

Perbandingan Metode Support Vector Machine dan Naive Bayes Classifier untuk Klasifikasi Survey Kepuasan Masyarakat terhadap Layanan UPT BLK Surabaya

Achmad Fadli Rafi¹, Paramitha Nerisafitra²

^{1,2}Teknik Informatika/ Teknik Informatika, Universitas Negeri Surabaya

¹achmad.18067@mhs.unesa.ac.id

²paramithanerisafitra@unesa.ac.id

Abstrak— Teknologi akan terus menerus mengalami perkembangan seiring dengan perkembangan zaman. Hal ini mengakibatkan keperluan manusia untuk dapat mengakses informasi secara cepat, akurat, dan se-efisien mungkin. Salah satu yang mengalami perkembangan adalah dalam pengolahan data. Klasifikasi data adalah teknik untuk membagi data menjadi beberapa kelas menggunakan fitur-fitur yang didapat dari data. *Support Vector Machine* dan *Naive Bayes Classifier* adalah contoh algoritma yang umum digunakan dalam proses klasifikasi.

Menurut penelitian yang terdahulu, perbandingan algoritma klasifikasi hanya menggunakan 4 metrik penilaian, yaitu: nilai akurasi, nilai *precision*, nilai *recall*, dan nilai *f1-score*, namun belum banyak penelitian yang juga menyertakan waktu eksekusi dari program tersebut. Dari latar belakang di atas, muncul pertanyaan bagaimana perbandingan performa kinerja dan waktu eksekusi kedua algoritma tersebut.

Proses penelitian dimulai dengan mengumpulkan dan mengolah data untuk dapat digunakan dalam proses klasifikasi. Setelah dataset dilakukan pra-pemrosesan, data dapat dibagi menjadi 3 untuk pelatihan, validasi, dan pengujian. Dalam 3 proses ini akan juga dihitung waktu eksekusi dari program. Berdasarkan hasil penelitian, Algoritma SVM unggul dalam 4 metrik dan waktu eksekusi dibandingkan dengan algoritma NBC.

Kata Kunci— Klasifikasi, Support Vector Machine, Naive Bayes Classifier, Confusion Matrix, Indeks Kepuasan Masyarakat.

I. PENDAHULUAN

Teknologi akan terus menerus mengalami perkembangan seiring dengan perkembangan zaman. Hal ini mengakibatkan keperluan manusia untuk dapat mengakses informasi secara cepat, akurat dan se-efisien mungkin. Salah satu yang mengalami perkembangan adalah dalam pengumpulan data. Angket adalah salah satu metode yang dapat dilakukan untuk mendapatkan suatu data. Menurut KBBI Versi V (2016), "Angket adalah daftar pertanyaan tertulis mengenai masalah tertentu dengan ruang jawaban bagi setiap pertanyaan". Biasanya dalam bentuk kertas yang berisi pertanyaan yang dibagikan ke narasumber atau dalam bentuk digital seperti Google Form yang dibagikan link-nya. Data dari angket kemudian akan diolah sebagai demikian hingga dapat diinterpretasikan oleh pembaca. Dari pengolahan data, terdapat banyak perkembangan metode untuk membagi atau memisahkan data menjadi bagian-bagian. Hal ini disebut klasifikasi data.

Klasifikasi data adalah teknik untuk membagi data menjadi beberapa kelas menggunakan fitur-fitur yang didapat dari data. Salah satu dari metode ini adalah Support Vector Machine yang merupakan salah satu algoritma yang digunakan untuk pengujian yang singkat. Support Vector Machine akan membuat suatu hyperplane untuk membagi kelas dari dataset. Sedangkan Naive Bayes Classifiers adalah tipe klasifikasi probabilitas yang mengasumsikan bahwa semua fitur yang membangun kelas itu independen. Jadi, Naive Bayes Classifiers mengasumsikan bahwa informasi tiap kelas yang didapat dari masing-masing variabel tidak berhubungan terhadap variabel lainnya. Masing-masing algoritma tersebut adalah contoh dari klasifikasi diskriminatif (SVM) dan generatif (NBC).

Menurut penelitian terdahulu, perbandingan algoritma klasifikasi hanya menggunakan 4 metrik penilaian, yaitu: nilai akurasi, nilai *precision*, nilai *recall*, dan nilai *f1-score*, namun belum banyak penelitian yang juga menyertakan waktu eksekusi dari program tersebut. Dari latar belakang di atas, muncul pertanyaan bagaimana perbandingan performa kinerja dan waktu eksekusi algoritma Support Vector Machine dibandingkan dengan Naive Bayes Classifiers untuk data Survey Kepuasan Masyarakat terhadap layanan UPT BLK Surabaya.

II. METODE PENELITIAN

A. Jenis Penelitian

Jenis Penelitian yang akan dilakukan adalah penelitian yang menggunakan pendekatan kuantitatif. Menurut Bryman(2012), penelitian kuantitatif adalah suatu strategi penelitian yang memfokuskan dalam kuantifikasi pengumpulan dan pengolahan data. Penelitian ini dibentuk dari pendekatan deduktif yang menekankan pengujian teori, yang dibentuk dari filosofi empiris dan positifis.

Bentuk penelitian atau desain penelitiannya adalah eksperimen. Menurut Prastowo(2011), desain penelitian eksperimen adalah metode penelitian yang berupaya mengkaji hubungan kualitas suatu kejadian dengan melakukan manipulasi terhadap objek penelitian, selain juga melakukan kontrol.

B. Populasi dan Sampel Penilaian

Populasi adalah keseluruhan subyek penelitian. Menurut Sugiyono(2013), Populasi merupakan wilayah generalisasi yang terdiri atas obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulan. Populasi dari penelitian ini adalah peserta pendidikan keahlian di UPT BLK Surabaya periode 2022. Total populasi dari data adalah 347 dengan rincian 250 penilaian sangat baik, 54 penilaian baik, 37 penilaian kurang baik, dan 6 penilaian tidak baik.

Sampel adalah bagian dari populasi. Teknik sampel yang dilakukan adalah teknik sampel jenuh. Menurut Sugiyono(2019), Teknik sampel jenuh adalah teknik pemilihan sampel apabila semua anggota populasi dijadikan sampel. Teknik ini dipilih disebabkan adanya prosedur untuk para peserta diwajibkan untuk mengisi kuesioner yang diberikan oleh pihak UPT BLK Surabaya setelah menerima pendidikan keahlian.

C. Instrumen Penelitian

Instrumen penelitian adalah suatu alat yang digunakan untuk mengukur fenomena alam maupun sosial yang diamati. Dalam penelitian ini, digunakan instrumen penelitian non tes berupa angket. Berikut adalah ukuran pernyataan yang digunakan untuk diambil datanya.

TABEL I
INSTRUMEN PERNYATAAN

Instrumen Pernyataan	Jumlah Item
Pendapat Saudara tentang kesesuaian persyaratan pelayanan dengan jenis pelayanannya.	1
Pemahaman Saudara tentang kemudahan prosedur pelayanan di unit ini	1
Pendapat Saudara tentang kecepatan waktu dalam memberikan pelayanan	1
Pendapat Saudara tentang kewajaran biaya/tarif dalam pelayanan	1
Pendapat Saudara tentang kesesuaian produk pelayanan antara yang tercantum dalam standar pelayanan dengan hasil yang diberikan	1
Pendapat Saudara tentang kompetensi/kemampuan petugas dalam pelayanan	1
Pendapat saudara perilaku petugas dalam pelayanan terkait kesopanan dan keramahan	1
Pendapat Saudara tentang kualitas sarana dan prasarana	1
Pendapat Saudara tentang penanganan pengaduan pengguna layanan	1

Dari instrumen pernyataan diatas, akan digunakan skala likert sebagai skala pengukuran. Skala likert adalah suatu penghitungan dengan pendekatan psikometrik untuk mengetahui pendapat responden terhadap pernyataan yang diberi oleh pembuat angket. Dalam penelitian ini, skala likert yang digunakan adalah skala likert yang memiliki 4 poin untuk menghindari pendapat netral dari responden.

D. Teknik Pengumpulan Data

Menurut Sugiyono (2019), teknik pengumpulan data adalah langkah yang paling utama dalam penelitian karena tujuan utama penelitian adalah mendapatkan data. Berikut pengumpulan data yang telah dilakukan.

1) Penyebaran Angket

Cara penyebaran angket ini adalah melalui Google Form dan membagikannya kepada penerima pendidikan keahlian setelah mengikuti kelas hingga selesai. Untuk beberapa kelas yang masih belum sempat menerima link Google Form, akan diberikan angket kertas yang dapat diisi setelah menerima pendidikan keahlian.

Gbr. 1 Angket digital menggunakan layanan Google Form yang telah disebarkan.

Responden akan diberikan 9 pernyataan yang akan dijawab menggunakan skala likert dengan empat nilai. Berikut adalah skala yang digunakan.

TABEL III
SKALA LIKERT

Jawaban	Skor
Sangat Setuju	1
Setuju	2
Kurang Setuju	3
Tidak Setuju	4

Angket yang telah didapat akan dikumpulkan dari berbagai kelas dan akan dijadikan menjadi satu dataset. Berikut adalah rincian data dari hasil angket.

TABEL IIIII
JUMLAH RESPONDEN

Label	Jumlah
Sangat Setuju	250
Setuju	54
Kurang Setuju	37
Tidak Setuju	6
Total	347

E. Teknik Analisis Data

Menurut Sugiyono (2019), analisis data adalah kegiatan yang analisis yang dilakukan oleh peneliti setelah data-data dari responden dan/atau sumber data lain telah didapatkan dalam penelitian kuantitatif. Analisis ini adalah proses pengelompokan, tabulasi/pemrosesan, dan penyajian data berdasarkan variabel yang telah didapatkan dan perlakuan perhitungan untuk menjawab pertanyaan dan menguji hipotesis yang telah diajukan di awal penelitian.

Adapun teknik yang digunakan untuk analisis data adalah analisis statistik deskriptif. Hal ini berguna untuk mengetahui jumlah populasi, jumlah dataset yang terbagi untuk dilatih dan diuji, serta kinerja dari algoritma yang digunakan.

Setelah data dikumpulkan, angket akan diberi penilaian untuk menjadi label nilai pelayanan. Penghitungan nilai tersebut adalah sebagai berikut:

$$SKM = \frac{\text{Total dari nilai persepsi per unsur}}{\text{Total unsur yang terisi}} \times 25 \quad (1)$$

Dari perhitungan nilai di atas, dapat dihitung label kinerja unit pelayanan. Berikut adalah Tabel konversi unit kerja pelayanan.

TABEL IVV
KONVERSI PENILAIAN KINERJA UNIT PELAYANAN

Nilai Persepsi	Nilai Interval	Nilai Internal Konversi	Mutu Pelayanan	Kinerja Unit Pelayanan
1	1.00 – 2.59	25.00 – 64.99	D	Tidak Baik
2	2.60 – 3.06	65.00 – 76.60	C	Kurang Baik
3	3.06 – 3.52	76.61 – 88.30	B	Baik
4	5.53 – 4.00	88.31 - 100	A	Sangat Baik

Setelah data dikumpulkan, data dapat diproses dengan hanya menyisakan 9 nilai persepsi per unsur dan label. Berikut adalah Tabel setelah diproses.

TABEL V
SAMPEL DATA UNSUR PENILAIAN

No	Unsur Penilaian									Label Pelayanan
	1	2	3	4	5	6	7	8	9	
1	3	3	3	4	3	3	2	3	2	Kurang Baik
2	4	4	4	4	4	4	4	4	4	Sangat Baik

No	Unsur Penilaian									Label Pelayanan
	1	2	3	4	5	6	7	8	9	
3	4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	4	4	4	4	4	4	3	4	Sangat Baik
5	4	3	3	4	3	4	3	3	4	Baik
6	4	3	4	3	3	4	4	3	4	Sangat Baik
7	4	4	4	4	4	4	4	4	4	Sangat Baik
8	4	4	4	4	4	4	4	4	4	Sangat Baik
9	4	3	3	4	4	4	4	3	4	Sangat Baik
10	4	4	4	4	4	4	4	4	4	Sangat Baik

Dari data di atas, data akan dibagi menjadi 3 dengan perbandingan 7 : 1.5 : 1.5 untuk dilatih, divalidasi dan diuji menggunakan dua algoritma yang ditentukan yaitu algoritma *Support Vector Machine* dan *Naïve Bayes Classifier*.

Data kemudian akan dilatih dan diuji menggunakan teknik confusion matrix untuk mengetahui kinerja dari masing-masing algoritma. Selain itu, semua tahap pelatihan, validasi, dan pengujian akan dihitung waktunya untuk mengetahui efisiensi dari algoritma.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Pada bab ini akan dilakukan penelitian dan pembahasan sesuai dengan rancangan penelitian yang telah dibuat pada bab III. Dalam penelitian ini digunakan Google Colaboratory sebagai dalam browser Microsoft Edge, Google Chrome, dan Mozilla Firefox. Adapun proses tahap penelitian adalah sebagai berikut.

1) *Persiapan File CSV*

Dalam langkah ini, dilakukan beberapa operasi untuk persiapan dataframe yang akan dilakukan penelitian. Hal ini dilakukan untuk mempermudah dokumentasi. Dalam file notebook ini, hanya akan dilakukan persiapan dataframe, untuk pengolahan data dilakukan di notebook lainnya. Modul dan *library* yang digunakan adalah *pandas* dan *drive* dari *google.colab*. *Pandas* digunakan untuk operasi dataframe. *Drive* digunakan untuk membaca dan menulis file dari atau ke dalam layanan *drive* yang disediakan oleh Google. Dalam langkah ini juga akan didefinisikan kriteria penilaian sesuai dengan Peraturan Menteri PAN-RB Republik Indonesia No. 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat di Unit Penyelenggara Pelayanan Publik sesuai dengan tabel iv.

2) Load CSV dan Label Penilaian

Dari hasil pengumpulan data dengan kuesioner yang disebar melalui platform *Google Forms*, didapatkan responden dengan jumlah 347 peserta. Dari data yang didapat ini, akan diambil 9 fitur yang diperlukan sekaligus memperlakukan penilaian untuk memberi label sebagai pengarah dalam proses klasifikasi. Setelah library dan modul diimpor, dapat dilakukan pemanggilan file CSV hasil ekspor dari *google form*.

	1. Bagaimana pendapat Saudara tentang kesesuaian pelayanan dengan jenis pelayannya.	2. Bagaimana pemahaman Saudara tentang kemudahan prosedur pelayanan di unit ini	3. Bagaimana pendapat Saudara tentang kecepatan waktu dalam memberikan pelayanan	4. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	5. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	6. Bagaimana pendapat Saudara tentang kualitas pelayanan yang diberikan
0	4	4	4	4	4	4
1	4	4	4	4	4	4
2	4	4	4	4	4	4

Gbr. 2 Impor file csv dari ekspor google form menjadi dataframe.

Dari Gambar di atas dapat dilihat bahwa banyak fitur yang tidak relevan dari penelitian ini. Dari 18 Fitur yang ada, hanya diperlukan 9 fitur.

	1. Bagaimana pendapat Saudara tentang kesesuaian pelayanan dengan jenis pelayannya.	2. Bagaimana pemahaman Saudara tentang kemudahan prosedur pelayanan di unit ini	3. Bagaimana pendapat Saudara tentang kecepatan waktu dalam memberikan pelayanan	4. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	5. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	6. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	7. Bagaimana pendapat Saudara tentang kesesuaian dan keramahan	8. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	9. Bagaimana pendapat Saudara tentang kualitas pelayanan yang diberikan
0	4	4	4	4	4	4	4	4	
1	3	3	3	4	3	4	3	3	
2	4	4	4	4	4	4	4	4	
3	3	2	4	3	3	3	3	3	
4	4	4	4	4	4	4	4	4	

Gbr.3 Dataframe dengan 9 fitur.

Dataframe yang dikeluarkan dari proses di atas, akan kemudian diperlakukan perhitungan nilai pelayanan sesuai dengan Peraturan Menteri PAN-RB Republik Indonesia No. 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat di Unit Penyelenggara Pelayanan Publik. Setelah diperlakukan perhitungan, akan digabungkan dengan *Dataframe* 9 fitur di atas dengan label "Label Penilaian".

1. Bagaimana pendapat Saudara tentang kesesuaian pelayanan dengan jenis pelayannya.	2. Bagaimana pemahaman Saudara tentang kemudahan prosedur pelayanan di unit ini	3. Bagaimana pendapat Saudara tentang kecepatan waktu dalam memberikan pelayanan	4. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	5. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	6. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	7. Bagaimana pendapat Saudara tentang kesesuaian dan keramahan	8. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	9. Bagaimana pendapat Saudara tentang kualitas pelayanan yang diberikan	Label Penilaian
4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
3	3	3	4	3	3	2	3	3	Kurang Baik
3	3	3	4	4	4	4	3	3	Baik
3	3	2	4	3	2	1	3	3	Kurang Baik
3	4	3	4	3	4	2	4	4	Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
3	3	3	4	2	3	4	3	2	Kurang Baik
4	4	4	4	4	4	4	4	4	Sangat Baik
4	4	3	4	4	3	4	3	4	Sangat Baik
3	4	4	4	4	4	4	4	3	Sangat Baik
4	4	4	4	4	4	4	4	4	Sangat Baik

Gbr.4 Dataframe dengan Label Penilaian

Setelah dilakukan penilaian, dapat diketahui bahwa terdapat 347 responden dengan rincian: 250 "Sangat Baik", 54 "Baik", 37 "Kurang Baik", dan 6 "Tidak Baik". Berikut adalah Gambar diagram distribusi penilaian responden. Tabel dataframe di atas kemudian diekspor untuk diperlakukan tahap penelitian lebih lanjut. File keluaran csv kemudian dapat digunakan dalam proses selanjutnya di dalam notebook perbandingan algoritma klasifikasi.

3) Impor Library dan Modul, Pemanggilan File CSV

Mulai dari langkah ini, akan dibuka *notebook* baru untuk penelitian perbandingan performa algoritma klasifikasi *Support Vector Machine* dan *Naive Bayes Classifier*. Seperti pada langkah sebelumnya, modul dan *library* akan diimpor. Modul yang digunakan adalah *drive* dari *google.colab*, *pandas*, *metrics* dari *sklearn*, *SVC* dari *sklearn.svm*, *GaussianNB* dari *sklearn.naive_bayes*, *train_test_split* dari *sklearn.model_selection*, dan *time*.

4) Persiapan Dataframe

Dalam langkah ini, dataframe akan kembali dibagi menjadi dataframe X yang berisi kolom-kolom penilaian dan dataframe y yang berisi label penilaian.

5) Train, Test, Split

Tabel dataframe yang telah dipisahkan, kemudian akan dilakukan pembagian untuk dilakukan pelatihan, validasi dan pengujian. Dalam proses ini, data akan dibagi secara acak menjadi tiga dengan rasio pelatihan : validasi : pengujian adalah 7 : 1.5 : 1.5 menggunakan modul *test_train_split* dari *library sklearn.model_selection*.

Proses pembagian ini akan menggunakan proses pembagian berstrata acak. Teknik ini digunakan untuk memastikan tiap sub-kelompok terutama "Tidak Baik" tetap terwakilkan dalam tiap bagian. Setelah dilakukan pembagian, akan didapatkan 243 data train, 52 data validation, dan 52 data test.

	1. Bagaimana pendapat Saudara tentang kesesuaian pelayanan dengan jenis pelayannya.	2. Bagaimana pemahaman Saudara tentang kemudahan prosedur pelayanan di unit ini	3. Bagaimana pendapat Saudara tentang kecepatan waktu dalam memberikan pelayanan	4. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	5. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	6. Bagaimana pendapat Saudara tentang kejelasan biaya/tarif dalam pelayanan	7. Bagaimana pendapat Saudara tentang kesesuaian dan keramahan	8. Bagaimana pendapat Saudara tentang kesesuaian produk pelayanan yang tercantum dalam standar dengan hasil yang diberikan	9. Bagaimana pendapat Saudara tentang kualitas pelayanan yang diberikan
8	4	4	4	4	4	4	4	4	
163	3	3	3	4	3	4	3	3	
71	4	4	4	4	4	4	4	4	
328	3	2	4	3	2	3	3	3	
86	4	4	4	4	4	4	4	4	

Gbr.5 Dataframe X_train.

Dari Gambar Tabel di atas, dapat dilihat pada nomor record yang ditampilkan adalah acak. *Dataframe* train, valid, dan test kemudian dapat disatukan fitur dan labelnya dan diekspor untuk dokumentasi.

Setelah data dibagi menjadi tiga bagian, data training digunakan untuk proses pelatihan kedua algoritma klasifikasi, data validasi untuk mengecek algoritma, dan data test untuk menguji kinerja kedua algoritma klasifikasi

6) Pelatihan Data, Validasi Data, dan Pengujian Data

Dalam langkah ini, dilakukan pelatihan, validasi, dan pengujian algoritma terhadap dataframe-dataframe yang telah dibagi. Dalam setiap langkah klasifikasi dilakukan penghitungan waktu pengerjaan untuk dievaluasi.

Dataframe training digunakan sebagai data penyocokan(fitting) kedua algoritma. Dalam penelitian ini, terdapat dua fase dalam tahap ini yaitu, pembuatan model algoritma dan penyocokkan data. Parameter-parameter yang digunakan adalah kernel linear untuk algoritma Support Vector Machine dan variance smoothing bernilai 1×10^9 untuk algoritma Naïve Bayes Classifier. Kedua algoritma dicocokkan menggunakan dataframe X_train terhadap y_train. Dalam langkah ini juga akan dihitung waktu eksekusi proses training atau fitting.

Model klasifikasi kedua algoritma kemudian dapat divalidasi menggunakan dataframe valid. Model kemudian divalidasi menggunakan nilai akurasi kedua model tersebut.

Setelah divalidasi, model klasifikasi dapat dilakukan proses pengujian. Model klasifikasi diuji menggunakan dataframe test. Model kemudian dievaluasi menggunakan metode confusion matrix dan penghitungan nilai akurasi, presisi, recall, dan f1-score. Untuk penghitungan nilai presisi, recall dan f1-score digunakan beratan makro.

B. Pembahasan

Dalam proses klasifikasi, data yang digunakan adalah data peserta pelatihan di UPT BLK Surabaya. Data akan dikonversi menggunakan nilai sebagai berikut.

TABEL VI
DISTRIBUSI RANGE PENILAIAN

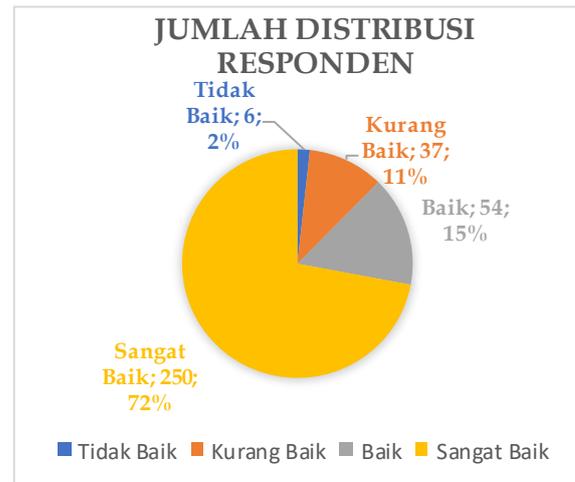
Label	Range Bawah	Range Atas	Range Delta
Sangat Baik	88,31	100	11,69
Baik	76,61	88,3	11,69
Kurang Baik	65	76,6	11,6
Tidak Baik	25	64,99	39,99

Dari Tabel distribusi di atas dapat dilihat bahwa label tidak baik memiliki range sekitar tiga kali lipat dibandingkan dengan label lainnya. Data kemudian diberi label menggunakan nilai konversi di atas. Berikut adalah rincian data tersebut.

TABEL VII
DISTRIBUSI RESPONDEN PENILAIAN

Label	Jumlah
Sangat Setuju	250
Setuju	54
Kurang Setuju	37

Label	Jumlah
Tidak Setuju	6
Total	347



Gbr.6 Diagram lingkaran jumlah distribusi responden

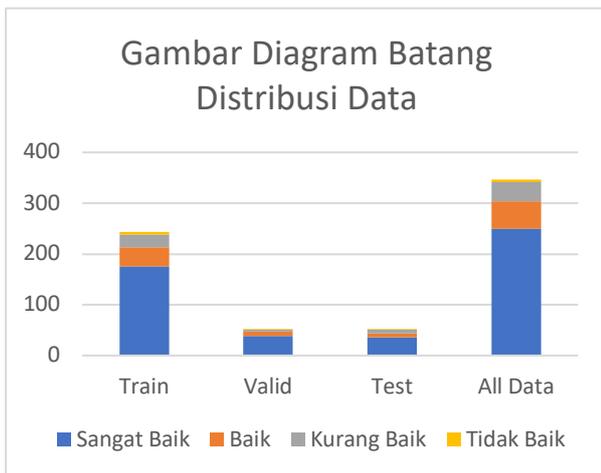
Data kemudian dibagi menjadi 3 bagian, yaitu data training, data validasi, dan data test dengan rasio 7 : 1.5 : 1.5. Berikut adalah rincian distribusi dan presentase data tersebut dibandingkan dengan data total.

TABEL VIII
DISTRIBUSI DATA DI 3 FASE PROSES KLASIFIKASI

Proses	Sangat Baik	Baik	Kurang Baik	Tidak Baik	Total
Train	175	38	26	4	243
Valid	39	8	4	1	52
Test	36	8	7	1	52
All Data	250	54	37	6	347

TABEL IX
PRESENTASE DISTRIBUSI DATA DI 3 FASE PROSES KLASIFIKASI

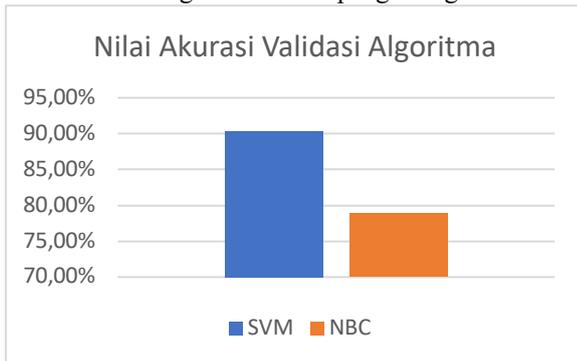
Proses	Sangat Baik	Baik	Kurang Baik	Tidak Baik	Total
Train	72,02%	15,64%	10,70%	1,65%	70%
Valid	75,00%	15,38%	7,69%	1,92%	15%
Test	69,23%	15,38%	13,46%	1,92%	15%
All Data	72,05%	15,56%	10,66%	1,73%	100%



Gbr.7 Diagram batang distribusi data dalam 3 fase proses klasifikasi

Dari rincian di atas dapat dilihat bahwa secara umum, total record untuk label “Sangat Baik” memiliki jumlah paling banyak dengan distribusi sekitar 70-75% dari total data. Hal ini berbanding terbalik dengan data dengan label tidak baik dengan distribusi sekitar 2% dari total data meskipun memiliki range nilai sekitar tiga kali lipat dibanding range lainnya.

Data selanjutnya akan dilakukan penyocokan(fitting) menggunakan dataframe train. Setelah proses fitting data, data dapat divalidasi dengan dilakukan penghitungan nilai akurasi.



Gbr.8 Diagram batang nilai akurasi validasi algoritma

Dari rincian dan diagram di atas, dapat dilihat bahwa dalam fase proses validasi, algoritma klasifikasi SVM memiliki nilai akurasi 90.38% dibandingkan algoritma klasifikasi NBC dengan 78.85%. Hal ini berarti, dalam fase validasi, algoritma SVM berperforma 14.62% lebih baik dibandingkan algoritma NBC.

Setelah proses validasi, kedua algoritma dapat dilakukan pengujian menggunakan dataframe test. Data kemudian dapat dievaluasi menggunakan confusion matrix dan 4 metrik penghitungan yaitu nilai akurasi, precision, recall dan f1-score.

TABEL X
CONFUSION MATRIX KLASIFIKASI ALGORITMA SVM

SVM	Data Prediksi			
	Sangat Baik	Baik	Kurang Baik	Tidak Baik
Data Sebenarnya				
Sangat Baik	36	0	0	0
Baik	0	6	2	0
Kurang Baik	0	1	6	0
Tidak Baik	0	0	0	1

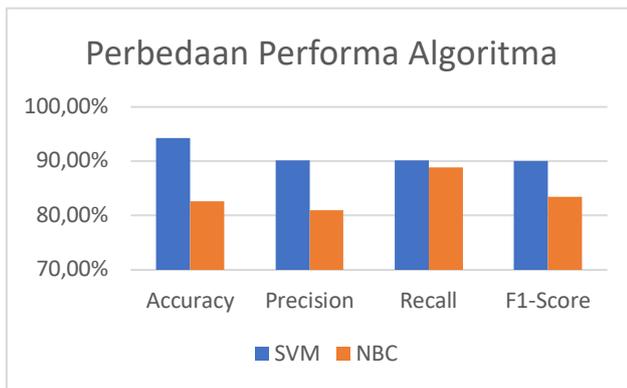
TABEL XI
CONFUSION MATRIX KLASIFIKASI ALGORITMA NBC

NBC	Data Prediksi			
	Sangat Baik	Baik	Kurang Baik	Tidak Baik
Data Sebenarnya				
Sangat Baik	29	7	0	0
Baik	0	6	2	0
Kurang Baik	0	0	7	0
Tidak Baik	0	0	0	1

Dari kedua Tabel di atas, dapat dilihat bahwa range tiga kali lipat dari label tidak baik memudahkan data untuk diklasifikasi. Algoritma SVM dan NBC berhasil memprediksi data yang memiliki label “Tidak Baik”. Algoritma SVM berhasil memprediksi data yang memiliki label “Sangat Baik” sedangkan algoritma NBC memprediksi 29 dari total 36. Kedua algoritma memiliki kesalahan dalam mengklasifikasikan data dengan label baik dan kurang baik. Klasifikasi algoritma SVM dapat memprediksi 6 dari 8 dari data set baik dan 1 false positive baik dari kurang baik, 6 dari 7 dari data set kurang baik dan 1 false positive kurang baik dari baik. Klasifikasi algoritma NBC dapat memprediksi 6 dari 8 dari data set baik dan 7 false positive baik dari sangat baik, 7 dari 7 dari data set kurang baik dan 2 false positive kurang baik dari baik.

TABEL XII
PERFORMA METRIK ALGORITMA KLASIFIKASI

Metrik	SVM	NBC
Accuracy	94,23%	82,69%
Precision	90,18%	80,98%
Recall	90,18%	88,89%
F1-Score	90,00%	83,47%



Gbr.9 Diagram batang perbandingan performa algoritma klasifikasi.

Dari Tabel metrik dan diagram di atas, dapat dilihat bahwa secara umum algoritma SVM berperforma lebih baik dibandingkan algoritma NBC. Algoritma SVM memiliki nilai akurasi 94.23% dibandingkan dengan 82.69% dari algoritma NBC, nilai precision 90.18% dibandingkan dengan 80.98% dari algoritma NBC, nilai recall 90.18% dibandingkan dengan 88.89% dari algoritma NBC, dan nilai f1-score 90.00% dibandingkan dengan 83.47% dari algoritma NBC.

TABEL XIII
PRESENTASE PERBANDINGAN ALGORITMA SVM DIBANDING NBC

MetriK	SVM to NBC
Accuracy	13,96%
Precision	11,36%
Recall	1,45%
F1-Score	7,82%

Dalam rincian Tabel di atas, dapat dilihat bahwa algoritma SVM unggul di empat kategori metrik penghitungan performa model klasifikasi. Algoritma SVM memiliki keunggulan 13.96%, 11.36%, 1.45%, dan 7.82% dibandingkan algoritma NBC pada metrik nilai akurasi, presisi, recall, dan f1-score secara berurut.

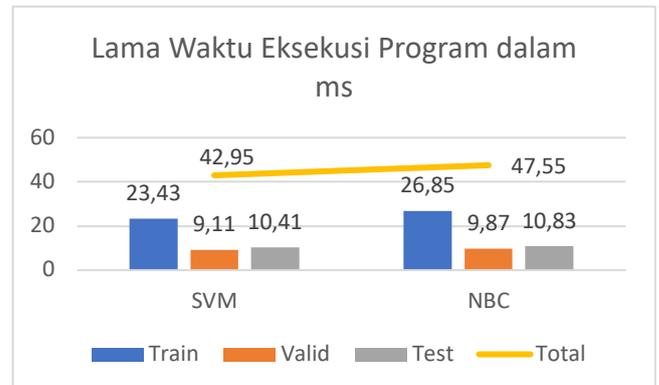
Selain metrik di atas, telah dilakukan penghitungan waktu eksekusi untuk tiap fase klasifikasi. Algoritma SVM melakukan 3 fase klasifikasi dalam waktu 42.94 milidetik dibandingkan NBC dengan waktu 47.55 milidetik.

TABEL XIV
WAKTU PROSES KLASIFIKASI DALAM MS(MILIDETIK)

Process	SVM in ms	NBC in ms
Train	23,43	26,85
Valid	9,11	9,87
Test	10,41	10,83
Total	42,95	47,55

TABEL XIV
INTERVAL WAKTU PROSES DALAM MS DAN PRESENTASE PERBANDINGANNYA

Process	NBC to SVM in ms	Percentage NBC to SVM
Train	3,42	14,61%
Valid	0,76	8,34%
Test	0,42	4,03%
Total	4,6	10,71%



Gbr.10 Diagram batang dan garis lama waktu eksekusi klasifikasi

Dari kedua Tabel di atas, dapat dilihat bahwa mayoritas perbedaan waktu NBC dibanding SVM adalah pada fase training dengan 3.42 milidetik lebih lama. Dengan total waktu perbedaan 4.6 milidetik, dapat dikatakan bahwa total waktu proses eksekusi klasifikasi algoritma NBC 10.71% lebih lama dibandingkan dengan algoritma SVM.

IV. KESIMPULAN

Berdasarkan hasil penelitian, analisis dan pembahasan data mengenai “Perbandingan Metode Support Vector Machine dan Naive Bayes Classifier untuk Klasifikasi Survey Kepuasan Masyarakat terhadap Layanan UPT BLK Surabaya”, peneliti dapat menarik kesimpulan sebagai berikut.

- 1) Hasil dari penelitian ini adalah algoritma Support Vector Machine unggul dalam 4 metrik dan waktu eksekusi dengan nilai akurasi 94.23%, nilai presisi 90.18%, nilai recall 90.18%, f1-score 90.00%. Algoritma Naive Bayes Classifier berperforma kurang dibandingkan dengan algoritma Support Vector Machine dengan nilai akurasi 82.69%, nilai presisi 80.98%, nilai recall 88.89%, f1-score 83.47. Kurangnya performa dalam algoritma Naive Bayes Classifier dapat dijelaskan karena algoritma Naive Bayes Classifier tidak membandingkan atau mencari relasi antar fitur (naive). Sehingga data yang tidak seimbang dapat memengaruhi performa algoritma Naive Bayes Classifier.
- 2) Waktu eksekusi penyocokan data oleh Naive Bayes Classifier yang hampir 15% lebih lambat atau 4.6 ms lebih lambat dengan waktu eksekusi 26.85 ms training, 9.87 ms validation, dan test 10.83 ms dengan total 47.55 ms dibandingkan dengan algoritma Support Vector Machine dengan waktu eksekusi 23.43 ms training, 9.11 ms validation, dan test 10.41 ms dengan total 42.95 ms.

UCAPAN TERIMA KASIH

Puji syukur kehadirat Allah SWT atas berkat dan rahmat-Nya, sehingga penulis dapat menyelesaikan penyusunan artikel jurnal ini sesuai dengan harapan. Selain itu, penulis juga mengucapkan terima kasih kepada semua pemangku kepentingan, karena penyusunan artikel jurnal ini, juga tidak dapat terselesaikan tanpa dukungan, bantuan, bimbingan dan kerjasama berbagai pihak dalam proses penyusunan laporan ini.

REFERENSI

- [1] W. McKinney, "pandas: a Foundational Python Library for Data," p. 1, 2011.
- [2] T. Kabir, "Notebooks, Visual Studio Code style," 08 11 2021. [Online]. Available: <https://code.visualstudio.com/blogs/2021/11/08/custom-notebooks>.
- [3] C. Cortes and V. Vapnik, "Support-Vector Network," *Machine Learning*, 20, pp. 273-297, 1995.
- [4] J. Han, M. Kamber and P. Jian, "Data Mining: Concepts and Techniques, 3rd Edition," in *The Morgan Kaufmann Series in Data Management Systems (Selected Titles)*, Elsevier, 2012.
- [5] scikit-learn, "About us," [Online]. Available: <https://scikit-learn.org/stable/about.html>.
- [6] Python Software Foundation, "General Python FAQ," [Online]. Available: <https://docs.python.org/3/faq/general.html>.
- [7] Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia, Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia Nomor 14 Tahun 2017 Tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik, Jakarta: Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia, 2017.
- [8] Menteri Pendayagunaan Aparatur Negara, Keputusan Menteri Pendayagunaan Aparatur Negara nomor: KEP/25/M.PAN/2/2004, Jakarta: Kementerian Pendayagunaan Aparatur Negara, 2004.
- [9] Suparyanto, "Merancang Kuesioner Penelitian," 2010. [Online]. Available: <https://dr-suparyanto.blogspot.com/2010/08/merancang-kuesioner-penelitian.html>.
- [10] Menteri Pendidikan dan Kebudayaan, Kamus Besar Bahasa Indonesia Edisi V, Jakarta: Kementerian Pendidikan dan Kebudayaan, 2016.
- [11] R. P. Pridiptama, W. Wasono and F. D. T. Amijaya, "Perbandingan Algoritma Support Vector Machine dan Naive Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda)," *Basis: Jurnal Ilmiah Matematika*, pp. 1-16, 2024.
- [12] A. Saputra, M. Subing and R. Pratama, "Perbandingan Metode Naive Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022," *TEKNOMATIKA, Vol.13, No.01*, pp. 22-31, 2023.
- [13] H. Apriyani and Kurniati, "Perbandingan Metode Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *Journal of Information Technology Ampera*, pp. 133-143, 2020.
- [14] H. D. Putra, L. Khairani and D. Hastari, "Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa," *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 120-125, 2023.
- [15] U. Khaira, R. Aryani and R. W. Hardian, "Perbandingan Algoritma Naive Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Kebijakan Kemdikbudristek Tentang Kuota Internet Selama Pandemi Covid-19," *Processor: Jurnal Ilmiah Sistem Informasi, Teknologi Informasi*, pp. 183-191, 2023.
- [16] N. Ikhwana, M. Nusrang and Sudarmin, "PERBANDINGAN METODE PCA-SVM DAN SVM UNTUK KLASIFIKASI INDEKS KEPUASAN MASYARAKAT TERHADAP LAYANAN PENDIDIKAN DI KABUPATEN JENEPONTO," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research Vol. 3 No. 3*, pp. 148-155, 2021.
- [17] M. W. Khoiriyah, I. H. Santi and R. D. Romadhona, "ANALISIS ALGORITMA C4.5 DAN NAIVE BAYES DALAM MENENTUKAN TINGKAT KEPUASAN PUBLIK DI RUPBASAN KELAS 2 BLITAR," *JIP (Jurnal Informatika Polinema)*, pp. 13-18, 2024.
- [18] J. Saputro, Asruddin, A. U. Bani and Mesran, "Penerapan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Program Studi," *Journal of Computer System and Informatics (JoSYC)*, pp. 196-202, 2022.
- [19] Y. Hariyanto, A. Primadewi and M. Hanafi, "Analisis Kepuasan Masyarakat terhadap Pelayanan Publik Menggunakan K-Means Clustering," *Journal of Information System Research (JOSH)*, pp. 1065-1074, 2025.
- [20] S. A. Wicaksono and I. K. D. Nuryana, "Perbandingan Klasifikasi Hasil Indeks Kepuasan Masyarakat Terhadap UPT Balai Latihan Kerja Surabaya Menggunakan Algoritma Decision Tree, Random Forest, K-Nearest Neighbor," *JINACS: Volume 05 Nomor 04*, pp. 552-559, 2024.