalam persiapan data adalah sebagai berikut, termasuk

1. Stopword Removal
Membuang kata-kata yang diabaikan pada sentiment analisis, biasanya yang berupa kata sambung dan kata keterangan.
2. Lemmatizer

Lemmatization adalah proses pengelompokan bentuk infleksi yang berbeda dari kata yang sama. Proses ini digunakan dalam linguistik komputasi, pemrosesan bahasa alami (NLP), dan chatbot.

3. Tokenize
Memecah sekumpulan karakter atau kalimat menjadi sebuah potongan karakter atau kata – kata sesuai dengan kebutuhan, biasa juga disebut tokenisasi.
4. 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.
5. Indonesia Stemming
Proses yang berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan baik awalan maupun akhiran
6. Filter Tokens (By Lenght).
Menghilangkan kata – kata dengan panjang karakter tertentu, biasanya kata yang memiliki hanya 2 karakter tidak memiliki arti.

Preprocessing pertama kali dilakukan secara otomatis menggunakan python melalui Google Collab. Pengolahan dataset selanjutnya dilakukan melalui tools Rapidminer.

Evaluasi dan Validasi

Pada tahap pengujian metode, peneliti menggunakan RapidMiner dalam melakukan proses eksperimen, data training yang digunakan adalah komentar video youtube dan twitter mengenai pro dan kontra penyelenggaraan Formula E Jakarta 2020 Kemudian dataset dikelompokkan menjadi 2 bagian yaitu opini positif dan negatif.

Peneliti menggunakan 10 Fold Cross Validation dalam melakukan validasi, dan untuk mengukur akurasi dilakukan dengan Confusion Matrix, yaitu mengukur akurasi *Support Vector Machine* dan *Naïve Bayes*.

Kemudian Kurva ROC digunakan untuk mengukur nilai AUC.

1. Confussion 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[5]:

Tabel 1. Tabel Confussion Matrix

Correct Classification	Clasified 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 [5].

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

3. LANDASAN TEORI

Analisis Sentimen

Analisis sentimen atau Opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [6].

Setiap hari, internet dibanjiri oleh miliaran data dari berbagai sumber. Sentiment analysis berperan sebagai alat yang dapat menghubungkan seluruh data tersebut. Dengan begitu, perusahaan dapat memperoleh masukan inti dari pengguna atau konsumen secara efisien.

Analisis sentimen merupakan salah satu bidang dari Natural Language Processing (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review. Dengan bantuan sentiment analysis, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.

Data tersebut dapat menjelaskan opini masyarakat mengenai produk, merek, layanan, politik, atau topik lainnya. Perusahaan, pemerintah, maupun bidang lainnya kemudian memanfaatkan data-data tersebut untuk membuat analisis marketing, review produk, umpan-balik produk, dan layanan masyarakat. Guna menghasilkan opini yang dibutuhkan, sentiment analysis tidak hanya harus bisa mengenali opini dari teks. Proses yang juga disebut sebagai opini mining ini juga perlu bekerja dengan mengenali tiga aspek berikut:

- Subjek: topik apa yang sedang dibicarakan.
- Polaritas: apakah opini yang diberikan bersifat positif atau negatif.

- Pemegang opini: seseorang yang mengeluarkan opini tersebut.

Sentiment analysis kemudian akan membedakan teks menjadi dua kategori, yakni fakta dan opini. Fakta merupakan ekspresi objektif mengenai sesuatu. Sementara opini adalah ekspresi subjektif yang menggambarkan sentimen, perasaan, maupun penghargaan terhadap suatu hal.

Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear.

SVM memiliki kelebihan yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Namun *Support Vector Machine* memiliki kekurangan terhadap masalah pemilihan parameter atau fitur yang sesuai. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi[7].

Hyperplane pemisah ideal yang memaksimalkan margin data pelatihan adalah apa yang ingin diidentifikasi oleh SVM. Margin (data pelatihan). Hyperplane terbaik dalam SVM adalah yang berada di tengah-tengah dua kelompok objek dari kelas yang berbeda. Jarak tegak lurus antara hyperplane dan objek terdekat disebut margin, dan memaksimalkannya identik dengan menemukan hyperplane yang optimal. Support vector adalah objek terdekat di kejauhan. Jarak tegak lurus antara vektor pendukung dari dua kelas adalah bagaimana beberapa sumber mendefinisikan margin. Karena memberikan nilai margin tertinggi, maka dianggap sebagai hyperplane terbaik[8].

Naïve Bayes

Algoritma *Naive Bayes* merupakan salah satu metode klasifikasi yang menetapkan teorema Bayes yang berdasarkan nilai probabilitas. apabila masih ada 2 buah

kejadian A dan kejadian B dalam suatu syarat maka nilai probabilitas kejadian A terhadap kejadian B bisa dirumuskan menggunakan persamaan 4[9].

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

Dimana $p(A|B)$ merupakan conditional probability kejadian A sehingga mengakibatkan perubahan kejadian B dan $p(B|A)$ merupakan conditional probability kejadian B sehingga mengakibatkan perubahan kejadian A. Sedangkan $p(A)$ dan $p(B)$ merupakan probabilitas kejadian A dan B. Jika A merupakan kelompok dan B merupakan data menurut persamaan 4 algoritma Naïve Bayes pada mengklasifikasikan sebuah data bisa ditulis misalnya persamaan 5[9].

$$\frac{p(kategori|dokumen)}{p(Dokumen)} = \frac{p(dokumen|kategori) \times p(kategori)}{p(Dokumen)}$$

Kategori direpresentasikan seperti c_j , dimana c_j ialah teks yang akan diklasifikasikan dan data direpresentasikan seperti d . Jadi $p(kategori)$ yang ialah probabilitas menurut kategori teks bisa ditulis $p(c_j)$ dan $p(data)$ yang ialah probabilitas data ditulis sebagai $p(d)$. Jadi persamaan 6 berubah menjadi seperti berikut[9].

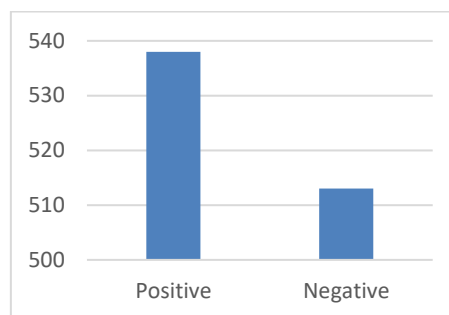
$$p(c_j|d) = \frac{p(d|c_j) \times p(c_j)}{p(d)}$$

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Hasil pengumpulan data yang didapat dari komentar video youtube dan twitter mengenai gelaran balap formula e di Jakarta, peneliti mengumpulkan 1051 data opini pengguna Youtube dan Twitter. Kemudian data tersebut dikumpulkan dalam format file xls, setelah itu data dikelompokkan menjadi 2 bagian yaitu data positif dan data negatif.

Pengelompokan data dilakukan oleh beberapa koresponden, sehingga menghasilkan data 538 opini positif dan 513 opini negatif.



Gambar 1. Statistik Hasil Pengumpulan Data

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. Stopword Removal

Membuang kata-kata yang diabaikan pada sentiment analisis, biasanya yang berupa kata sambung dan kata keterangan.

Tabel 2. Hasil Stopword Removal

Sebelum	Sesudah
Gak usah nonton ticket dah habis	nonton ticket dah habis

2. Lemmatizer

proses pengelompokan bentuk infleksi yang berbeda dari kata yang sama.

3. Tokenize

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

4. Transform 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 3. Hasil Transform Cases

Sebelum	Sesudah
Sebagai warga negara Indonesia sangat bangga dengan even Formula E	Sebagai warga negara Indonesia sangat bangga dengan even formula e

5. Indonesia Stemming

Proses yang berfungsi mengubah kata perkata menjadi sebuah kata dasar, dengan cara menghilangkan imbuhan baik awalan maupun akhiran.

Tabel 4. Hasil Stemming

Sebelum	Sesudah
Kegiatan formula e ini sangat penting untuk disukseskan bersama ini aku suka dan suka sekali	giat formula e ini sangat penting untuk sukses sama ini aku suka dan suka kali

6. Filter Tokens (By Length)

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

Berikut gambar proses pengolahan data pada tools rapidminer.



Gambar 2. Proses Pengolahan Data Rapidminer

Metode pembobotan kata, yang mencoba menentukan bobot pada setiap kata yang akan digunakan sebagai fitur, akan digunakan untuk mengubah data prapemrosesan dari kata

menjadi angka. Semakin banyak teks yang perlu diproses, semakin banyak fitur yang akan digunakan. Ada dua langkah dalam proses pada tahap ini: TF (Term Frequency) dan IDF (Inverse Document Frequency).

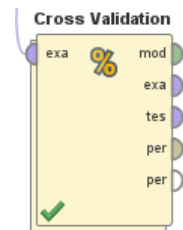
TF mengukur seberapa sering setiap kata muncul dalam sebuah dokumen; semakin sering sebuah istilah muncul dalam sebuah teks, semakin tinggi nilai TF-nya. Df adalah jumlah nilai dokumen untuk setiap kata, yang berbanding terbalik, yang berarti bahwa nilai IDF dari sebuah kata yang hanya sesekali muncul dalam dokumen lebih tinggi daripada nilai istilah yang sering muncul.



Gambar 3. Proses Tf Idf

Proses Validasi Algoritma

Selanjutnya adalah penggunaan algoritma *Naive Bayes* dan *Support Vector Machine* untuk melakukan klasifikasi data yang dihubungkan dengan pengujian 10-fold cross validation dimana proses ini untuk mengevaluasi proses kerja algoritma tersebut dengan membagi data secara acak ke dalam 10 fold untuk mendapatkan 10 data yang sama, kemudian data tersebut digunakan 9 fold untuk data training dan 1 fold untuk data testing.



Gambar 4. Proses Cross Validation dengan Rapidminer

Langkah selanjutnya adalah prosedur evaluasi, yang menilai seberapa baik perhitungan dan pemodelan yang dilakukan

setelah menambahkan metode *Naive Bayes* dan *Support Vector Machine*. Validasi Silang digunakan untuk memberikan angka akurasi, yang dapat digunakan untuk menentukan seberapa baik kinerja model algoritme yang dipilih (prosedur Validasi Silang ditunjukkan pada Gambar 4). Model pengujian *Support Vector Machine* dari alat Rapidminer ditunjukkan di bawah ini.



Gambar 5. Model Pengujian *Support Vector Machine* dari di alat Rapidminer

Dibawah ini adalah model pengujian *Naive Bayes* dengan menggunakan tools Rapidminer



Gambar 6. Model Pengujian *Naive Bayes* dari di alat Rapidminer

Evaluasi

Dalam penelitian ini, analisis sentimen dilakukan dengan menggunakan teknik *Support Vector Machine* dan *Naive Bayes*. Kemudian, peneliti menggunakan Akurasi dan AUC (Area Under Curve) untuk menghitung evaluasi. Selain itu, peneliti akan membandingkan hasil akurasi *Naive Bayes* dan *Support Vector Machine*.

Dari tahapan - tahapan pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 1051 komentar opini pengguna Youtube dan Twitter mengenai pengadaan formula e di jakarta, maka hasil Akurasi Algoritma *Support Vector Machine* yaitu sebesar 65.33%, sedangkan untuk nilai AUC sebesar 0.839.

Berikut ini adalah Tabel Confusion Matrix Algoritma *Support Vector Machine*.

Tabel 5. Confusion Matrix Algoritma *Support Vector Machine*

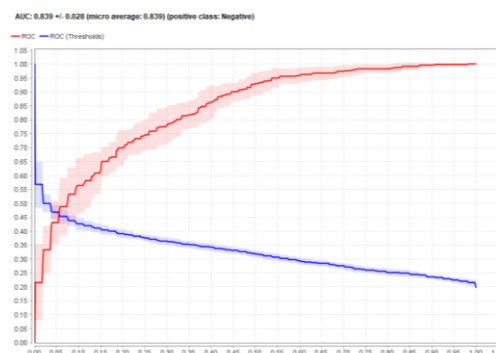
accuracy : 65.33% +/-2.34%(micro average: 65.33%)

	True Positive	True Negative	
Pred Positive	512	349	59.89%
Pred Negative	14	163	92.09%

	97.38%	31.84%	
--	--------	--------	--

Pada tabel 5 Confusion Matrix dapat dilihat, sebanyak 513 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class positif, sebanyak 14 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 349 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 163 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 7. Grafik Area Under Curve (AUC) *Support Vector Machine*

Berikutnya adalah pengujian yang sudah dilakukan dengan menggunakan dataset sebanyak 1051 komentar opini pengguna Youtube dan Twitter mengenai pengadaan formula e di jakarta, maka hasil Akurasi *Naive Bayes Machine* yaitu sebesar 70.57 %, sedangkan untuk nilai AUC sebesar 0.523.

Dibawah ini adalah Tabel Confusion Matrix Algoritma *Naive Bayes*.

Tabel 6. Confusion Matrix Algoritma *Naive Bayes*

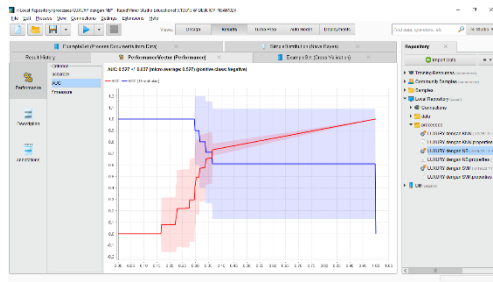
accuracy : 70.57% +/-4.85%(micro average: 70.58%)

	True Positive	True Negative	
Pred Positive	385	158	70.90%
Pred Negative	150	354	70.24%
	71.96%	69.14%	

Pada tabel 6 Confusion Matrix dapat dilihat, sebanyak 385 data diprediksi class positif ternyata sesuai, yaitu masuk kedalam class

positif, sebanyak 150 data yang diprediksi class negatif ternyata termasuk kedalam prediksi class positif.

Dan sebanyak 158 data yang diprediksi class positif ternyata masuk dalam class negatif, kemudian 354 data di prediksi class negatif sesuai yaitu termasuk kedalam prediksi class negatif.



Gambar 8. Grafik Area Under Curve (AUC) Naïve Bayes

5. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan, bisa disimpulkan bahwa hasil akurasi Naïve Bayes lebih baik dari algoritma Support Vector Machine dengan akurasi sebesar 70.57%, tetapi hasil pengujian pengklasifikasian AUC algoritma Support Vector Machine lebih baik dari algoritma Naïve Bayes dengan nilai keakuratannya dapat dikategorikan sebagai Good Classification.

DAFTAR PUSTAKA

- [1] F. Del Pero, L. Berzi, C. Antonia Dattilo, dan M. Delogu. (2021). *Environmental sustainability analysis of Formula-E electric motor*. Proc. Inst. Mech. Eng.
- [2] R. Tanjung. (2021). *Instruksi Gubernur DKI Jakarta Tentang Penyelenggaraan Balap Formula E Dalam Tinjauan Siyasa Islam*. Al Ahkam,
- [3] Jatmiko, H. B., Kurniadi, N. T., & Maulana, D. (2022). *Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta*. Journal Automation Computer Information System.

- [4] Hossain, S. I., Akhand, M. A. H., Shuvo, M. I. R., Siddique, N., & Adeli, H. (2019). *Optimization of university course scheduling problem using particle swarm optimization with selective search*. Expert systems with applications.

- [5] Setiawan, K., Rahmatullah, B., Burhanuddin, B., Paryanti, A. B., & Fauzi, F. (2020). *Komparasi Metode Naive Bayes Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil Esemka*. JISAMAR (Journal of Information System, Applied, Management, Accounting and Research).

- [6] Liu, B. (2010). *Handbook of Natural Language Processing, chapter Sentiment Analysis and Analysis, 2nd Edition*. Chapman & Hall / CRC Press, New York

- [7] Z. Hao, L. Shaohong, and S. Jinping. (2011). *Unit Model of Binary SVM with DS Output and its Application in Multi-class SVM*. Fourth International Symposium on Computational Intelligence and Design.

- [8] Z. Yan. (2010). *A SVM model for data mining and knowledge discovering of mine water disasters*. 8th World Congress on Intelligent Control and Automation.

- [9] Azhari, M., Situmorang, Z. and Rosnelly, R., (2021). *Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes*. JURNAL MEDIA INFORMATIKA BUDIDARMA.