

ANALISA PERBANDINGAN TINGKAT PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI JALUR MINAT SMA

Oki Arifin¹⁾, Theopilus Bayu Sasongko²⁾

^{1), 2)} Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta
Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55281
Email : okiarifin@amikom.ac.id¹⁾, theopilus.27@amikom.ac.id²⁾

Abstrak

Untuk membimbing peserta untuk memilih jalur peminatan yang tepat sangat penting dalam jenis pembelajaran yang ada di sekolah agar sebagai dasar untuk melanjutkan keperguruan tinggi. Banyak metode yang bisa dilakukan untuk memilih jalur minat. Salah satu metode yang digunakan adalah teknik data mining. Tapi tidak semua algoritma data mining memiliki kinerja yang baik dalam mengklasifikasi pemilihan jalur minat di jenjang SMA. Oleh karena itu dalam penelitian ini akan membahas tentang analisis perbandingan tingkat performansi algoritma klasifikasi Support Vector Machine (SVM) dan Naïve Bayes Classifier. Analisis perbandingan yang dimaksud adalah perbandingan tingkat akurasi dari kedua algoritma menggunakan 10-fold cross validation. SVM adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. Sedangkan Naïve Bayes Classifier adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Pada penelitian ini digunakan data training guna pembentukan model klasifikasi yaitu nilai peserta didik saat mendaftar di salah satu SMA pada tahun pelajaran 2013-2014 yang berjumlah 288 siswa. Siswa yang berlabel jurusan IPA berjumlah 150 siswa dan siswa yang berlabel jurusan IPS berjumlah 138 siswa. Dengan menggunakan alat bantu RapidMiner studio 8.0, dapat disimpulkan bahwa algoritma SVM (Support Vector Machine) dengan kernel anova dan parameter C sebesar 5.0 relatif lebih unggul dibandingkan algoritma Naïve Bayes Classifier.

Kata kunci: Algoritma Support Vector Machine, Naïve Bayes Classifier, klasifikasi jalur minat

1. Pendahuluan

Untuk membimbing peserta untuk memilih jalur peminatan yang tepat sangat penting dalam jenis pembelajaran yang ada di sekolah agar sebagai dasar untuk melanjutkan keperguruan tinggi. Banyak metode yang bisa dilakukan untuk memilih jalur minat. Salah satu metode yang digunakan adalah teknik data mining. Data mining (DM) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam

database[1], Sedangkan menurut Mirkin (2011) data mining didefinisikan sebagai suatu proses untuk mencari pola dari sekumpulan data yang terdapat di dalam database untuk kemudian dianalisis sehingga menghasilkan suatu informasi tertentu untuk dimanfaatkan pada proses selanjutnya[2]. Untuk proses klasifikasi dalam data mining bisa menggunakan metode klasifikasi atau supervised learning. Klasifikasi merupakan suatu pekerjaan yang melakukan penilaian terhadap suatu objek data untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang tersedia[3]. Klasifikasi dibagi menjadi lima kelompok berdasarkan teori yang diadopsi atau teori yang menjadi dasar teknik klasifikasi. Lima pengelompokan klasifikasi itu adalah classifier Bayes's theorem, distance-based classifier, discriminant classifier, neural networks classifier, dan decision tree classifier[4]. Tapi tidak semua algoritma data mining memiliki kinerja yang baik dalam mengklasifikasi pemilihan jalur minat di jenjang SMA.

Ada beberapa penelitian yang sudah dilakukan mengenai perbandingan metode SVM dan Naïve Bayes Classifier diantaranya adalah penelitian yang dilakukan oleh Shoparak, dkk (2014) untuk deteksi microaneurysm secara otomatis melalui gambar pupil non dilatasi menggunakan metode SVM dan Naïve Bayes. Hasilnya metode SVM memiliki hasil presisi terbaik dari pada Naïve Bayes dalam deteksi microaneurysm secara otomatis sebagai tanda awal terjadinya retinopathy diabetik yang dapat digunakan oleh dokter spesialis mata[5]. Penelitian lain yang dilakukan Mayy M. Al-Tahrawi (2015) membahas tentang teks kategorisasi Arab menggunakan regresi logistik. Dataset yang diambil dari 1500 artikel berita arab yang dibagi menjadi 5 kategori yaitu seni, ekonomi, politik, sains dan olahraga. Langkah-langkah yang dilakukan dalam penelitian tersebut diantaranya preprocessing, tokenization, menghilangkan kata-kata yang bukan bahasa arab, menghilangkan nomor, karakter spesial, tanda baca, menghilangkan kata awalan dan akhiran serta proses stemming. Setelah itu tahap pengklasifikasian menggunakan regresi logistik, Support Vector Machine, Naive Baiyes dan GIS. Dari penelitian, terbukti Support Vector Machines memiliki tingkat akurasi lebih tinggi dibandingkan dengan pengklasifikasi lainnya dalam hal ini regresi logistik, Naïve Baiyes dan GIS[6]. Selanjutnya penelitian yang dilakukan oleh Nurul Anisa Sri Winarsih (2016) membandingkan

metode *Naïve Bayes* dan *Support Vector Machine* untuk klasifikasi emosi pada teks bahasa Indonesia. Hasilnya akurasi metode SVM lebih tinggi dibanding *Naïve Bayes*[7]. Sehingga muncul pertanyaan “Bagaimana tingkat performansi antara model yang dihasilkan oleh metode *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* untuk klasifikasi jalur minat SMA?”

Pada penelitian ini digunakan data dari dua sekolah SMA sebanyak 288 siswa yang diambil pada tahun ajaran 2013-2014. Siswa yang berlabel IPA berjumlah 150 siswa dan siswa yang berlabel jurusan IPS berjumlah 138 siswa. *Dataset* terdiri dari data nilai psikologi (iq, logika rasional, konkrit operasional, abstrak konseptual, analisa sintesa, logika verbal, logika numerik, daya ingat), Nilai UN pada jenjang sebelumnya (Matematika, IPA) dan nilai rata-rata *raport* kelas X (IPA, IPS) yang diambil dari SMA di Jawa Tengah.

Tujuan penelitian ini adalah melakukan analisis perbandingan tingkat performansi yang dihasilkan oleh teknik atau model *data mining* yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* sehingga diperoleh metode manakah yang memiliki tingkat akurasi yang lebih baik untuk mengklasifikasi jalur minat di jenjang SMA.

A. Algoritma *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah suatu metode atau algoritme untuk melakukan klasifikasi maupun prediksi [8]. Prinsip kerja dari metode ini adalah mencari ruang pemisah yang paling optimal dari suatu *dataset* dalam kelas yang berbeda. Dalam kehidupan sehari-hari, kita sering diperhadapkan pada persoalan-persoalan yang tidak *linear* / data yang tidak dapat benar-benar dipisahkan secara *linear* yaitu suatu kondisi dimana tidak ada sebuah garis atau bidang yang dapat dibuat untuk menjadi pemisah antar kelas data. Pada permasalahan ini ada 2 langkah yang dapat dilakukan, yaitu:

1. Menggunakan *soft margin hyperplane*.

Tujuan dari *soft margin hyperplane* adalah mengubah data yang bersifat *non linear* ke dalam bentuk *linear* dengan tetap mempertahankan bidang pembatas tetap fleksibel.

Formulasi pada *soft margin hyperplane* yang menggunakan variabel slack (ξ) dirumuskan dengan Persamaan (1)[9].

$$\begin{aligned} x_i \cdot w + b &\geq 1 - \xi \text{ untuk } y_i = \text{kelas 1} \\ x_i \cdot w + b &\leq -1 + \xi \text{ untuk } y_i = \text{kelas 2} \end{aligned} \quad (1)$$

Dengan Persamaan (1), maka pencarian bidang pemisah terbaik dapat diformulasikan menjadi Persamaan (2)[9].

$$\min \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n \cdot \xi_i \right) \quad (2)$$

Dengan n fungsi pemisah $y_i(x_i \cdot w + b) \geq 1 - \xi \forall i = 1, \dots, n$

C adalah parameter yang menentukan besarnya pinalti yang dikarenakan oleh kesalahan pengklasifikasian data.

2. Mencari *hyperplane* pemisah *linear* dalam ruang dimensi yang baru (*feature space*).

Mengubah *input space* (*dot product*) kedalam bentuk *feature space* sering dikenal dengan teknik kernel *trick* $\Phi(x_i), \Phi(x_j)$ yang kemudian berkembang menjadi fungsi Kernel $K(x_i, x_j)$ [9]. Mengubah dari *input space* ke *feature space* mengakibatkan komputasi yang sangat besar, karena ada kemungkinan dimensi *feature space* yang sangat banyak bahkan tak terhingga. Oleh karena itu SVM menjembatani hal itu dengan fungsi *Kernel*.

Pada Penelitian ini digunakan lima kernel meliputi kernel *dot*, *radial*, *polynomial*, *neural*, dan *anova* (*analysis of variable*).

B. *Naïve Bayes Classifier* (NBC)

Naïve Bayes Classifier (NBC) merupakan pengklasifikasian probabilitas sederhana berdasarkan pada teorema *bayes*. Teorema *bayes* dikombinasikan dengan “*Naïve*” yang berarti setiap atribut atau *variable* bersifat bebas (*independent*). *Naïve Bayes Classifier* dapat dilatih dengan efisien dalam pembelajaran terawasi, proses *Naïve Bayes Classifier* mengasumsikan bahwa ada atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain dikelas yang sama[10]. *Naïve Bayes Classifier* termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk dapat mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar yang akan dijadikan label kelas data masukan tersebut. *Naïve Bayes Classifier* merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma *Naïve Bayes Classifier* juga mampu menangani set data yang memiliki banyak atribut [11]. Persamaan (3) dari *Naïve Bayes Classifier* sebagai berikut:

$$P(C_i | X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (3)$$

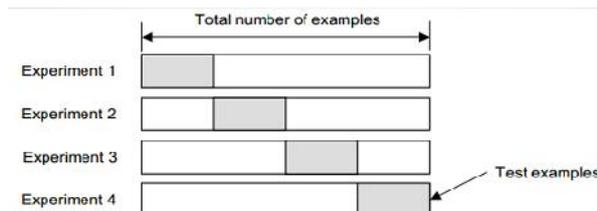
Keterangan :

- X : Kriteria suatu kasus berdasarkan masukan
- C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas
- $P(C_i|X)$: Probabilitas kemunculan label kelas C_i dengan kriteria masukan X
- $P(X|C_i)$: Probabilitas kriteria masukan X dengan label kelas C_i
- $P(C_i)$: Probabilitas label kelas C_i
- $P(X)$: Nilai P(X) selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian $P(X|C_i)P(C_i)$ dengan

memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi.

C. K-Fold Cross Validation

K-fold cross validation adalah teknik untuk mengestimasi performansi dari model pelatihan yang telah dibangun[12]. Metode ini membagi data *training* dan data *testing* sebanyak *k* bagian data. Fungsi dari *k-fold cross validation* adalah agar tidak ada *overlapping* pada data *testing*. Berikut ilustrasi sederhana dari *k-fold cross validation* ditunjukkan pada Gambar 1 dibawah ini.



Gambar 1. Ilustrasi K-Fold Cross Validation

Pada ilustrasi yang tertera pada gambar 1 dapat dijabarkan bahwa percobaan menggunakan *4-fold cross validation*. Hal ini tertera pada jumlah eksperimen yang dilakukan Kotak yang berwarna abu-abu merupakan *test set* dan sisanya (kotak berwarna putih) merupakan *training set*. Sebagai contoh diberikan data sebanyak 40 *instances* pada ilustrasi gambar 3, maka eksperimen pertama kotak yang berwarna abu-abu dari *instances* 1 hingga *instances* dengan urutan yg ke 10 yang berjumlah 10 *instances* merupakan *test set* dan sisanya *instances* urutan ke 11 hingga urutan ke 30 yang berjumlah 30 *instances* (kotak yang berwarna putih) merupakan *training set*. Dari eksperimen pertama didapatkan hasil nilai rata-rata *error*. Kemudian dilanjutkan cara yang sama untuk eksperimen kedua seperti pada eksperimen pertama hingga eksperimen ke empat. Setelah itu dihitung rata-rata keseluruhan estimasi *error* dari setiap eksperimen yang telah dilakukan hingga akhir.

D. Akurasi, Presisi, Recall, dan AUC

Akurasi dapat didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Presisi menunjukkan tingkat ketepatan atau ketelitian dalam pengklasifikasian. Sedangkan *recall* berfungsi untuk mengukur proporsi positif aktual yang benar diidentifikasi. Untuk mengukur akurasi, presisi, dan *recall* biasanya digunakan *confusion matrix*. *Confusion matrix* adalah alat ukur berbentuk *matrix* yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai. Berikut akan disajikan bentuk *confusion matrix* pada Tabel 1 dibawah ini:

Tabel 1. Bentuk Confusion Matrix Dari Dua Kelas

Confusion Matrix		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

Pada Tabel 1 nilai TP (*true positive*) dan TN (*true negative*) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, presisi, dan *recall*. Jika label prediksi keluaran bernilai benar (*true*) dan nilai sebenarnya bernilai salah (*false*) disebut sebagai *false positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*false*) dan nilai sebenarnya bernilai benar (*true*) maka hal ini disebut sebagai *false negative* (FN)[13]. Berikut formulasi untuk menghitung akurasi, presisi, dan *recall* pada pembentukan model klasifikasi ditunjukkan pada Persamaan (4), Persamaan (5), dan Persamaan (6)[12].

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (4)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

AUC (*Area Under Curve*) adalah luas area dibawah kurva. Luas dari AUC selalu berada diantara nilai 0 hingga 1. AUC dihitung berdasarkan gabungan luas trapesium titik-titik (*sensitivity* dan *specificity*) [9]. Berikut adalah standar Tabel kategori pengklasifikasian berdasarkan nilai AUC pada Tabel 2 dibawah ini:

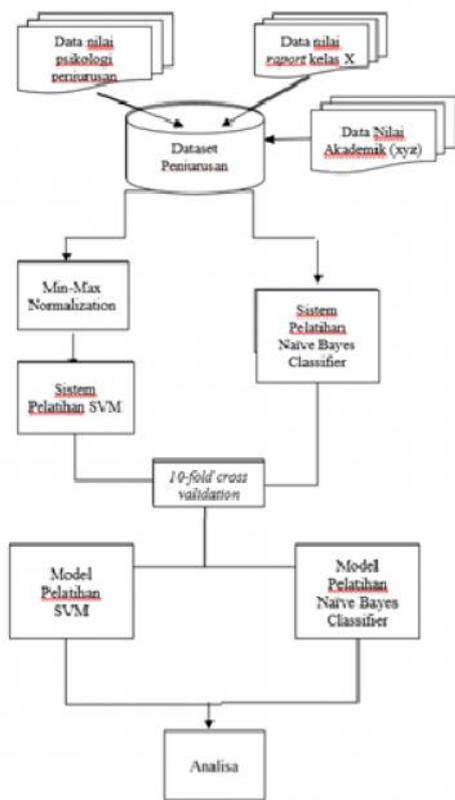
Tabel 2. Kategori Klasifikasi Berdasarkan Nilai AUC

Nilai AUC	Kategori Pengklasifikasian
0.90 - 1.00	Excellent
0.80 - 0.90	Good
0.70 - 0.80	Fair
0.60 - 0.70	Poor
0.50 - 0.60	Fail

2. Pembahasan

A. Metode Penelitian

Metode Penelitian yang dilakukan untuk membandingkan dan menganalisa algoritma SVM dengan *Naïve Bayes Classifier* tercantum seperti pada skema Gambar 2 dibawah ini:



Gambar 2. Skema Metodologi Penelitian

Metode yang digunakan untuk melakukan perbandingan tingkat performansi algoritma SVM dengan *Naive Bayes Classifier* adalah dengan melakukan pengujian tingkat performansi di dua buah dataset peminatan.

Metode penelitian dimulai dari pengumpulan data penelitian di dua buah SMA. *Dataset* pengujian pertama (*dataset* penjurusan abc) terdiri dari data nilai psikologi (iq, logika rasional, konkrit operasional, abstrak konseptual, analisa sintesa, logika verbal, logika numerik, daya ingat), Nilai UN pada jenjang sebelumnya (Matematika, IPA) dan nilai rata-rata raport kelas X (IPA, IPS) yang diambil dari SMA di Jawa Tengah pada tahun ajaran 2013-2014 yang berjumlah 288 siswa. Siswa yang berlabel IPA berjumlah 150 siswa dan siswa yang berlabel jurusan IPS berjumlah 138 siswa. Potongan (*sample*) *dataset* penjurusan abc dapat dilihat pada Tabel 3 dibawah ini:

Tabel 3. Sample Dataset Penjurusan ABC

NAMA	NILAI UN		RATA-RATA RAPORT		NILAI PSIKOLOGI PEMINATAN									
	MATEM	IPA	MIPA	IPS	IQ	LOGIKA RASIONAL	KONKRIT OPERASIONAL	ABSTRAK KONSEPTUAL	ANALISA SINTESA	LOGIKA VERBAL	LOGIKA NUMERIK	DAYA INGAT	PEMIN ATAN	
AAM YULIANTO	9,75	8,25	8,25	7,96	109	109	109	109	127	127	109	109	IPA	
ABU ARSUHMAN	9,5	8,25	8,34	8,16	104	104	106	108	107	106	107	106	IPA	
ADITIA PUTRI	9,0	8,25	8,22	8,11	101	109	106	106	106	106	127	127	IPA	
ADIB ZANONDI	9,5	7,25	8,33	8,1	99	109	109	106	90	106	106	106	IPS	
ADYANTO	8,75	8	7,67	7,84	99	127	109	90	90	106	90	106	IPS	
ADY YUSUF AGEL	8,5	8,25	8,36	8,1	109	109	109	127	106	106	106	106	IPA	
AGHIANINGGA S	9,75	8	8,7	8,52	107	109	109	109	127	109	127	127	IPS	

Dataset pengujian kedua (*dataset* penjurusan xyz) terdiri dari data nilai psikologi (iq, logika rasional, konkrit operasional, abstrak konseptual, analisa sintesa, logika verbal, logika numerik, daya ingat), Nilai UN pada jenjang sebelumnya (Matematika, IPA), nilai rata-rata raport (Matematika, IPA), dan nilai tes akademik. Potongan (*sample*) *dataset* penjurusan xyz seperti pada Tabel 4 dibawah ini:

Tabel 4. Sample Dataset Penjurusan XYZ

NAMA	NILAI UN		RATA-RATA RAPORT		NILAI PSIKOLOGI PEMINATAN								PEMIN ATAN
	MATEM	IPA	MIPA	IPS	IQ	LOGIKA RASIONAL	KONKRIT OPERASIONAL	ABSTRAK KONSEPTUAL	ANALISA SINTESA	LOGIKA VERBAL	LOGIKA NUMERIK	DAYA INGAT	
Vico Blanda N	9,75	8,5	7,5	110	109	109	109	127	127	106	106	109	IPA
Yuliani Karika Sari	9,25	8,75	7,5	110	109	109	127	106	106	106	127	127	IPS
Muhammad Anon K	10	8,75	8,25	109	108	107	109	106	127	106	106	107	IPA
Himali Setyaningrum	9	8,5	8	105	109	109	109	106	106	127	106	127	IPA
Amr Ulfah	10	9	8,25	110	109	109	109	109	127	109	106	106	IPS
Amna Syifa Mardani P	9,5	8,25	7,25	11	109	109	109	106	106	106	106	109	IPS
Kharisma Dewangga	10	8,5	8	108	109	109	109	127	127	106	107	107	IPS

Pada proses klasifikasi dengan menggunakan algoritma SVM dimulai dengan proses normalisasi data. Proses normalisasi yang digunakan menggunakan *min-max normalization* dengan nilai minimal adalah -1 dan nilai maksimal adalah 1. Digunakan model normalisasi ini karena jauh lebih baik dibandingkan *scaling method* yang lain[8]. Potongan (*sample*) *dataset* abc setelah dilakukan normalisasi menggunakan min max normalization seperti pada Tabel 5 dibawah ini:

Tabel 5. Normalisasi Sample Dataset Penjurusan ABC

Nilai UN	Rata-rata Raport		Kemampuan Khusus									Analisa Peminatan	
	Mat	IPA	MIPA	IPS	IQ	Log_Ras	Kon_Op	Abst_Kon	Anal_Si	Log_Ver	Log_Num		Daya Ingat
1,00	0,69	0,78	0,45	0,95	1,00	1,00	0,03	0,00	1,00	1,00	1,00	1,00	IPA
0,11	0,38	0,23	1,00	1,00	1,00	-1,00	-1,00	0,03	1,00	0,03	1,00	1,00	IPS
1,00	0,69	0,20	-0,24	0,16	-1,00	-1,00	0,03	0,00	-1,00	1,00	1,00	1,00	IPA
0,11	-0,25	0,22	-0,00	-0,61	-1,00	-1,00	0,03	-1,00	-1,00	0,03	-1,00	-1,00	IPS
-0,14	0,25	-0,72	-0,60	-1,00	1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	IPS
0,71	0,38	0,13	0,40	0,18	1,00	1,00	0,03	0,95	1,00	0,03	1,00	1,00	IPA
1,00	1,00	1,00	0,21	0,40	1,00	1,00	0,03	0,95	1,00	1,00	1,00	1,00	IPS

Proses sistem pelatihan *Naive Bayes Classifier* tidak dilakukan normalisasi. Data dari Proses pelatihan dari algoritma SVM dan *Naive Bayes Classifier* kemudian dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Pembagian proses data ini menggunakan metode *10-fold cross validation* untuk menguji performansi model dari SVM yang dikomparasikan dengan *Naive Bayes Classifier*. Hasil proses pengujian ini diperoleh performansi dari metode algoritma yang dibandingkan. Kriteria pengujian komparasi metode ini meliputi uji akurasi, uji presisi, uji *recall* dan nilai AUC (*area under curve*). Pengujian komparasi kedua metode ini kemudian dilakukan analisis.

B. Hasil dan Pembahasan

Pada pengujian pertama membandingkan algoritma SVM (*Support Vector Machine*) dan *Naive Bayes Classifier* pada *dataset* penjurusan abc yang berjumlah 288 data dengan 12 atribut penentu. Pembagian dan pengujian *dataset* penjurusan abc menggunakan *10-fold*

cross validation. Tujuan dari pengujian pertama ini adalah melakukan komparasi dan menganalisis algoritma model SVM dan model *Naïve Bayes Classifier* berdasarkan tingkat akurasi, *presisi*, *recall*, dan nilai AUC. Model SVM yang diterapkan pada *kernel anova* dengan nilai C (faktor pinalti) sebesar 5.0 karena parameter dengan nilai ini menghasilkan tingkat akurasi, *presisi*, *recall* terbaik pada algoritma SVM[13].

Hasil Pengujian komparasi SVM dengan *Naïve Bayes Classifier* diperlihatkan pada Tabel 6 dibawah ini:

Tabel 6. Hasil Pengujian Dataset Penjurusan ABC

Algoritma	Akurasi (%)	Presisi (%)	Recall (%)	AUC (0-1)
SVM (anova,5.0)	96.88	98.62	94.95	0.996 (Excellent)
Naïve Bayes	88.87	89.63	87.69	0.964 (Excellent)

Dari data hasil pengujian algoritma SVM (*anova 5.0*) dengan *Naïve Bayes Classifier* dengan menggunakan *dataset* penjurusan abc dipaparkan bahwa tingkat performansi (akurasi, *presisi*, *recall*, dan auc) algoritma SVM dengan *kernel anova* dan nilai C 5.0 lebih besar daripada algoritma *Naïve Bayes Classifier*.

Pada pengujian kedua, implementasi pengujian komparasi metode SVM dengan *Naïve Bayes Classifier* dilakukan pada *dataset* penjurusan xyz yang berjumlah 280 data dengan 12 atribut penentu. Pembagian dan pengujian menggunakan metode *10-fold cross validation*. Hasil pengujian algoritma SVM dengan *Naïve Bayes Classifier* pada *dataset xyz* dapat dilihat pada Tabel 7 dibawah ini:

Tabel 7. Hasil Pengujian Dataset Penjurusan XYZ

Algoritma	Akurasi (%)	Presisi (%)	Recall (%)	AUC (0-1)
SVM (anova,5.0)	97.14	99.44	95.88	0.999 (Excellent)
Naïve Bayes	92.86	93.03	95.85	0.987 (Excellent)

Dari data hasil pengujian SVM (*kernel anova*, nilai C = 5.0) dengan *Naïve Bayes Classifier* pada *dataset* penjurusan xyz pada Tabel 4, didapatkan beberapa informasi yaitu tingkat performansi (akurasi, *presisi*, *recall*, dan auc) algoritma SVM dengan *kernel anova* dan nilai C 5.0 lebih besar daripada algoritma *Naïve Bayes Classifier*, terdapat peningkatan performansi kinerja SVM dan *Naïve bayes* ketika dilakukan pengujian pada *dataset xyz*.

Setelah dilakukan pengujian SVM dengan *Naïve Bayes Classifier* terhadap dua *dataset*, maka Rata-rata (mean) hasil pengujian dapat dilihat pada Tabel 8 dibawah ini:

Tabel 8. Hasil Pengujian Rata-Rata (Mean)

Algoritma	Mean Akurasi (%)	Mean Presisi (%)	Mean Recall (%)	AUC (0-1)
SVM (anova,5.0)	97.01	99.03	95.41	0.997 (Excellent)
Naïve Bayes	90.86	91.33	91.77	0.975 (Excellent)

Pada Tabel 8 dapat ditarik kesimpulan bahwa algoritma SVM dengan *kernel anova* dan *factor pinalti* (C) sebesar 5.0 memiliki rata-rata performansi klasifikasi jalur minat yang lebih unggul daripada *Naïve Bayes Classifier*.

3. Kesimpulan

Dari hasil penelitian dan pembahasan yang sudah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Penelitian dilakukan dengan melakukan komparasi algoritma SVM dengan *Naïve Bayes Classifier*.
2. Digunakan dua *dataset* penjurusan yaitu *dataset* penjurusan abc yang berjumlah 288 siswa dan *dataset* penjurusan xyz yang berjumlah 280 siswa.
3. Tingkat performansi yang diukur dengan akurasi, *presisi*, *recall*, dan nilai AUC (*area under curve*) algoritma SVM yang menggunakan *kernel anova* dan faktor pinalti (C) sebesar 5.0 lebih besar daripada performansi algoritma *Naïve Bayes Classifier*.
4. Pengujian dengan menggunakan *dataset xyz* pada SVM maupun *Naïve Bayes Classifier* memiliki hasil Performansi yang lebih baik daripada pengujian yang dilakukan pada *dataset abc*.

Beberapa Saran kedepan untuk dapat diperbaiki dan dikembangkan dari penelitian ini adalah:

1. Menggunakan percobaan *dataset*, atribut, jumlah data yang lebih beragam.
2. Dilakukan uji statistik untuk menguji keabsahan performansi.

Daftar Pustaka

- [1] Turban, E., Aronson, J.E. dan Liang, T.P, *Decision Support System and Intelligent System*, edisi 7 jilid 1, Cv Andi Offset : Yogyakarta, 2005.
- [2] Mirkin, B., "Data Analysis, Mathematical Statistics, Machine Learning, Data Mining: Similarities and Differences", *Int. Conf. Adv. Comput. Sci. Inf. Syst.*, Vol.2, pp. 1-8, 2011.
- [3] Prasetyo, E., *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, 1st ed : Cv Andi Offset : Yogyakarta, 2012.
- [4] T. Justin, R. Gajšek, V. Štruc, and S. Dobrišek, "Comparison of different classification methods for emotion recognition," in *MIPRO, 2010 Proceedings of the 33rd International Convention*, 2010, pp. 700-703.
- [5] Sopharak, A., Uyyanonvara, B., dan Barman, S., "Comparing SVM and Naive Bayes Classifier for Automatic Microaneurysm Detections", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol.8, No.5, 797-800, 2014.

- [6] Al-tahrawi, M. M, "Arabic Text Categorization Using Logistic Regression", *I.J.Intelligent System and Applications*, 06, 71–78,2015.
- [7] Winarsih, Nurul., A., S., "Perbandingan Metode Naive Bayes Dan Support Vector Machine Untuk Klasifikasi Emosi Pada Teks Bahasa Indonesia", Skripsi, Jurusan Teknik Informatika, Udinus : Semarang, 2016.
- [8] J. Agustinus, "Sistem Deteksi Intrusi Jaringan dengan Metode Support Vector Machine", M. Eng, Thesis, Jurusan Ilmu Komputer FMIPA UGM, Yogyakarta, 2012.
- [9] I.H. Witten, E. Frank, and M.A. Hall, "Data Mining Practical Machine Learning Tools and Techniques", Third Edition, Elsevier Publisher, USA, 2011.
- [10] Setiawan,W., "Sistem Pakar Diagnosa Penyakit Mata Menggunakan Naïve Bayes Classifier", *Seminar Nasional Sains dan Teknologi* , 2014.
- [11] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J.-F. Chang, and L. Hua, "Data Mining in Healthcare and Biomedicine: A Survey of The Literature", pp. 2431–2448, 2012.
- [12] J. Han, and M. Kamber, "Data mining: Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers, San Fransisco, 2013.
- [13] B. Theopilus, "Klasifikasi Jalur Minat Siswa Menggunakan Algoritme Support Vector Machine (SVM) (Kasus : SMA Negeri 1 dan SMA Negeri 2 Sragen)", Tesis, Jurusan Teknik Elektro dan Teknologi Informasi, UGM, Yogyakarta, 2014.

Biodata Penulis

Oki Arifin, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2012. Memperoleh gelar Master of Computer Science (M.Cs.) Program Pasca Sarjana Magister Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2016. Saat ini menjadi Dosen di Universitas AMIKOM Yogyakarta.

Theopilus Bayu Sasongko, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi Universitas Kristen Duta Wacana, lulus tahun 2012. Memperoleh gelar Master of Engineering (M.Eng) Program Pasca Sarjana Magister Teknologi Informasi Universitas Gajah Mada Yogyakarta, lulus tahun 2014. Saat ini menjadi Dosen di Universitas AMIKOM Yogyakarta.