

KAJIAN PENERAPAN MODEL C45, *SUPPORT VECTOR MACHINE (SVM)*, DAN *NEURAL NETWORK* DALAM PREDIKSI KENAIKAN KELAS

LUSI ARIYANI

Program Studi Teknik Informatika, Fakultas Teknik, Matematika, dan IPA
Universitas Indraprasta PGRI

Abstrak. Penilaian hasil belajar merupakan prediksi kenaikan kelas bagi siswa Sekolah Menengah Kejuruan (SMK) untuk melanjutkan ke tingkatan kelas berikutnya. Banyak mata pelajaran yang diujikan. Dari hasil mata pelajaran yang diujikan, maka sekolah dapat melihat rata-rata sehingga dapat diketahui siswa yang naik kelas atau tidak. Dengan cara tersebut siswa maupun guru atau wali kelas dapat memprediksi mata pelajaran dan nilai yang mempengaruhi kenaikan kelas. Selama ini Sekolah Menengah Kejuruan (SMK) Bina Taqwa belum memiliki pola-pola prediksi kenaikan kelas sebagai acuan untuk memprediksi jumlah siswa naik kelas dan yang tidak. Prediksi kenaikan kelas yang dilakukan saat ini masih manual, data yang diambil dari nilai semester siswa di akhir tahun. Prediksi hampir sama dengan klasifikasi yang akan terjadi di masa mendatang. Sehingga akan menghambat tentang peringkat sekolah dalam mengatasi kenaikan kelas siswa. Kendala tersebut dapat diatasi dengan analisis yang diuji menggunakan 3 buah metode algoritma yaitu algoritma C4.5, algoritma *Support Vector Machine* dan *Neural Network*. Dari hasil pengujian dengan mengukur kinerja ketiga algoritma tersebut diketahui bahwa algoritma C45 memiliki nilai *accuracy* paling tinggi. Sehingga dapat diterapkan untuk permasalahan prediksi kenaikan kelas.

Kata Kunci : Siswa, Kenaikan Kelas, Algoritma C4.5, Algoritma *Support Vector Machine*, *Neural Network*.

Abstract. Evaluation of the result from student's studies could be an expectation for the student to go to the next step to continue the next grade at vocational high school. Too many subject ate to be done by students. From the result of the subject which being tested, school can get the average, then school will decided their students can continue to the next grade or not. The prediction for decided about students can go to the next grade or not till this time still in manual and data takes by the result from the end of semester. All predection almost the same with classification which will happen in the future it can be a constraint for the school to manage the rank to solve how to decided about the rank level for the student. The constraint can be solved with analysis which using 3 algorithm C45, algorithm Support Vector Machine and Neural Network. From the result of the research with analysis three of them we'll know that algorithma Support Vector Machine have high in accuration. Then we can use in class to solve the predection problem about students up to the next grade.

Keywords: the students, next grade, Algorithm C45, Algorithm Support Vector Mechine, Neural Network.

PENDAHULUAN

Penilaian hasil belajar merupakan prediksi kenaikan kelas bagi siswa Sekolah Menengah Kejuruan (SMK) untuk melanjutkan ke tingkatan kelas berikutnya. Banyak mata pelajaran yang diujikan. Dari hasil mata pelajaran yang diujikan, maka sekolah dapat melihat rata-rata sehingga dapat diketahui siswa yang naik kelas atau tidak.

Untuk mengatasi masalah tersebut, maka diuji menggunakan 3 buah metode algoritma yaitu algoritma C4.5, algoritma SVM dan Neural Network. Dari ketiga buah metode tersebut akan dikaji metode mana yang paling akurat digunakan untuk mengukur kenaikan kelas pada siswa/I Sekolah Menengah Kejuruan (SMK) Bina Taqwa Cimanggis Depok.

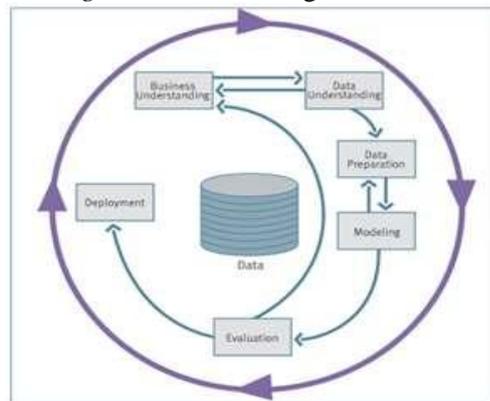
TINJAUAN PUSTAKA

Prediksi

Merupakan proses keilmuan untuk memperoleh pengetahuan secara sistematis berdasarkan bukti fisis. Ilmuwan melakukan pengamatan serta membentuk hipotesis dalam usahanya untuk menjelaskan fenomena alam. Prediksi yang dibuat berdasarkan hipotesis tersebut diuji dengan melakukan eksperimen. Jika suatu hipotesis lolos uji berkali-kali, hipotesis tersebut dapat menjadi suatu teori ilmiah.

Pengertian Data Mining

Data Mining adalah sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari keseluruhan proses, preprosesing data, pembentukan model, model evaluasi, dan tahap akhir penyebaran model (Larose, 2005). Gambar 1. merupakan proses *Data Mining* CRISP-DM sebagai berikut :



Gambar 1. Proses CRISP-DM (Larose, 2005)

1. *Business/Research Understanding Phase* (Fase Pemahaman Bisnis)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. *Data Understanding Phase* (Fase Pemahaman Data)
 - a. Mengumpulkan data
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan.
3. *Data Preparation Phase* (Fase Pengolahan Data)
 - a. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya.

- b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. *Modeling Phase* (Fase Pemodelan)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Kalibrasi atur model untuk mengoptimalkan hasil.
 - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
 - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan *data mining* tertentu.
 5. *Evaluation Phase* (Fase Evaluasi)
 - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan eektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
 6. *Deployment Phase* (Fase Penyebaran)
 - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Contoh sederhana penyebaran : Pembuatan laporan
 - c. Contoh kompleks penyebaran : Penerapan proses *data mining* secara parallel pada departemen lain.

Algoritma C4.5

Konsep Algoritma C4.5

C4.5 adalah algoritma yang digunakan untuk klasifikasi data yang dapat mengolah data/atribut numerik, algoritma ini dapat mengatasi nilai atribut yang hilang, dan dapat mengatasi data kontinyu dan *pruning*/penyederhanaan. Hasil dari proses klasifikasi berupa aturan yang dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dan secara umum digunakan untuk membangun pohon keputusan dengan melakukan tahapan sebagai berikut: pilih atribut sebagai akar (*root*), buat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang, kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah sebuah metode seleksi yang membandingkan parameter standar seperangkat nilai diskrit yang disebut kandidat set, dan mengambil salah satu yang memiliki akurasi klasifikasi terbaik (Dong, Xia, Tu, dan Xing, 2007). *Support Vector Machine* (SVM) adalah salah satu alat yang paling berpengaruh dan kuat untuk memecahkan klasifikasi (Burgess, 1998). Konsep *Support Vector Machine* (SVM) dapat dijelaskan secara sederhana sebagai usaha mencari *Hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada input *space*. Untuk *n*-dimensional space, input data x_i ($i=1, \dots, k$), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2.

Neural Network

Neural Network mempunyai relasi dengan synapse yang mengelilingi *neuron-neuron* lainnya. Susunan syaraf tersebut dipresentasikan dalam *Neural Network* berupa graf yang terdiri dari simpul (*neuron*) yang dihubungkan dengan busur, yang berkorespondensi dengan *synapse*. Sejak tahun 1950-an, *Neural Network* telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga regresi dengan atribut target *continue* (Vecellis, 2009:200).

Rapidminer 5.1

Rapid Miner dikhususkan untuk penggunaan data *mining*. Model yang disediakan juga cukup lengkap, seperti model *Bayesian Modelling*, *Tree Induction*, *Neural Network* dan lain-lain.

METODE

1. Penelitian Eksperimental

Penelitian eksperimental merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variabel atau atribut.

2. Penelitian Perbandingan atau Studi Komparasi yakni dengan membandingkan antara tiga macam algoritma yaitu algoritma *C4.5*, *SVM (Support Vector Machine)* dan *Neural Network*.

Pengumpulan Data

Mengumpulkan data dan informasi dilakukan metode pengumpulan data sebagai berikut:

1. Pengumpulan data primer

Melakukan wawancara langsung dengan pihak Sekolah SMK Bina Taqwa .

2. Pengumpulan data sekunder

Mengamati data, membaca, mempelajari dan mengutip dari buku literatur, serta sumber-sumber yang berhubungan erat dengan penelitian ini.

Instrument

1. Penelitian menggunakan Data sekunder berupa data siswa yang digunakan sebagai instrumentasi guna memperoleh data dalam proses penentuan prediksi kenaikan kelas.

2. Data disajikan dalam bentuk Tabulasi model dan variabel masing- masing sebanyak 236 siswa terdiri dari siswa kelas X, XI, dan XII tahun ajaran 2012/2013.

Teknis Analisis Data

Teknik Analisis data menggunakan Data Kuantitatif berupa kaidah-kaidah matematika terhadap anda atau numerik.

Dalam penelitian ini menggunakan model CRISP-DM (*Cross Standart Industries for Data Mining*).

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menentukan akurasi kelayakan prediksi kenaikan kelas yang dibandingkan dengan menggunakan metode algoritma *C4.5*, *Support Vector Mechine*, dan *Neural Network*. Setelah itu membandingkan nilai akurasi ketiga metode tersebut, dalam menentukan hasil penelitian ini menggunakan data *training* berjumlah 236 data dan data *testing* berjumlah 83 data.

Perhitungan Algoritma C4.5

Data *training* adalah untuk menentukan apakah seorang siswa/i naik kelas atau tidak naik kelas. Berikut akan dibahas prediksi apakah siswa/i naik kelas atau tidak naik kelas, menggunakan metode klasifikasi.

Langkah untuk membuat pohon keputusan, yaitu :

1. Tabel 3.3 adalah data *training* beserta kelasnya
2. Hitung nilai *entropy*. Dari data training diketahui jumlah kasus ada 236, siswa yang termasuk kelas Naik Kelas 231 *record* dan Tidak Naik Kelas 4 *record* sehingga didapat *entropy*:

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\ &= (-231/236 \cdot \log_2 (231/236)) + (-4/236 \cdot \log_2 (4/236)) \\ &= 0.1239 \end{aligned}$$

3. Hitung nilai *entropy* per atribut terlebih dahulu dengan rumus sama dengan di atas

1. Pekerjaan Orang Tua
 - a. Wiraswasta

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\ &= (-1/63 \cdot \log_2 (1/63)) + (-62/63 \cdot \log_2 (62/63)) \\ &= 0.1176 \end{aligned}$$

2. Kehadiran

- a. 0%-24%

$$\begin{aligned} Entropy(S) &= \sum_{i=1}^n -p_i \cdot \log_2 p_i \\ &= (-0/33 \cdot \log_2 (0/33)) + (-33/33 \cdot \log_2 (33/33)) \\ &= 0.0000 \end{aligned}$$

dst

4. Hitung nilai *gain* untuk tiap atribut, lalu tentukan nilai *gain* tertinggi. Yang mempunyai nilai *gain* tertinggi itulah yang akan dijadikan akar dari pohon. Misalkan untuk atribut status perkawinan = menikah, didapat nilai *gain*:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

- a. Perhitungan Gain status Pekerjaan Orang Tua

$$\begin{aligned} Gain(S,A) &= 0.1239 - (63/236(0.1176) + 142/236(0.1477) + \\ &31/236(0.0000)) \\ &= 0.0612 \end{aligned}$$

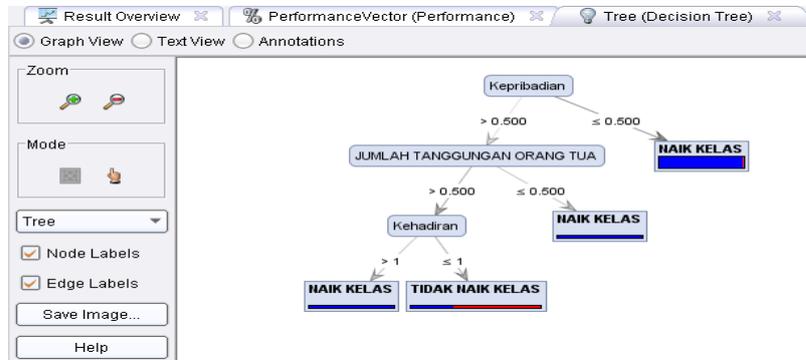
Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai gain tertinggi. Hasil perhitungan seluruh atribut terlihat pada Tabel 4.1.

Tabel 4.1 nilai *entropy* dan *gain* untuk menentukan simpul akar

Atribut	kasus	naik kelas	tidak naik kelas	entropy	gain
pekerjaan orang tua	236	232	4	0.1239	0.0037
Wiraswasta	63	62	1	0.1176	
Tetap	142	139	3	0.1477	
Tidak tetap	31	31	0	0.0000	
Kehadiran	236	232	4	0.1239	0.0094
0%-24%	33	33	0	0.0000	
25%-49%	17	16	1	0.3228	
50%-74%	24	24	0	0.0000	
>=75%	162	159	3	0.1330	
Orang tua	236	232	4	0.1239	0.0014
Ada	223	219	4	0.1297	
Yatim	13	13	0	0.0000	
Penghasilan orang tua	236	232	4	0.1239	0.0273
Tinggi	7	6	1	0.5917	
Sedang	48	48	0	0.0000	
Rendah	84	81	3	0.2223	
Sangat rendah	97	97	0	0.0000	
Jumlah tanggungan orang tua	236	232	4	0.1239	0.0157
Banyak	6	6	0	0.0000	
Sedang	125	121	4	0.2043	
Sedikit	105	105	0	0.0000	
Kepribadian	236	232	4	0.1239	0.0283
Baik	228	226	2	0.0725	
Cukup	7	5	2	0.8631	
Kurang	1	1	0	0.0000	
Kegiatan pengembangan diri	236	232	4	0.1239	0.0011
Baik	226	222	4	0.1283	
Cukup	10	10	0	0.0000	

Dari hasil perhitungan *entropy* dan *gain* yang didapat pada Tabel 4.1, terlihat bahwa atribut kepribadian mempunyai nilai gain tertinggi yaitu 0.0283. Oleh karena itu maka kepribadian merupakan simpul akar pada pohon keputusan.

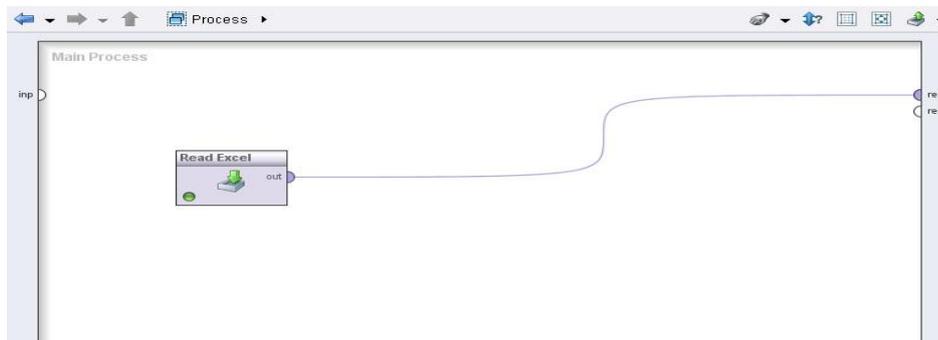
Penghitungan tidak dilakukan secara keseluruhan, namun hasil dari pohon keputusan akan terlihat setelah data siswa di aplikasikan pada rapid miner dan akan terlihat pada gambar 2.



Gambar 2. Pohon Keputusan hasil perhitungan dengan metode C4.5

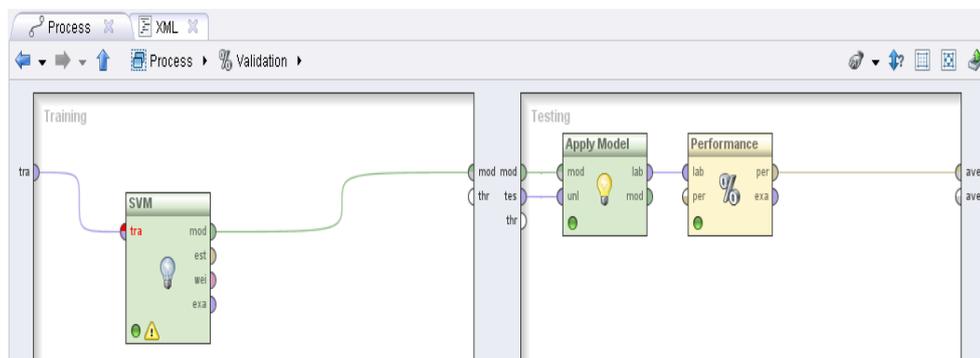
Algoritma Support Vector Machine

Gambar 3. adalah pembuatan model algoritma *Support Vector Machine* (SVM) diawali dengan pembacaan file data (*Read Excell*). *Data training* disimpan dalam satu file *Excell 2003*.

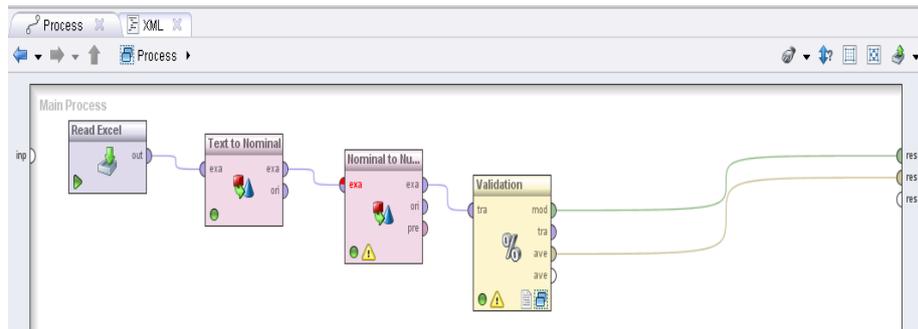


Gambar 3. Model Algoritma *Support Vector Machine* (SVM)

Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Support Vector Machine* (SVM).



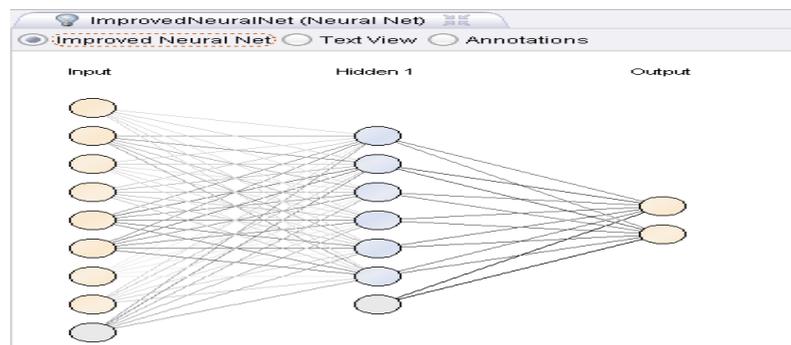
Gambar 4. Model Algoritma *Support Vector Machine* (SVM)
Data tersebut kemudian divalidasi



Gambar 5. Desain Model *Support Vector Machine*

Algoritma Neural Network.

Gambar 6. adalah *neural network* yang dihasilkan dari pengolahan data *training* dengan metode *neural network* adalah *multilayer perceptron* yang dihasilkan dari data *training*. Terdiri dari tiga *layer*, yaitu *Input layer* terdiri dari delapan simpul, sama dengan jumlah atribut prediktor ditambah satu simpul bias. Pada pembahasan ini digunakan satu *hidden layer* yang terdiri dari enam simpul ditambah satu simpul bias. Di bagian *output layer* terdapat dua simpul yang mewakili atribut kelas yaitu naik kelas dan tidak naik kelas.



Gambar 6. *Neural Network* yang dihasilkan dengan metode *neural network*

Untuk setiap data pada data *training*, dihitung input untuk simpul berdasarkan nilai input dan jaringan saat itu. Bobot awal untuk *input layer*, *hidden layer*, dan bias diinisialisasi secara acak. Simpul bias terdiri dari dua, yaitu pada *input layer* yang terhubung dengan simpul-simpul pada *hidden layer*, dan pada *hidden layer* yang terhubung pada *output layer*. Setelah semua nilai awal diinisialisasi, kemudian dihitung masukan, keluaran, dan *error*. Selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid. Setelah didapat nilai dari fungsi aktivasi, hitung nilai *error* antara nilai yang diprediksi dengan nilai yang sesungguhnya. Setelah nilai *error* dihitung, selanjutnya dibalik ke *layer* sebelumnya (*backpropagated*). Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi. Hasil perhitungan akhir *backpropagation* fungsi aktivasi untuk simpul pada *hidden layer* terdapat pada Tabel 1. Kolom pertama pada Tabel 1. merupakan atribut yang dinyatakan berupa simpul pada *input layer* seperti pada Gambar 6. Sedangkan Kolom satu sampai enam mewakili jumlah simpul pada *hidden layer*.

Tabel 1. Nilai bobot akhir untuk *hidden layer*

Atribut	Hidden Layer (Sigmoid)					
	1	2	3	4	5	6
Pekerjaan	0.588	0.725	0.577	0.603	0.620	0.665
Orang Tua						
Kehadiran	1.343	1.669	1.245	1.205	1.432	1.420
Orang Tua	-0.623	-0.798	-0.565	-0.553	-0.644	-0.634
Usia	0.901	1.107	0.968	1.002	0.998	0.970
Penghasilan	-1.703	-2.276	-1.681	-1.805	-1.991	-1.960
Orang Tua/Wali						
Jumlah	-1.914	-2.217	-1.784	-1.914	-2.058	-2.041
Tanggung Orang Tua						
Kepribadian	0.303	0.358	0.335	0.370	0.308	0.247
Remark	-0.657	-0.930	-0.589	-0.704	-0.820	-0.765
Threshold	1.339	1.776	1.328	1.391	1.497	1.452

Tabel 2. adalah nilai akhir fungsi aktivasi pada *output layer*. Kolom pertama pada Tabel 2. menyatakan *class*, yaitu atribut kelas yang dinyatakan dengan simpul pada *output layer* seperti pada gambar 6. Nilai yang terdapat pada kolom berlabel angka satu sampai enam adalah nilai bias terbaru yang terdapat pada relasi antara simpul pada *hidden layer* dan simpul pada *output layer*.

Tabel 2. Nilai Bobot Akhir untuk *Output Layer*

Class	output (sigmoid)						threshold
	1	2	3	4	5	6	
Naik	2.189	2.782	2.091	2.240	2.378	2.413	-4.078
Tidak	-2.191	-2.762	-2.110	-2.208	-2.385	-2.438	4.078
Naik							

Evaluasi dan Validasi

Dalam pembuatan model klasifikasi, dapat digunakan banyak metode. Dalam penulisan ini misalkan, metode yang digunakan, yaitu algoritma C4.5, SVM, dan *neural network*, setelah melakukan analisis menggunakan komparasi ketiganya dan mengukur metode mana yang paling akurat. Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas, dan interpretabilitas. Penelitian ini bertujuan untuk melihat akurasi analisis prediksi kenaikan kelas pada siswa dalam menilai kelayakan siswa dibandingkan dengan menggunakan algoritma C4.5, SVM, dan *neural network*, kemudian menganalisa akurasi dengan membandingkan ketiga metode tersebut.

Pengujian Model

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan data uji yang berasal dari data *training*. Data dalam penelitian ini 236 data maka digunakan metode *cross validation* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode C4.5 sebesar 96.83%, metode SVM sebesar 97.88%, dan metode *Neural Network* sebesar 97.35%.

1. Confusion Matrix

Tabel 1 adalah perhitungan berdasarkan data *training* pada Tabel 4.9, diketahui dari 236 data, 183 diklasifikasikan tidak *true* naik kelas sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 4 data diprediksi tidak naik kelas tetapi ternyata naik kelas, 0 data naik kelas diprediksi sesuai, dan 2 data diprediksi tidak naik kelas ternyata naik kelas.

Tabel 3. Model confusion Matrix untuk C45

	true NAIK KELAS	true TIDAK NAIK KELAS	class precision
pred. NAIK KELAS	183	4	97.88%
pred. TIDAK NAIK KELAS	2	0	0.00%
class recall	98.92%	0.00%	

Tabel 4.5 adalah *confusion matrix* untuk metode SVM. Diketahui dari 236 data, 185 diklasifikasikan naik kelas sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 4 data diprediksi tidak naik kelas tetapi ternyata naik kelas, 0 data tidak naik kelas diprediksi sesuai, dan 0 data diprediksi tidak naik kelas ternyata naik kelas.

Tabel 4. Model confusion matrik untuk metode SVM

	true NAIK KELAS	true TIDAK NAIK KELAS	class precision
pred. NAIK KELAS	185	4	97.88%
pred. TIDAK NAIK KELAS	0	0	0.00%
class recall	100.00%	0.00%	

Dengan metode *neural network*, menghasilkan kondisi seperti pada Tabel 4.7 Diketahui dari 236 data, 183 diklasifikasikan naik kelas sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 3 data diprediksi naik kelas tetapi ternyata tidak naik kelas, 1 data tidak naik kelas diprediksi sesuai, dan 2 data diprediksi tidak naik kelas ternyata naik kelas.

Tabel 5. Model Confusion Matrix untuk metode neural network

	true NAIK KELAS	true TIDAK NAIK KELAS	class precision
pred. NAIK KELAS	183	3	98.39%
pred. TIDAK NAIK KELAS	2	1	33.33%
class recall	98.92%	25.00%	

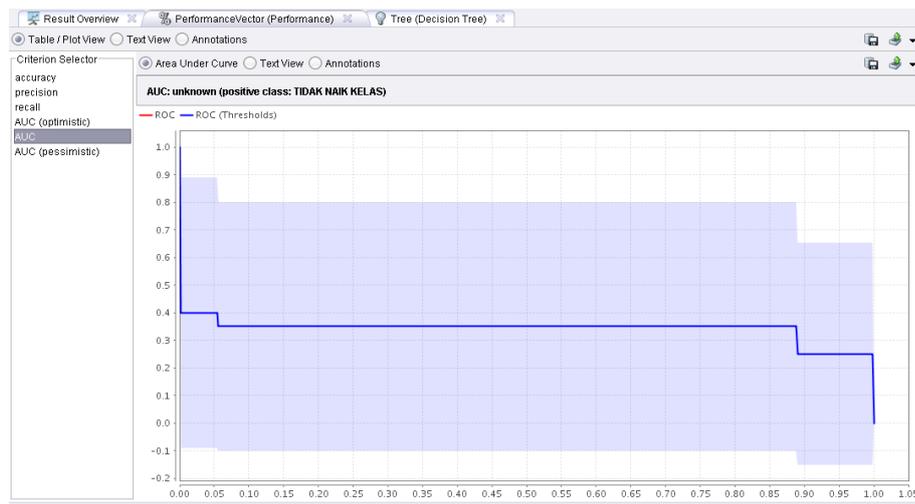
Dari tiga tabel *confusion matrix*, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, *sensitivity*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, *sensitivity*, dan *recall* yang telah dihitung untuk metode C4.5, SVM, dan *neural network* dapat dilihat pada Tabel 4.7.

Tabel 6. Komparasi Nilai Accuracy, Precision, dan Recall

	C4.5	SVM	Neural netork
Accuracy	96.83%	97.88%	97.35%
Precision	0.00%	0.00%	33.33%
Recall	0.00%	0.00%	25.00%

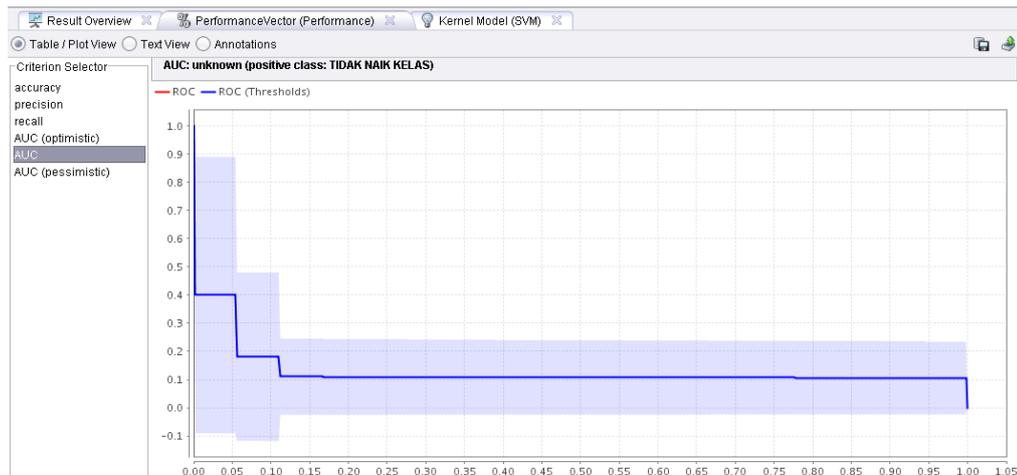
2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan ketiga metode komparasi bisa dilihat pada Gambar IV.6 yang merupakan kurva ROC untuk algoritma C45.



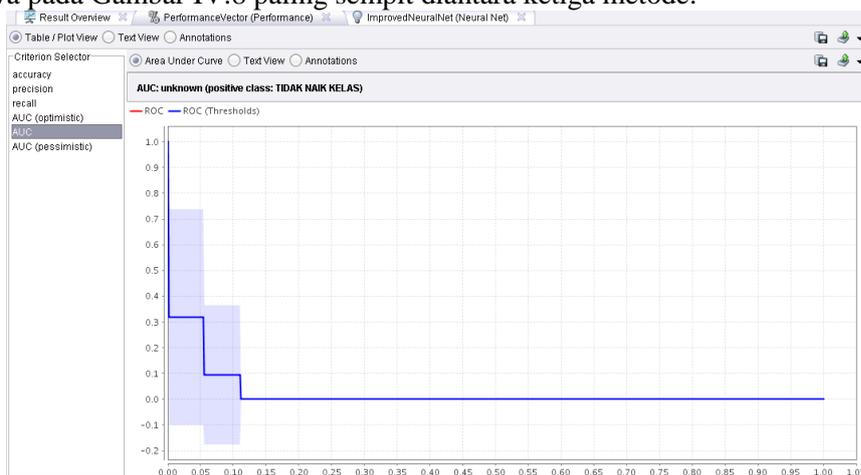
Gambar 7. Kurva ROC dengan algoritma C4.5

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan ketiga metode komparasi bisa dilihat pada Gambar IV.7 yang merupakan kurva ROC untuk algoritma SVM.



Gambar 8. Kurva ROC dengan Metode SVM

Seperti terlihat pada Gambar IV.6, Gambar IV.7, dan Gambar IV.8, area di bawah kurva pada Gambar IV.8 paling sempit diantara ketiga metode.



Gambar 9. Kurva ROC dengan Metode Neural Network

Perbandingan hasil perhitungan nilai AUC untuk metode C4.5, SVM, dan neural network dapat dilihat pada Tabel 4.8.

Tabel 7. Komparasi Nilai AUC

	C4.5	SVM	Neural Network
AUC	0.906	0.948	0.932

Analisis Hasil Komparasi

Model yang dihasilkan dengan metode C4.5, SVM, dan neural network diuji menggunakan metode Cross Validation, terlihat perbandingan nilai accuracy, precision, sensitivity, dan recall pada Tabel 4.9, untuk metode SVM memiliki nilai accuracy, precision, sensitivity, dan recall yang paling tinggi, diikuti dengan metode neural network, dan yang terendah adalah C45.

Tabel 8. Komparasi Nilai Accuracy dan AUC

	C4.5	SVM	Neural network
--	------	-----	----------------

<i>Accuracy</i>	96.83%	97.88%	97.35%
AUC	0.906	0.948	0.932

Tabel 4.9 membandingkan *accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy* dan AUC SVM paling tinggi. Untuk metode *neural network* dan C45 juga menunjukkan nilai yang sesuai.

Perancangan Sistem

Data baru diterapkan pada algoritma yang memiliki akurasi paling tinggi, dalam hal ini adalah algoritma SVM. Data baru yang digunakan sebanyak sepuluh record diuji dengan menggunakan confusion matrix dan diperoleh akurasi dan persisi sebesar 90%.

Tabel 9. Tabel Nilai Akurasi dan Persisi Data Baru Algoritma SVM

<i>Accuracy</i> : 90%			
<i>Precision</i> : 100%			
		Naik Kelas	Tidak Naik Kelas
Prediction Naik Kelas	7	1	87.50%
Prediction Tidak Naik Kelas	2	0	100%
Class recall	66.67%	100%	

Dari tabel 4.10 diketahui bahwa data baru yang digunakan sebanyak 10 record. Sebanyak 7 record diprediksi secara Naik Kelas dan 1 record diprediksi secara benar Tidak Naik Kelas. *Interface* sebagai implementasi hasil penelitian ini dirancang menggunakan bahasa pemrograman Java

Berikut ini rancangan model prediksi kenaikan kelas sebagai berikut:



Gambar 10. Tampilan implementasi model Prediksi Kenaikan Kelas

Implikasi Penelitian

Dari hasil evaluasi ternyata Algoritma SVM terbukti paling akurat dibanding Algoritma C4.5 dan *neural network*. Ketiga metode klasifikasi tersebut diterapkan pada data siswa yang naik kelas. Dengan hasil ini, menunjukkan bahwa metode *decision tree* merupakan metode yang cukup baik dalam pengklasifikasian data, dengan demikian algoritma SVM dapat memberikan pemecahan untuk permasalahan penentuan penerimaan siswa yang dapat naik kelas. Untuk mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen strategik, model ini dapat diterapkan pada sekolah menggunakan *software* RapidMiner. Penelitian semacam ini

dapat dikembangkan pada unit bisnis serupa atau yang lain. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode *naïve bayes*. Karena dunia bisnis, khususnya sekolah semakin berkembang maka kajian semacam ini dapat dilakukan secara baik.

PENUTUUP

Simpulan

Dari pengukuran kinerja ketiga algoritma yang telah dilakukan berdasarkan jumlah data maka dapat disimpulkan bahwa algoritma *Support Vector Machine* memiliki kemampuan dalam pengambilan keputusan untuk menentukan prediksi kenaikan kelas. Maka hasil penelitian dari percobaan yang telah dilakukan dapat disimpulkan bahwa:

1. Ketiga algoritma data mining (C4.5, *Support Vector Mechine*, dan *Neural Network*) dapat digunakan dalam prediksi kenaikan kelas. Ketiga algoritma ini dikomparasi kemudian diuji akurasi. Tingkat akurasi tertinggi lah yang digunakan dalam menentukan prediksi kenaikan kelas pada SMK Bina Taqwa Cimanggis - Depok.
2. Algoritma *Support Vector Machine* memiliki tingkat akurasi tertinggi diantara ketiga algoritma tersebut. Algoritma inilah yang di implementasikan dalam menentukan prediksi kenaikan kelas pada SMK Bina Taqwa Cimanggis – Depok.

Saran

Agar penerapan data mining berjalan lancar di SMK Bina Taqwa, maka penulis menyarankan hal hal sebagai berikut:

Menentukan kelompok data mining yang sesuai berdasarkan tugasnya, lalu menentukan algoritma yang sesuai dengan pengelompokan data mining sehingga hasil yang diharapkan lebih akurat dan efisien dalam penerapan sistem aplikasi data mining.

1. SMK Bina Taqwa merupakan sekolah yang memiliki data-data besar, oleh karena itu data-data yang banyak dapat di olah kemudian di carikan polanya sehingga menemukan pengetahuan baru agar data-data yang banyak tidak terbuang sia-sia. Sehingga data-data tersebut dapat dibuat berubah menjadi sebuah pengetahuan baru melalui data mining.
2. Implementasi data mining beserta algoritmanya tidak harus terpaku dengan algoritma yang sudah, misalnya *Support Vector Machine*. Penelitian berikutnya dapat dikembangkan dengan menggunakan algoritma lain.

DAFTAR PUSTAKA

- Burges C. J. 1998 Burges C. J. 1998. **A Tutorial On Support Vector Machines For Pattern Recognition**. Boston: Kluwer Academic Publishers.
- Dong. Y., Xia. Z., Tu. M., dan Xing. G. 2007 Dong. Y., Xia. Z., Tu. M., dan Xing. G. 2007. **An Optimization Method For Selecting Parameters In Support Vector Machines**. *Sixth International Conference On Machine Learning And Applications*, 1.
- Gorunescu F. 2011 Gorunescu F. 2011. **Data Mining Concepts, Models and Techniques**. Berlin Heidelberg: Springer Verlag.
- Han. J., dan Kember. M. 2006 Han. J., dan Kember. M. 2006. **Data Mining Concepts adn Techniques**. San Fransisco: Morgan Kauffman
- Huang. K., Yang. H., King. I., dan Lyu. M. 2008 Huang. K., Yang. H., King. I., dan Lyu. M. 2008. **Machine Learning Modeling Data Locally And Globally**.

- Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag Gmbh.
- Kusrini, dan Luthfi. T. E. 2009 Kusrini, dan Luthfi. T. E. 2009. **Algoritma Data Mining**. Yogyakarta: Penerbit Andi.
- Larose, Daniel. T. 2005. Larose, Daniel. T. 2005. **Discovering Knowledge in Data**. New Jersey: John Willey & Sons, Inc.
- Liao 2007 Liao 2007, **Recent Advances in Data Mining of Enterprise Data: Algorithms and Application**, World Scientific Publishing, Singapore.
- Maimon, Oded. 2005 Maimon, Oded. 2005, **Data Mining and Knowledge discovery Handbook**, Springer, Newyork.
- Mukhlis,Khusnul M. 2012 Mukhlis,Khusnul M. 2012, **Diagnosa Kemungkinan Pasien Terkena Stroke dengan menggunakan Metode Naïve Bayes dan Metode Jaringan Syaraf Tiruan Berbasis Web**, Surabaya.
- Myatt. G. J. 2007 Myatt. G. J. 2007. **Making Sense of Data A Practical Guide to Exploratory Data Analysis and Data Mining**. New Jersey: Wiley-Interscience.
- Nugroho. A. S. 2008 Nugroho. A. S. 2008. **Support Vector Machine: Paradigma Baru Dalam Softcomputing**. *Konferensi Nasional Sistem dan Informatika*, 92-99.
- Quinlan. 1993 Quinlan. 1993, Algoritma C4.5
- Sa'diyah N. N. A. 2012 Sa'diyah N. N. A. 2012, **Komparasi Algoritma C4.5, Naive Bayes, dan Neural Network Untuk Memprediksi Penyakit Jantung**.
- Vercellis C. 2009 Vercellis C. 2009. **Business Intelligent: Data Mining and Optimizzation for Decision Making**. Southern Gate, Chichester, West Sussex, United Kingdom : John Wiley & Sons Ltd.