



**Prediksi Penjurusan IPA, IPS dan BAHASA dengan
Menggunakan Machine Learning**

Edy Widodo

Program Studi Teknik Informatika, Universitas Pelita Bangsa
Korespondensi email: ewidodo@pelitabangsa.ac.id

Abstrak

Data Mining is an artificial intelligence (Artificial Intelligence) that is very useful and useful for research scientists who include machine learning, statistics and databases. Thus data mining is a research method with the aim of uncovering hidden patterns. It can also be interpreted that data mining is a very useful and useful research pattern for turning data into information. With this application, it will be easier for teachers, parents, guardians of students and students to be more practical and accurate in determining the direction of Natural Sciences, Social Sciences and Languages. Which will be the weight of the calculation of the majors determination are the SMP (National Examination) UN scores, TKD scores (Basic Ability Tests) in high school, IQ scores, students 'interest values and parents' desire values (being another alternative). Thus the results of these predictions will result in the appropriate and accurate major with the ability of students and girls related.

Informasi Artikel

Diterima: 26-02-2020
Direvisi: 5-03-2020
Dipublikasikan: 25-04-2020

Keywords

Naive bayes, decision tree, support vector machine dan Machine Learning

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu data yang selama ini tidak kita ketahui secara manual. Bahwa kata *mining* disini berarti usaha untuk mendapatkan sejumlah besar material data dasar. Sehingga *Data Mining* itu memiliki akar yang panjang dari bidang ilmu kecerdasan tiruan (*Artificial Intelligent*), *machine learning*, statistik dan database. Oleh sebab itu *Data Mining* adalah proses menerapkan metode ini untuk data

dengan tujuan mengungkap pola-pola tersembunyi. Juga dapat diartikan bahwa *Data Mining* itu adalah proses untuk penggalian pola-pola dari data. *Data mining* juga menjadi alat yang semakin penting untuk mengubah data menjadi informasi. Hal ini sering digunakan dalam berbagai praktek profil, seperti pemasaran, pengawasan, penipuan deteksi dan penemuan ilmiah.

Data Mining telah digunakan selama bertahun-tahun oleh bisnis, ilmuwan dan pemerintahan untuk menjangkau volume data seperti catatan perjalanan penumpang penerbangan, data sensus dan supermarket scanner data untuk menghasilkan laporan riset pasar. Alasan utama untuk menggunakan data mining adalah untuk membantu dalam analisis koleksi pengamatan perilaku. Beberapa teknik yang sering disebut-sebut dalam literatur *Data Mining* dalam penerapannya antarlain : *clustering, classification, association rule mining, neural network, genetic algorithm* dan lain-lain. Yang membedakan persepsi terhadap *Data Mining* adalah perkembangan teknik-teknik *Data Mining* untuk aplikasi pada database skala besar.

Dengan berbagai permasalahan dan sangat sulit sekali untuk menentukan siswa-siswi dalam menentukan penjurusan IPA, IPS dan BAHASA di sebuah pendidikan menengah atas (SMA), baik dari segi

kemampuan , akademik dan tes IQ sekalipun masih banyak kendala yang memang belum dapat terselesaikan sampai saat ini, sehingga untuk menentukan penjurusan tersebut banyak sekali kendala yang dialami. Sistem yang akan saya gunakan ini mudah-mudahan dapat membantu para guru, siswa-siswi dan orang tua untuk menentukan penjurusan yang sangat tepat dan sesuai dengan kemampuan anak didik.

Sistem yang akan dipergunakan oleh *Machine Learning* dengan aplikasi *Rapid Miner* untuk menentukan penjurusan IPA, IPS dan BAHASA secara komputerisasi, agar menghasilkan penjurusan yang akurat. Tujuan memprediksi penjurusan dengan *Machine Learning* (Moertini, V. S. (2002)) ini untuk mencari solusi yang selama ini menggunakan perhitungan manual yang hasilnya tidak relevan, tetapi untuk mencari dan hasil yang sesuai dengan minat, bakat dan keinginan dari siswa-siswi yang terkait bukan paksaan dari siapapun.

Dengan menganalisa dan observasi maka dengan judul “ *Prediksi Penjurusan IPA, IPS dan BAHASA Menggunakan Machine Learning* “ (Vercellis, C. (2009), dapat menjadi acuan untuk perumusan menentukan penjurusan sesuai dengan minat dan bakat. Tapi tidak bermaksud untuk merubah atau mengganti kebijakan sekolah SMA terkait. Penelitian ini untuk mencari solusi yang terbaik serta dapat membantu meringankan kinerja para guru untuk menentukan penjurusan yang baik, benar, cepat dan akurat sesuai dengan minat dan bakat siswa.

Beberapa metode yang akan digunakan sebagai tolak ukur penelitian ini adalah menggunakan *naive bayes, decision tree* dan *SVM (Support Vector Machine)* Aydin, I., Karakose, M., & Akin, E. (2011). Hasil akhir dari penelitian ini nantinya akan sangat berguna karena menghasilkan hasil out put yang memang sangat akurat sekali.

Algoritma naive bayes merupakan salah satu dari metode pengklasifikasi yang sangat sederhana yang menerapkan *teorema bayes* dengan variabel yang bebas (*independen*), yang mana satu variabel dengan variabel lainnya saling terkait dan terhubung sesuai dengan yang diinginkan oleh peneliti. Algoritma decision tree merupakan metode klasifikasi khusus dan dapat memprediksi data penjurusan yang sangat kuat dengan memisahkan dua kelas dengan kelas yang lainnya (Jianwei Han, 2001).

- a. Nilai UN (Ujian Negara) SMP
- b. Nilai TKD (Tes Kemampuan Dasar) yang diadakan oleh SMA Negeri satu Setu yang meliputi :
 - Nilai tes bahasa Indonesia
 - Nilai tes Bahasa Inggris
 - Nilai tes Matematika
 - Nilai tes Ilmu Pengetahuan Alam
 - Nilai tes Ilmu Pengetahuan Sosial
- c. Nilai tes Minat atau tes IQ (*Intelligence Quotient*)
- d. Nilai tes wawancara guru BK (Bimbingan Konseling) dari SMA
- e. Permintaan orang tua wali murid

Tabel. 1. Panduan nilai penjurusan

NO.	PENJURUSAN	NILAI
1	IPA	77 - 100
2	IPS	74 - 76
3	BAHASA	73 kebawah

Algoritma SVM (*Support Vector Machine*) (Aydin, I., Karakose, M., & Akin, E. (2011). merupakan metode yang dapat mengklasifikasikan data yang memiliki dimensi yang lebih tinggi dari pada vektor input dan dapat memisahkan dengan feature space yang berbeda. Dengan beberapa metode tersebut akan dapat menghasilkan beberapa masukan untuk menghasilkan output yang diinginkan dengan menggunakan tool *rapid miner*.

Dalam melakukan penelitian ini yang perlu dan sangat penting dipersiapkan adalah semua data siswa sebagai data uji. Peneliti berusaha mengumpulkan semua data siswa yang terkait dengan nilai dan penjurusan. Dengan demikian peneliti membutuhkan beberapa atribut yang memang sangat dibutuhkan untuk mendapatkan hasil yang baik dengan menggunakan beberapa metode. Atribut yang peneliti gunakan adalah :

I. Metodologi

Agar dapat menghasilkan penjurusan yang tepat, akurat serta sesuai dengan minat dan bakat serta keinginan siswa-siswi terkait dengan menggunakan Machine Learning, maka diperlukan data-data sebagai tujuan masalah yaitu :

1. Data set yang perlu digunakan dalam metode ini adalah nilai UN (Ujian Nasional) SMP, nilai TKD (Tes Kemampuan Dasar), nilai hasil tes IQ (*Intelligence Quotient*) dan nilai minat anak didik termasuk keinginan dari orang tua, data set yang memang harus digunakan adalah semua data-data siswa-siswi kelas X SMA baik MIPA, IPS DAN BAHASA.
2. Metode untuk memprediksi penentuan penjurusan IPA, IPS dan BAHASA menggunakan Machine Learning dengan perbandingan menggunakan aplikasi *Rapid Miner*.
3. Dengan menggunakan beberapa metode penelitian untuk mendapatkan hasil yang tepat sesuai dengan harapan siswa maupun orang tua siswa terkait, metode tersebut adalah *Naïve Bayes*,

Decision Tree dan Support Vector Machine.

Dalam melakukan penelitian dengan menggunakan data mining sebagai data tes yaitu semua siswa-siswi SMA yang di hitung selama 3 tahun. Data tes tersebut meliputi penjurusan IPA, IPS dan BAHASA dengan menggunakan data-data siswa yang diterima di SMA terkait dalam 3 periode tahun ajaran 2015 - 2016, tahun ajaran 2016 – 2017 dan tahun ajaran 2017 - 2018. Data tersebut di olah menjadi beberapa kelompok sebagai data tes dan selanjutnya untuk penelitian menggunakan beberapa algoritma yaitu *Naive Bayes, Decision Tree dan SVM (Support Vector Machine)*. Hasil dari pada penelitian tersebut dihasilkan bahwa semakin data tes sedikit maka hasilnya akan lebih baik dan bila data tes semakin banyak maka nilai acurasinya berkurang dalam bentuk presentasi.



Gambar 1. Data siswa selama 3 tahun

Tabel.2. Jumlah siswa

Jurusan	Jumlah Siswa
IPA	510
IPS	263
BAHASA	316
TOTAL	1089

III. Metodologi dan Pengambilan Data

Dari hasil penelitian tersebut dapat diartikan bahwa Machine Learning sangat dibutuhkan dalam penyelesaian untuk mencari nilai kebenaran, yang memang sangat mempengaruhi tingkat keberhasilan dalam

melakukan penelitian adalah kesalahan penjurusan yang memang dipaksakan dan tidak sesuai dengan hasil tes. Sehingga peneliti sangat sulit untuk mendapatkan hasil yang baik dan benar.

Klasifikasi adalah salah satu pembelajaran yang paling umum di data *mining*, dan dapat didefinisikan sebagai bentuk analisis data untuk mengekstrak data atau model yang akan digunakan untuk penelitian dengan beberapa atribut. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas dalam statistik. Klasifikasi data terdiri dari 2 proses yaitu tahap pembelajaran dan tahap pengklasifikasian.

Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi sedangkan tahap pengklasifikasian merupakan tahapan model klasifikasi untuk memprediksi label kelas dari suatu data. Tahapan dari klasifikasi dalam data *mining* terdiri dari :

1. Pembangunan model dalam tahapan ini dibuat untuk menyelesaikan masalah klasifikasi *class* atau *atribut* dalam data, model ini dibangun berdasarkan training set sebuah contoh data dari permasalahan yang dihadapi. Training set tersebut sudah mempunyai informasi yang lengkap, baik atribut maupun classnya.
2. Penerapan model, pada tahapan ini yang sudah dibangun sebelumnya digunakan untuk menentukan atribut / class dari sebuah data baru yang atribut / classnya belum diketahui sebelumnya.
3. Evaluasi, pada tahapan ini hasil dari penerapan model pada tahapan sebelumnya di evaluasi menggunakan parameter terukur untuk menentukan apakah model tersebut dapat diterima.

Banyak algoritma yang dapat di gunakan dalam pengklasifikasian data, namun

dalam penelitian ini hanya akan membandingkan tiga algoritma saja, yakni *naive bayes*, *decision tree* dan *SVM* (*Support Vector Machine*)

1. Naive Bayes

Naive bayesian klasifikasi adalah suatu klasifikasi berpeluang sederhana berdasarkan aplikasi teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen). Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu kejadian tertentu dari suatu kelompok tidak berhubungan dengan kehadiran atau ketiadaan dari kejadian lainnya.

Naive Bayesian dapat digunakan untuk berbagai macam keperluan antara lain untuk klasifikasi dokumen, deteksi spam atau filtering spam, dan masalah klasifikasi lainnya. Dalam hal ini lebih disorot mengenai penggunaan teorema *Naive Bayesian* untuk spam *filtering*. *Naive Bayes* merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas.

Konsep dasar yang digunakan oleh *Naive bayes* adalah *Teorema Bayes*, yaitu teorema dalam statistika untuk menghitung peluang, *Bayes Optimal Classifier* menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

Perhitungan data nominal digunakan untuk menghitung :

- PA (yes|E) : yaitu peluang jurusan yang di inginkan IPA dan keputusan akhirnya IPA
- PA (no|E): yaitu Peluang jurusan yang di inginkan IPA dan keputusan akhirnya IPS
- PB (yes|E): yaitu Peluang hasil Psikotest IPA dan keputusan akhirnya IPA

-PB (no|E): yaitu Peluang hasil Psikotest IPS dan keputusan akhirnya IPS

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X_i)} \quad (1)$$

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

2. Decision Tree

Pohon keputusan atau disebut juga dengan *Decision tree* merupakan teknik yang sangat populer dalam data mining saat ini. Metode *decision tree* memang sangat mudah dan berfungsi untuk menentukan keputusan suatu kasus karena tidak memerlukan proses pengolahan pengetahuan terdahulu sehingga dapat menyelesaikan dengan sangat sederhana kasus-kasus yang mempunyai dimensi yang besar.

Keuntungan dalam metode *decision tree* tingkat akurasi yang sangat baik, dengan syarat data tes merupakan data yang benar-benar valid. Terbukti bahwasannya metode ini sering digunakan dalam berbagai bidang, antara lain kesehatan, financial, astronomi hingga pendidikan dan juga mungkin di bidang-bidang lainnya. Contohnya saja dalam bidang penelitian yang sedang lakukan dalam studi kasus penjurusan siswa-siswi SMA yaitu penjurusan IPA, IPS dan BAHASA. Karena penjurusan di SMA

sangat perlu dilakukan untuk menentukan tingkat keahlian dibidangnya masing-masing, sehingga siswa didik dapat fokus dengan materi dan pembelajaran yang sesuai dengan penjurusannya.

Peneliti saat menggunakan data tes Uji nilai untuk mencari penjurusan yang tepat sasaran yaitu menggunakan nilai UN SMP (Ujian Nasional SMP), TKD (Tes Kemampuan Dasar) yang dilakukan di SMA terkait dan nilai IQ (*Intelligence Quotient*).Nyatanya, metode ini banyak diaplikasikan dalam berbagai bidang, antara lain kesehatan, financial, astronomi hingga pendidikan. ataupun bidang-bidang lain yang saya sendiri belum pernah tahu. Kajian teori dari *decision tree* tidak saya bahas, karena saya yakin diluar sana sudah banyak yang membahasnya. Saya akan langsung membahas dalam bentuk implementasi nyata dalam sebuah studi kasus. Studi kasus yang saya angkat adalah Penjurusan Siswa SMA.

Sekarang, saatnya kita membuat *decision tree* sebagai system pendukung keputusan untuk penjurusan siswa Sekolah Menengah Atas (SMA), mengapa aplikasi ini diperlukan ? Penjurusan siswa SMA merupakan kewajiban wali kelas untuk menentukan jurusan apa yang tepat bagi siswa-siswa didiknya pada tingkat SMA, penjurusan dilakukan ketika memasuki jenjang akhir semester II atau semester genap kelas X. Pastinya, penentuan penjurusan didasarkan atas 3 nilai utama dari raport, yaitu nilai IPA, IPS dan Bahasa.

Seorang wali kelas bias saja menentukan penjurusan berdasarkan nilai tertinggi dari ketiga nilai. Namun, apa jadinya jika nilai IPA yang diberikan guru IPA terlalu ketat? akibatnya siswa yang masuk jurusan IPA pasti lebih sedikit. Bagaimana solusinya? Kita akan mengacu pada sebuah data penjurusan siswa SMA yang tidak berdasarkan nilai tertinggi, data ini kita sebut dengan data training. Data training

biasanya didapatkan melalui hasil penjurusan tahun lalu.

Pemilihan atribut untuk mejadi *root node* internal *node* sebagai atribut tes berdasarkan atas ukuran *impurity* dari masing-masing atribut. Ukuran-ukuran *impurity* yang umumnya digunakan adalah *information gain*, *rain ratio* dan *gain index*. Atribut yang memiliki *impurity* tertinggi akan terpilih sebagai atribut tes. *Information gain* merupakan suatu ukuran korelasi pada model parametik yang menggambarkan ketergantungan antara 2 perubah acak X dan Y. *Information Gain* memiliki rumus :

$$Gain(A)=I(S_1, S_2, \dots, S_m) - E(A)$$

$$I(S_1, S_2, \dots, S_m)f(x) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2)$$

Keterangan :

S_i : Jumlah sampel dalam himpunan sampel S (berisi S sampel) yang masuk kelas C

m : Banyaknya nilai yang berbeda atribut label kelas yang akan mendefinisikan kelas yang berbeda, C_i (i=1,2,...,m)

P_i : Peluang bahwa suatu sampel akan masuk ke kelas C_i dan dengan $\frac{S}{S_i}$

E(A) : Entropy A

Secara statistik, *entropy* menyatakan ukuran ketidakpastian secara probabilistik. *Entropy* A memiliki rumusan :

$$E(A)= \sum_{j=1}^v \frac{\delta_{1j}+S_{2j}+\dots+S_{mj}}{s} I(S_{1j}, S_{2j}, \dots, S_{mj}) \quad (3)$$

v :Banyaknya nilai / katagori yang berbeda yang dimiliki attribute A

S_{1j} : Banyaknya sampel pada attribute A yang masuk katagori atribut A

$$\frac{S_{1j}+S_{2j}+\dots+S_{mj}}{s} \quad (4)$$

Menyatakan proposisi jumlah *simple attribute* A kategori *j* terhadap jumlah *simple* total. *Gain ratio* merupakan modifikasi dari *information gain* untuk mengurangi bias atribut yang memiliki banyak cabang. *Gain ratio* memiliki sifat :

1. Bermilai besar bila data menyebar rata
2. Bermilai kecil bila semua data masuk dalam satu cabang

$$\text{Gain Ratio} = \frac{\text{Gain}}{\text{Split Info}} \quad (5)$$

Dimana rumus split info seperti pada rumus diatas dengan *m* menyatakan banyak split. Jenis split yang dipilih adalah split yang memiliki nilai *Gain Ratio* yang besar.

3. SVM (*Support Vector Machine*)

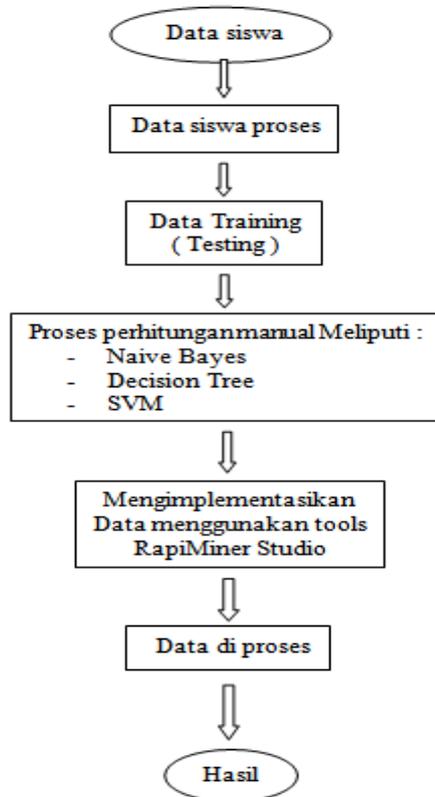
Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. Sebagai salah satu metode *pattern recognition*, usia SVM terbilang masih relatif muda. Walaupun demikian, evaluasi kemampuannya dalam berbagai aplikasinya menempatkannya sebagai *state of the art* dalam *pattern recognition*, dan dewasa ini merupakan salah satu tema yang berkembang dengan pesat. 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*. Tulisan ini membahas teori dasar SVM dan aplikasinya dalam *bioinformatics*, khususnya pada analisa ekspresi gen yang diperoleh dari analisa *micro array*.

Pengertian yang lainnya adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*)

berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik.

Generalisasi didefinisikan sebagai kemampuan suatu metode (SVM, *neural network*, dsb.) untuk mengklasifikasikan suatu *pattern*, yang tidak termasuk data yang dipakai dalam fase pembelajaran metode itu. Vapnik menjelaskan bahwa *generalization error* dipengaruhi oleh dua faktor: error terhadap *training set*, dan satu faktor lagi yang dipengaruhi oleh dimensi VC (*Vapnik-Chervokinensis*). Strategi pembelajaran pada *neural network* dan umumnya metode *learning machine* difokuskan pada usaha untuk meminimalkan *error* pada *training-set*. Strategi ini disebut *Empirical Risk Minimization* (ERM). Adapun SVM selain meminimalkan *error* pada *training-set*, juga meminimalkan faktor kedua. Strategi ini disebut *Structural Risk Minimization* (SRM), dan dalam SVM diwujudkan dengan memilih *hyperplane* dengan margin terbesar. Berbagai studi empiris menunjukkan bahwa pendekatan SRM pada *Support Vector Machine* memberikan *error generalisasi* yang lebih kecil daripada yang diperoleh dari strategi ERM pada *neural network* maupun metode yang lain.

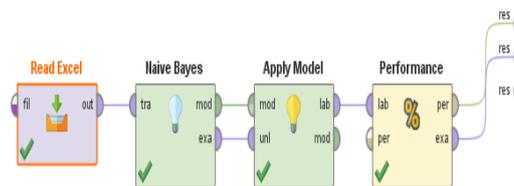
SVM dapat diimplementasikan relatif mudah, karena proses penentuan *support vector* dapat dirumuskan dalam *QP problem*. Dengan demikian jika kita memiliki *library* untuk menyelesaikan *QP problem*, dengan sendirinya SVM (*Support Vector Machine*) dapat diimplementasikan dengan mudah. Selain itu dapat diselesaikan dengan metode sekuensial.



Gambar 2. Alir Penelitian

IV. Hasil Penelitian

Hasil penelitian yang peneliti jalankan menggunakan data tes yang memang aktual dan nyata, yang meliputi seluruh siswa selama 3 tahun terakhir dengan 3 penjurusan di SMA yaitu IPA, IPS dan BAHASA.



Gambar.3. Hasil Tes Menggunakan Rapid Miner

accuracy: 100.00%

	true IPA	true BAHASA	true IPS	class precision
pred. IPA	6	0	0	100.00%
pred. BAHASA	0	11	0	100.00%
pred. IPS	0	0	3	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar. 4. Hasil Akurasi

Tabel.3. Hasil Nilai Tes Siswa

NAMA	Jen.Kel	UN	IQ	RATA2	Jurusan
ACHMAD CANDRA N	Laki-laki	72,75	109	80,49	IPA
ADE OCTAFIANI	Perempuan	68,38	114	39,08	BAHASA
ADE SUKMA MELATI	Perempuan	60,63	96	33,92	BAHASA
ADITYA FERDIYANSYAH	Laki-laki	52	90	68,14	BAHASA
ADITYA NUR RAHMAN	Laki-laki	51,13	95	68,11	BAHASA
ADITYA PRATAMA PUTRA	Perempuan	70,63	103	77,16	IPA
AENI OKTAVIANI	Perempuan	73,5	111	40,36	BAHASA
AFNI OCTAVIA	Laki-laki	64	95	34,9	BAHASA
AGIL MUHAMMAD FRIZDA	Laki-laki	54,5	102	69,62	BAHASA
AGUM MAHRUR ROBBY	Laki-laki	68,75	119	39,92	BAHASA
AGUSTINUS MASARINUS	Laki-laki	66	109	77,05	IPS
AHMAD FAJRI PARWADIS	Laki-laki	72,13	98	77,04	IPA
AHMAD SINGGIH	Laki-laki	63,63	101	75,78	IPA
AHSANIA NABILLA	Perempuan	67,75	108	76,73	IPS
AISYAH BARKAH	Perempuan	54,75	98	71,73	BAHASA
AJENG KARTIKA PUTRI	Perempuan	82,5	116	87,36	IPA
AJHENG YUNIAR PERNANDA	Perempuan	64,5	95	35,07	BAHASA
AL MILLA SYIFA AULIA	Perempuan	53,63	90	30,73	BAHASA
ALDIAD MADRUSDI	Laki-laki	64,88	110	76,43	IPS
ALDO VAREL BRAMASTA	Laki-laki	76,13	105	81,52	IPA

Dari hasil penelitian dengan data tes yang sudah ditentukan adalah :

Tabel.4. Hasil Tes Menggunakan Naive Bayes

No.	Data Tes Jumlah Siswa	Metode	Penjurusan			Akurat
			IPA	IPS	BHS	
1	20	Naive Bayes	6	3	11	100%
	99		33	24	31	88,89%
	399		132	93	98	80,95%
	699		259	207	181	92,56%
	1089		498	250	317	97,80%

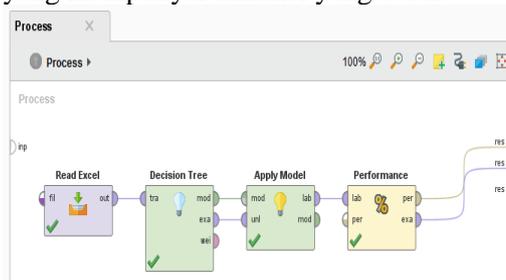
Tabel.5 Siswa Salah Jurusan

Ket. Salah Penjurusan			
IPA	IPS	BHS	%
0	0	0	0%
1	0	10	11,11%
14	8	54	19,05%
35	0	17	7,44%
12	12	0	2,20%

Jadi metode *naive bayes* sangat bagus dan akurat jika dipergunakan dalam menentukan penjurusan di sekolah SMA

Metode Decision Tree

Pohon keputusan atau disebut dengan Decision Tree merupakan teknik yang sangat populer dalam data mining saat ini. Metode ini memang sangat mudah dan berfungsi untuk menentukan keputusan suatu kasus karena tidak memerlukan proses pengolahan pengetahuan terlebih dahulu sehingga dapat menyelesaikan dengan sangat sederhana kasus-kasus yang mempunyai dimensi yang besar.



Gambar. 5. Proses Decision Tree

Tabel.6 Hasil Akurasi Decision Tree

accuracy: 100.00%

	true IPA	true BAHASA	true IPS	class precision
pred. IPA	6	0	0	100.00%
pred. BAHASA	0	11	0	100.00%
pred. IPS	0	0	3	100.00%
class recall	100.00%	100.00%	100.00%	

Description : PerformanceVector:

Accuracy : 100.00%

ConfusionMatrix:

True:	IPA	BAHASA	IPS
IPA:	6	0	0
BAHASA:	0	11	0
IPS:	0	0	3

Tabel.7. Hasil Tes Menggunakan Decision Tree

No.	Data Tes Jumlah Siswa	Metode	Penjurusan			Akurat
			IPA	IPS	BHS	
2	20	Decision Tree	6	3	11	100%
	99		33	20	40	93,94%
	399		131	74	144	87,47%
	699		189	93	171	64,81%
	1089		470	25	213	65,01%

Tabel 8. Siswa Salah Jurusan

Ket. Salah Penjurusan			
IPA	IPS	BHS	%
0	0	0	0%
2	3	1	6,06%
19	23	8	12,53%
105	131	10	35,19%
40	237	104	34,99%

Jika metode decision tree semakin banyak data tes hasilnya kurang akurat bila digunakan untuk menentukan penjurusan pada penjurusan di SMA.

SVM (Support Vector Machine)

Proses data testing menggunakan metode SVM dengan tools Rapid Miner.

Tabel.9. Hasil Tes Menggunakan SVM

accuracy: 55.00% +/- 15.00% (mikro: 55.00%)

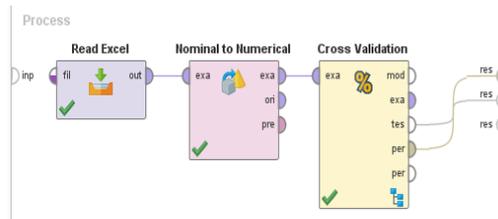
	true IPA	true BAHASA	true IPS	class precision
pred. IPA	0	0	0	0.00%
pred. BAHASA	6	11	3	55.00%
pred. IPS	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	

PerformanceVector:
accuracy: 55.00% +/- 15.00% (mikro: 55.00%)

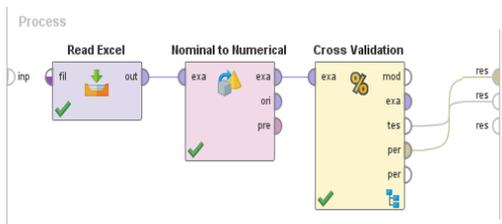
ConfusionMatrix:
True: IPA BAHASA IPS
IPA: 0 0 0
BAHASA: 6 11 3
IPS: 0 0 0

Classification_error: 45.00% +/- 15.00% (mikro: 45.00%)

ConfusionMatrix:
True: IPA BAHASA IPS
IPA: 0 0 0
BAHASA: 6 11 3
IPS: 0 0 0



Gambar. 6. Proses Support Vector Machine



classification_error: 32.23% +/- 2.86% (mikro: 32.23%)

	true IPA	true BAHASA	true IPS	class precision
pred. IPA	425	55	93	74.17%
pred. BAHASA	52	207	63	64.29%
pred. IPS	33	55	106	54.64%
class recall	83.33%	65.30%	40.46%	

Gambar.7 Hasil SVM

Tabel.10. Hasil Semua Tes Menggunakan SVM

No.	Data Tes (Jumlah Siswa)	Metode	Penjurusan			Akurat
			IPA	IPS	BHS	
3	20	SVM	0	0	11	55,00%
	99		33	20	40	93,94%

399	131	74	144	87,47%
699	189	93	171	64,81%
1089	425	106	207	67,77%

Metode	Presentasi
Naive Bayes	97,80%
Decision Tree	65,01%
SVM (Support Vector Machine)	67,77%

Tabel.11. Hasil Kesalahan Penjurusan Menggunakan SVM

Ket. Salah Penjurusan			
IPA	IPS	BHS	%
0	0	9	45%
1	3	1	6%
19	23	8	13%
105	131	10	35%
46	237	104	32%

Dengan demikian hasil penelitian dengan data yang sama maka hasilnya tidak stabil bahkan tidak sesuai dengan tujuan penjurusan pada SMA.

Performance Vector

PerformanceVector:
accuracy: 67.77% +/- 2.86% (mikro: 67.77%)

ConfusionMatrix:
Classification_error: 32.23% +/- 2.86% (mikro: 32.23%)

ConfusionMatrix

Jadi hasil daripada penelitian menggunakan 3 metode yang berbeda yaitu *Naive bayes*, *Decision Tree* dan *SVM (Support Vector Machine)* dan sebagai data testing adalah semua data nama-nama siswa SMA selama 3 tahun kurang 1089 siswa-siswi dari 3 jurusan.

Semua data tes tersebut di oleh sedemikian sehingga dapat seperoleh data penjurusan sesuai dengan yang diinginkan, menggunakan tool Rapid Miner sebagai berikut :

- a. Metode *Naive Bayes* (97,80 %)

- b. Metode *Decision Tree* (65,1 %)
 c. Metode SVM (*Support Vector Machine*) (67,77 %)

DAFTAR REFERENSI

Tabel. 12. Hasil 3 Metode Penelitian

No.	Data Tes	Hasil Penelitian		
		Naive bayes	Decision Tree	SVM SVM (Support Vector Machine)
1	19	100%	100%	55,00%
2	99	88,89%	93,94%	58,67%
3	399	80,95%	87,47%	61,87%
4	700	92,56%	64,81%	65,21%
5	1089	97,80%	65,01%	67,77%

Hasil dari peneliti ini dapat disimpulkan bahwa dengan 3 metode tersebut yang terbaik dan terakurat penelitiannya adalah **Naive Bayes**.

V. Kesimpulan

Dari hasil penelitian yang diperoleh yaitu :

1. *Naive bayes* data tes yang digunakan semakin banyak maka hasilnya semakin baik dan akurat daan sesuai dengan penjurusan pada sekolah SMA terkait.
2. *Decision tree* data tes semakin banyak hasilnya tidak relevan daan tidak sesuai dengan kriteria yang diharapkan.
3. SVM (*Support Vektor Machine*) bila data yang digunakan semakin banyak maka hasilnya tidak stabil karena data tes sedikit maupun banyak hasilnya tidak jauh berbeda. Artinya metode tersebut tidak sesuai untuk menentukan penjuruan di SMA.

Dari hasil 3 metode tersebut maka dapat peneliti simpulkan yang terbaik dan valid untuk menentukan penjurusan di sekolah SMA yaitu metode *Naive Bayes*.

- 1.Varghese Bindiya, dkk. Clustering Student Data to Characterize Performance Patterns. *International Journal of Advance Computer Science and Application*.138-140
- 2.Alpaydin, E. (2010). *Introduction To Machine Learning*. London: Massachusetts Institute Of Technology.
- 3.Abraham, A., Grosan, C., & Ramos, V. (2006). *Swarm Intelligence In Data Mining*. Verlag Berlin Heidelberg: Springer.
- 4.Aydin, I., Karakose, M., & Akin, E. (2011). A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Computer Engineering Department* , 120-129.
- 5.Berndtssom, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide For Students In Computer Science And Information Systems*. London: Springer.
6. Bramer, M. (2007). *Principles Of Data Mining*. Verlag London: Springer
- 7.Bromuri, S., Schumacher, M. I., & Ruiz, J. (2011). Monitoring Gestational Diabetes Mellitus With Cognitive Agents And Agent Environments. *Acm International Conferences On Web Intelligence And Intelligent Agent Technology* , 409-414.
8. Burges, C. J. (1998). *A Tutorial On Support Vector Machines For Pattern Recognition*. Boston: Kluwer Academic Publishers.
9. Dawson, C. W. (2009). *Projects In Computing And Information System A Student's Guide*. England: Addison-Wesley.
10. Dong, Y., Xia, Z., Tu, M., & Xing, G. (2007). An Optimization Method For Selecting Parameters In Support Vector Machines. *Sixth International Conference On Machine Learning And Applications* , 1.Fei, S. W., Miao, Y. B., & Liu, C. L. (2009). Chinese Grain Production Forecasting Method Based On Particle Swarm Optimization-Based Support Vector Machine. *Recent Patents On Engineering* 2009 , 3, 8-12.
- 11.Gomes, T. A., Prudenci, R. B., Soares, C., Rossi, A. L., & Andre, C. (2012). Combining meta-learning and search techniques to select parameters for support vectormachines. *T.A.F. Gomesetal./Neurocomputing* , 3-13

12. Gorunescu, F. (2011). *Data Mining Concepts, Models And Techniques*. Verlag Berlin Heidelberg: Springer
13. Larose, D. T. (2007). *Data Mining Methods And Models*. New Jersey: A John Wiley & Sons.
14. Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (Vol. Second Edition). New York Dordrecht Heidelberg London: Springer.
15. Moertini, V. S. (2002). Data Mining Sebagai Solusi Bisnis. *Integral, Vol. 7 No. 1, April 2002*, 4.
16. Nugroho, A. S. (2008). Support Vector Machine: Paradigma Baru Dalam Softcomputing. *Konferensi Nasional Sistem Dan Informatika*, 92-99.
17. Nuwangi, S., Oruthotaarachchi, C. R., Tilakaratna, J., & Caldera, H. A. (2010). Utilization Of Data Mining Techniques In Knowledge Extraction For Diminution Of Diabetes. *2010 Second Vaagdevi International Conference On Information Technology For Real World Problems*, 3-8.
18. Tamura, H., & Tanno, K. (2008). Midpoint Validation Method For Support Vector Machine With Margin Adjustment Technique. *The 3rd International Conference On Innovative Computing Information*, 1-4.
19. Vercellis, C. (2009). *Business Intelligence Data Mining And Optimization For Decision Making*. United Kingdom: A John Wiley And Sons, Ltd., Publication.
20. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And Techniques*. Burlington, Usa: Morgan Kaufmann Publishers.
21. Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature Selection And Parameter Optimization For Support Vector Machines: A New Approach Based On Genetic Algorithm With Feature Chromosomes. *School Of Computer Science And Technology*, 5197-5204.