

Jurnal Ilmiah

DASI

DATA MANAJEMEN DAN TEKNOLOGI INFORMASI



STMIK AMIKOM
YOGYAKARTA

VOL. 16 NO. 3 SEPTEMBER 2015
JURNAL ILMIAH
Data Manajemen Dan Teknologi Informasi

Terbit empat kali setahun pada bulan Maret, Juni, September dan Desember berisi artikel hasil penelitian dan kajian analitis kritis di dalam bidang manajemen informatika dan teknologi informatika. ISSN 1411-3201, diterbitkan pertama kali pada tahun 2000.

KETUA PENYUNTING

Abidarin Rosidi

WAKIL KETUA PENYUNTING

Heri Sismoro

PENYUNTING PELAKSANA

Kusrini

Emha Taufiq Luthfi

Hanif Al Fatta

Anggit Dwi Hartanto

STAF AHLI (MITRA BESTARI)

Jazi Eko Istiyanto (FMIPA UGM)

H. Wasito (PAU-UGM)

Supriyoko (Universitas Sarjana Wiyata)

Jano Hendarto (FMIPA-UGM)

Sri Mulyana (FMIPA-UGM)

Winoto Sukarno (AMIK "HAS" Bandung)

Rum Andri KR (AMIKOM)

Arief Setyanto (AMIKOM)

Krisnawati (AMIKOM)

Ema Utami (AMIKOM)

ARTISTIK

Amir Fatah Sofyan

TATA USAHA

Lya Renyta Ika Puteri

Murni Elfiana Dewi.

PENANGGUNG JAWAB :

Ketua STMIK AMIKOM Yogyakarta, Prof. Dr. M. Suyanto, M.M.

ALAMAT PENYUNTING & TATA USAHA

STMIK AMIKOM Yogyakarta, Jl. Ring Road Utara Condong Catur Yogyakarta, Telp. (0274) 884201 Fax. (0274) 884208, Email : jurnal@amikom.ac.id

BERLANGGANAN

Langganan dapat dilakukan dengan pemesanan untuk minimal 4 edisi (1 tahun) pulau jawa Rp. 50.000 x 4 = Rp. 200.000,00 untuk luar jawa ditambah ongkos kirim.

DAFTAR ISI

HALAMAN JUDUL.....	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
Perlindungan Data Terhadap Serangan Menggunakan Metoda Tebakan Pada Sistem Operasi Linux.....	1-8
Akhmad Dahlan (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Perlindungan Data Terhadap Serangan Menggunakan Metoda Tebakan Pada Sistem Operasi Linux.....	9-17
Ali Mustopa (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Integrasi Sistem Informasi Laboratorium Dengan Menggunakan Pendekatan <i>Service Oriented Architecture (Soa)</i>	18-26
Andika Agus Slameto (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Analisis dan Implementasi Algoritma Kriptografi Kunci Publik Rsa dan Luc Untuk Penyandian Data.....	27-36
Bayu Setiaji (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Kajian Infrastruktur Sistem Informasi Berbasis Sistem Multimedia.....	37-45
Dina Maulina (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Pemanfaatan Konsep Ontology Dalam Interaksi Sistem <i>Collaborative Learning</i>	46-52
Emigawaty (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Penerapan Algoritma <i>Learning Vector Quantization</i> Untuk Prediksi Nilai Akademis Menggunakan Instrumen Ams (<i>Academic Motivation Scale</i>).....	53-58
Hartatik (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Perancangan Sistem Audio On Demand Berbasis Jaringan Tcp/Ip di STMIK AMIKOM Yogyakarta.....	59-67
Hastari Utama (Teknik Informatika STMIK AMIKOM Yogyakarta)	
Analisis Perbandingan Aplikasi Web Berdasarkan <i>Quality Factors</i> dan <i>Object Oriented Design Metrics</i>	68-78
Jamal ¹⁾ , Ema Utami ²⁾ , Armadyah Amborowati ³⁾ (^{1,2})Magister Teknik Informatika, ³ Teknik Informatika STMIK AMIKOM Yogyakarta)	
Evaluasi Sumber Daya Teknologi Informasi di SMK Negeri 3 Magelang.....	79-86
Maria Harpeni Eko Meladewi ¹⁾ , Abidarin Rosidi ²⁾ , Hanif Al Fatta ³⁾ (^{1, 2, 3})Magister Teknik Informatika STMIK AMIKOM Yogyakarta)	

Uji Performa Implementasi Software-Based Openflow Switch Berbasis Openwrt Pada Infrastruktur Software-Defined Network.....	87-95
Rikie Kartadie ¹⁾ , Barka Satya ²⁾ (¹⁾ Teknik Informatika, ²⁾ Manajemen Informatika STMIK AMIKOM Yogyakarta)	
Analisis Keakuratan Metode Ahp dan Metode Saw Terhadap Sistem Pendukung Keputusan Penerimaan Beasiswa	96-100
Saifulloh ¹⁾ , Noordin Asnawi ²⁾ (^{1,2)} Teknik Informatika STT Dharma Iswara Madiun)	
Perbandingan Kinerja Algoritma Nbc, Svm, C 4.5 Dan Nearest Neighbor : Kasus Prediksi Status Resiko Pembiayaan Di Bank Syariah.....	101-106
Sumarni Adi (Teknik Informatika STMIK AMIKOM Yogyakarta)	

PERBANDINGAN KINERJA ALGORITMA NBC, SVM, C 4.5 DAN NEAREST NEIGHBOR : KASUS PREDIKSI STATUS RESIKO PEMBIAYAAN DI BANK SYARIAH

Sumarni Adi

*Teknik Informatika STMIK AMIKOM Yogyakarta
email: sumarniadi@gmail.com*

Abstract

Perkembangan teknologi informasi yang pesat telah mempengaruhi cara penilaian resiko pembiayaan yang semula dengan cara human judgment bergeser ke arah cara yang formal dan objektif yaitu melalui scoring pembiayaan. Banyak algoritma yang dapat membantu dalam membangun model klasifikasi pembiayaan. Pada perkembangan terbaru, teknik-teknik yang terdapat di dalam data mining mulai banyak digunakan khususnya teknik klasifikasi dan prediksi telah menjadi teknik yang populer. Ada beberapa algoritma dalam teknik klasifikasi diantaranya Naïve Bayes Classification (NBC), Support Vector Machine (SVM), C 4.5, dan Nearest Neighbor yang dapat digunakan untuk membangun model klasifikasi.

Kinerja algoritma dalam klasifikasi menjadi satu pertimbangan dalam pemilihan algoritma untuk memprediksi resiko pembiayaan. Penelitian ini mengukur kinerja dan membandingkan hasil pengukuran tingkat akurasi algoritma. Keempat algoritma tersebut menghasilkan akurasi model yang berbeda untuk dataset yang sama. Algoritma yang berbeda dapat memberikan keakuratan yang berbeda pula. Semakin tinggi nilai akurasi yang dihasilkan maka semakin akurat pula algoritma tersebut digunakan untuk prediksi pembiayaan nasabah berikutnya. Menggunakan cross validation, Algoritma NBC memberikan rata-rata tingkat keakuratan sebesar 66.95%, SVM sebesar 63.71%, C 4.5, sebesar 66.74 % dan Nearest Neighbor sebesar 63.03%. Sedangkan dengan bootstrap validation, Algoritma NBC memberikan rata-rata tingkat keakuratan sebesar 64.79%, SVM sebesar 61.25%, C 4.5, sebesar 65.91 % dan Nearest Neighbor sebesar 62.26 %.

Keywords:

Klasifikasi, Pembiayaan, NBC, SVM, C4.5, Nearest Neighbor

Pendahuluan

Salah satu pelayanan dalam dunia perbankan adalah pemberian pembiayaan kepada nasabah yang memenuhi syarat perbankan. pembiayaan merupakan sumber utama penghasilan bagi bank syariah dan juga sekaligus sumber resiko operasi bisnis terbesar, karena sebagian besar dana operasional bank diputar dalam bentuk pembiayaan [1].

Salah satu unsur dalam pembiayaan adalah adanya akad dan kesediaan membayar dari mudharib kepada shahibul maal [1]. Transaksi pem-bayaran pembiayaan tidak selamanya berjalan sesuai dengan perjanjian. Ketidاكلancaran pemba-yaran oleh mudharib dapat memunculkan pembiayaan bermasalah. Pembiayaan bermasalah ini dapat disebabkan oleh faktor eksternal seperti kondisi ekonomi yang tidak kondusif dan mudharib yang “nakal”, atau faktor internal yaitu kurang-mampuan pihak bank dalam menilai resiko calon mudharib. Faktor eksternal sulit dikontrol oleh pihak bank, namun faktor internal dapat dikontrol oleh pihak bank.

Perkembangan teknologi informasi yang pesat telah mempengaruhi cara penilaian resiko yang semula dengan cara *human judgment* bergeser ke arah cara yang formal dan objektif yaitu melalui *pembiayaan scoring*. Tujuan dari *scoring* pembiayaan ini adalah agar dapat diklasifikasi sehingga

membantu pihak bank mengkuantifikasi resiko finansial sehingga keputusan dapat diambil dengan cepat dan lebih akurat [2]. Banyak teknik *scoring* pembiayaan yang dapat membantu dalam pembangunan model klasifikasi. Pada perkembangan terbaru, teknik-teknik yang terdapat di dalam *data mining* mulai banyak digunakan. Khususnya teknik *scoring* telah menjadi teknik yang populer karena algoritma yang dihasilkan mudah diinterpretasikan dan divisualisasikan [2].

Ada beberapa algoritma *scoring* yang digunakan untuk klasifikasi yaitu NBC, SVM, C4.5, Nearest Neighbor untuk membangun model *klasifikasi*. Keempat algoritma tersebut menghasilkan akurasi yang berbeda untuk data set yang sama. Algoritma yang berbeda dapat memberikan keakuratan yang berbeda pula. Penelitian ini membahas perbandingan akurasi algoritma yang dihasilkan oleh performansi NBC, SVM, C 4.5 dan Nearest Neighbor untuk prediksi resiko pembiayaan.

Tinjauan Pustaka

Menurut [3] berdasarkan penelitiannya terhadap kemampuan Naïve Bayesian classifier untuk mengklasifikasi web pages dalam hal *binary classifications* dan *multi classifications*. Hasil penelitian

ini menunjukkan bahwa Bayesian *classifiers* memiliki kemampuan yang baik.

Sedangkan penelitian yang dilakukan oleh [4] menyatakan bahwa C4.5 lebih baik dengan akurasi yang mencapai 99,8 % dalam melakukan *information discovery* terhadap data pada data *medical record* untuk menentukan diagnosa penyakit cardiovascular (CVD). Pertanyaan yang kemudian muncul adalah bagaimana tingkat akurasi NBC jika dibandingkan dengan C4.5 dan juga jika dibandingkan dengan SVM dan *Nearest Neighbor* dalam menghasilkan model prediksi resiko pembiayaan?

Menurut [4] klasifikasi data memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian data class yang disebut *learned model*. Model tersebut dibangun dengan menganalisa *record* database. Setiap *record* diasumsikan menjadi *predefined class* yang ditentukan oleh suatu atribut yang disebut class label atribut. Akibat terdapat class label maka tahap ini juga dikenal dengan *supervised learning*. Selanjutnya, pada tahap kedua adalah tahap pengklasifikasian, model yang telah dihasilkan akan digunakan untuk melakukan klasifikasi terhadap *unknown* data (testing).

Pada proses *learned* model digambarkan dalam bentuk *classification rule* atau formula matematika yang biasa dikenal dengan algoritma.

1. Algoritma Naive Bayes Classifier

Naive bayes classifier (NBC) merupakan salah satu metode pada teknik klasifikasi dan termasuk dalam classifier statistik yang dapat memprediksi probabilitas keanggotaan class. NBC berprinsip pada teori bayes. NBC mengasumsikan bahwa nilai atribut pada sebuah class adalah independen terhadap nilai pada atribut yang lain. Naive bayes atau simple bayesian classifier memiliki prosedur sebagai berikut [4] :

1. Setiap sample data direpresentasikan dengan n-dimensional *feature vector*, $X=(X_1, X_2, \dots, X_n)$, dengan n dibuat dari *sample* n atribut, berturut-turut A_1, A_2, \dots, A_n .

2. Diandaikan terdapat m *class*, C_1, C_2, \dots, C_m . Diberikan sebuah data *sample*, X (yang tidak diketahui *class* labelnya), kemudian *classifier* akan memprediksi X ke dalam *class* yang memiliki probabilitas posterior tertinggi, Naive bayes *classifier* akan menentukan *sample* X ke dalam class C_i jika dan hanya jika

$$P(C_i | X) > P(C_j | X) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (1)$$

3. *Class* C_i adalah nilai terbesar, yang disebut dengan maksimum posterio hypothesis dengan teorema bayes :

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)} \quad (2)$$

4. $P(X)$ adalah konstan untuk semua *class*. Jika probabilitas *class* prior tidak diketahui, secara umum diasumsikan bahwa *class* adalah sama, yaitu $P(C_1)=P(C_2)=\dots=P(C_m)$, dan selanjutnya menghitung nilai $P(X|C_i)$ dan menghitung nilai $P(X|C_i)P(C_i)$. Probabilitas class prior diestimasi dengan $P(C_i) = \frac{s_i}{s}$, dimana s_i adalah jumlah training sample pada class C_i , dan s adalah jumlah *training sample*.

5. Apabila *dataset* terdiri dari banyak atribut, akan mengakibatkan komputasi yang rumit untuk menghitung $P(X|C_i)$. Untuk mengurangi komputasi, naive bayes mengasumsikan pada pembuatan *class independen*. Sehingga nilai pada atribut dikondisikan bersifat independen antara atribut yang satu dengan atribut yang lain, serta diantara atribut tidak terdapat relasi dependensi.

$$P(X | C_i) = \prod_k P(X_k | C_i) \quad (3)$$

6. Probabilitas $P(X_1 | C_i), P(X_2 | C_i), \dots, P(X_n | C_i)$, dapat diestimasi dari training sample, dimana

a. Jika A_k adalah kategorikal, maka $P(X_k | C_i)$

$$= \frac{s_{ik}}{s_i} \quad (4)$$

s_{ik} adalah jumlah dari training *sample* pada *class* C_i yang mempunyai nilai X_k untuk A_k dan s_i adalah jumlah *training sample* yang termasuk ke dalam *class* C_i .

b. Jika A_k bernilai kontinyu, maka diasumsikan mempunyai sebuah gaussian distribusi

$$P(x_k | C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) = \frac{1}{\sqrt{2\pi\sigma_{C_i}}} e^{-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2}}$$

$g(x_k, \mu_{C_i}, \sigma_{C_i})$ adalah fungsi gaussian untuk atribut A_k dengan μ_{C_i} dan σ_{C_i} adalah *mean* dan standard deviasi untuk atribut A_k pada *training sample class* C_i .

7. Untuk mengklasifikasikan *sample* X yang tidak diketahui, $P(X | C_i) P(C_i)$ dievaluasi untuk setiap *class* C_i . *Sample* X ditetapkan untuk *class* C_i jika dan hanya jika $P(C_i | X) > P(C_j | X)$ untuk $1 \leq j \leq m, j \neq i$

Dengan kata lain, ditetapkan sebagai *class* C_i untuk $P(C_i | X)$ yang bernilai maksimum.

2. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Langkah-langkahnya sebagai berikut [5] :

- Pilih variabel sebagai akar
- Buat cabang untuk tiap-tiap nilai
- Bagi kasus dalam cabang

- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk memilih variabel sebagai akar, didasarkan pada nilai *gain* tertinggi dari variabel-variabel yang ada. Untuk menghitung *gain* digunakan rumus pada Persamaan 5.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \dots (5)$$

Keterangan :

- S : Himpunan kasus
- A : Variabel
- n : jumlah partisi variabel A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

sementara itu untuk menghitung nilai entropi dapat dilihat pada Persamaan 6.

$$Entropy(S,A) = \sum_{i=1}^n - p_i \times \log_2 p_i \dots (6)$$

Keterangan :

- S : Himpunan Kasus
- A : Fitur
- n : jumlah partisi S
- pi : proporsi dari Si terhadap S

3. Algoritma *Nearest Neighbor*

Algoritma *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Adapun rumus untuk melakukan penghitungan kedekatan antara dua kasus seperti pada Persamaan 7 :

$$Similarity(T,S) = \frac{\sum_{i=1}^n f(T_i,S_i) \times w_i}{w_i} \dots (7)$$

Keterangan :

- T : Kasus Testing
- S : Kasus Training
- n : jumlah atribut dalam setiap kasus
- i : atribut individu antara 1s.d. n
- f : fungsi *similarity* atribut i antara kasus T dan kasus S
- w : bobot yang diberikan pada atribut ke-i

4. Algoritma Support Vector Machine (SVM)

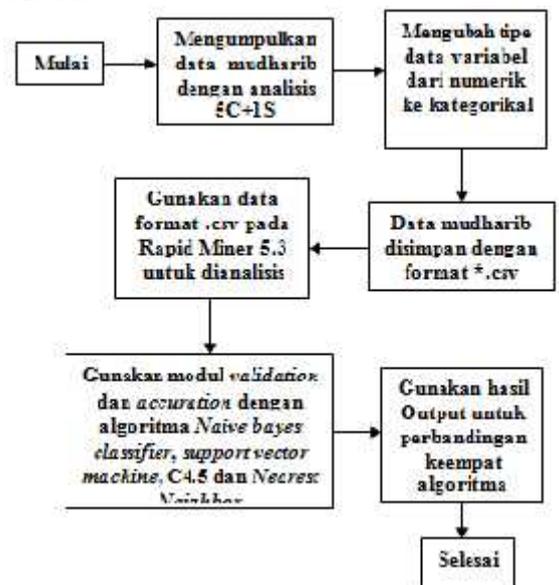
Support Vector Machine (SVM) termasuk dalam classifier statistik yang dapat memprediksi probabilitas keanggotaan class dengan cara optimasi dan akan bekerja dengan baik pada data dimensi tinggi [6]. Persamaan 8 merupakan persamaan optimasi problem.

$$L_1 = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \lambda_i (y_i (w \cdot x_i + b) - 1) \dots (8)$$

Metode Penelitian

Penelitian yang dilakukan untuk membandingkan algoritma NBC, SVM, C4.5 dan *Nearest Neighbor* yang ditunjukkan pada Gambar 1, terdiri atas 3 langkah yaitu :

1. Mengumpulkan data Nota Analisis pembiayaan mudharib jenis murabahah (konsumtif) di bank syariah kurun waktu 2005-2006, jumlah data-nya adalah 463 *record*. Data tersebut diekstraksi menggunakan analisis 5C+1S (*Character, Capacity, Capital, Collateral, Condition, dan Syariah*) untuk penentuan kelas aman atau tidak aman. Variabel yang diolah ada 5 yaitu data pendidikan, data gaji, data pekerjaan, data angsuran, dan data nilai jaminan.
2. Mengubah tipe data variabel yang masih numerik ke kategorikal, yaitu data gaji, data angsuran, dan data nilai jaminan
3. Setelah ekstraksi selesai kemudian diimport ke format CSV (*Comma Separated Value, *.csv*) untuk kemudian diolah menggunakan Rapid Miner 5.3
4. Data dalam format *.csv dapat dianalisis menggunakan modul Rapid Miner 5.3. Modul yang digunakan adalah modul *evaluation bootstrapping validaton* dan *modeling* dengan algoritma *Naïve bayes classifier, support vector machine, C4.5 dan Nearest Neighbor*. Algoritma diuji menggunakan tipe data kate-gorikal, kecuali SVM hanya dapat menggunakan tipe data numerik. Output yang dihasilkan mengandung persentase akurasi, presisi dan recall yang menjadi pembanding antara keempat algoritma tersebut.



Gambar 1. Skema Penelitian

Hasil dan Pembahasan

Data mudharib yang dikumpulkan dari nota analisis bank syariah dengan menggunakan analisis 5C+1S ditunjukkan pada gambar 2.

ID	Pendidikan	Pekerjaan	Gaji	Jaminan	Tempo	Pinjaman	Tujuan Pinjam	Angsuran	Marginal	Kelas
704390416	Sarjana	PNS Umum	2214500	100000000	00	1750000	RENOVASI RUMAH	520519.5412	17020172.47	aman
1050070629	Sarjana	PNS Umum	1398700	5000000	00	9650000	BARANG	1874241.77	15400005.07	aman
704390416	SMA	PNS Umum	3887000		00	14000000	MODAL USAHA	2950291.804	94500011.5	aman
1050070629	Sarjana	PNS Umum	1757500	5000000	00	4000000	BARANG	1194761.07	10010006.58	aman
1050070629	Sarjana	PNS Umum	2547700		00	5000000	BARANG	1228204.05	10750001.28	aman

Gambar 2. Data Mudharib hasil analisis 5C+1S

Data pada gambar 2 kemudian dilakukan konversi tipe data numerik ke kategorikal, yang ditunjukkan pada gambar 3.

Sebelum dikonversi, beberapa variabel dilakukan fungsi agregasi, yaitu variabel Tempo, Pinjaman, dan margin. Ketiga variabel ini dilakukan agregasi menjadi variabel angsuran, dengan cara menjumlahkan nilai pinjaman dengan nilai margin kemudian dibagi dengan nilai tempo.

ID	Pendidikan	Pekerjaan	Gaji	Angsuran	Nilai Jaminan	Kelas
1050070629	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
7044291538	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050060436	SMA	PNS Umum	Sedang	Sedang	Kecil	Tidak Aman
1050074990	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050049194	Sarjana	PNS Umum	Kecil	Kecil	Kecil	Aman
7043990416	Sarjana	PNS Umum	Sedang	Kecil	Kecil	Aman

Gambar 3. Data hasil konversi dari numerik ke kategorikal

Untuk variabel gaji dikategorikan menjadi 3 kategori yaitu:

1. Kecil, rasio Rp.1.000.000 – Rp.5.000.000
2. Sedang, rasio >5.000.000 – Rp.9.000.000
3. Besar, rasio >9.000.000 – Rp.13.000.000

Untuk variabel angsuran dikategorikan menjadi 3 kategori yaitu:

1. Kecil, rasio Rp.500.000 – Rp.4.500.000
2. Sedang, rasio >4.500.000 – Rp.8.500.000
3. Besar, rasio >8.500.000 – Rp.12.500.000

Untuk variabel nilai jaminan dikategorikan menjadi 3 kategori yaitu:

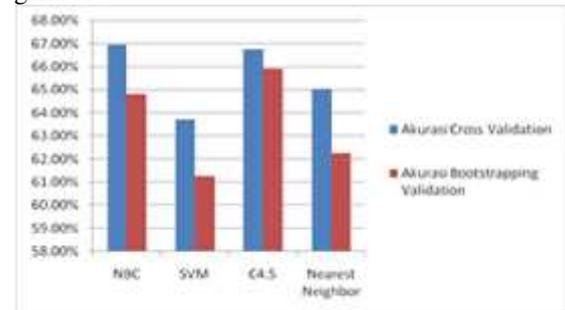
1. Kecil, rasio Rp 0 – Rp.100.000.000
2. Sedang, rasio >Rp.100.000.000 – Rp.200.000.000
3. Besar, rasio >Rp.200.000.000 – Rp.300.000.000

Hasil persentase perbandingan akurasi keempat algoritma menggunakan Rapid Miner 5.3 ditunjukkan pada Table 1.

Tabel 1. Persentase perbandingan akurasi keempat algoritma

Algoritma	Akurasi	
	Cross Validation	Bootstrapping Validation
NBC	66.95 %	64.79 %
SVM	63.71 %	61.25 %
C4.5	66.74 %	65.91 %
Nearest Neighbor	65.03 %	62.26 %

Dari Tabel 1 dapat digambarkan grafik hasil persentase perbandingan akurasi keempat algoritma menggunakan Rapid Miner 5.3 ditunjukkan pada gambar 4.

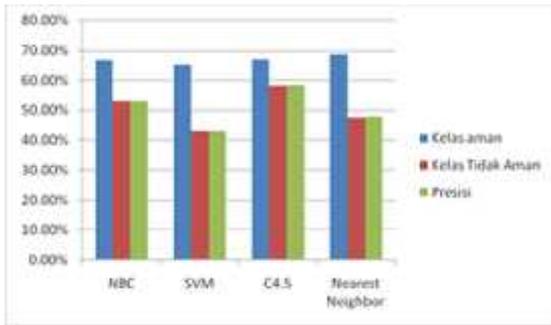


Gambar 4. Grafik perbandingan akurasi keempat algoritma

Pada Tabel 2 dirincikan presentase nilai presisi setiap kelasnya pada akurasi bootstrapping dari keempat algoritma, dan juga ditunjukkan dalam bentuk grafik pada Gambar 5.

Tabel 2. Presentase Presisi bootstrapping keempat algoritma

Algoritma	Kelas aman	Kelas Tidak Aman	Presisi
NBC	66.89 %	53.09 %	53.09 %
SVM	65.22 %	43.18 %	43.18 %
C4.5	67.03 %	58.21 %	58.50 %
Nearest Neighbor	68.66 %	47.47 %	48.01 %

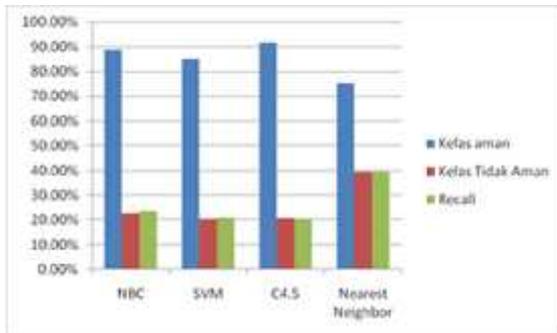


Gambar 5. Grafik Presentase Presisi bootstrapping keempat algoritma

Sedangkan pada Tabel 3 merupakan persentase nilai recallnya dan juga ditunjukkan dalam grafik pada Gambar 6.

Tabel 3. Presentase Recall bootstrapping keempat algoritma

Algoritma	Kelas aman	Kelas Tidak Aman	Recall
NBC	88.66 %	22.63 %	23.64 %
SVM	85.07 %	20.00 %	21.03 %
C4.5	91.64 %	20.53 %	20.22 %
Nearest Neighbor	75.22 %	39.47 %	39.57 %

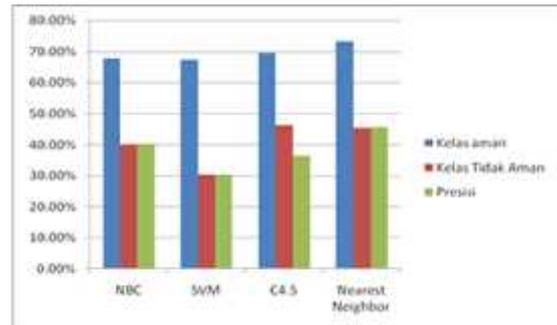


Gambar 6. Grafik Presentase Recall bootstrapping keempat algoritma

Pada Tabel 4 dirincikan presentase nilai presisi setiap kelasnya pada akurasi cross validation dari keempat algoritma, dan juga ditunjukkan dalam bentuk grafik pada Gambar 7.

Tabel 4. Presentase Presisi Cross Validation keempat algoritma

Algoritma	Kelas aman	Kelas Tidak Aman	Presisi
NBC	67.86 %	40.00 %	40.00 %
SVM	67.39 %	30.43 %	30.43 %
C4.5	69.44 %	46.30 %	36.52 %
Nearest Neighbor	73.23 %	45.65 %	45.75 %

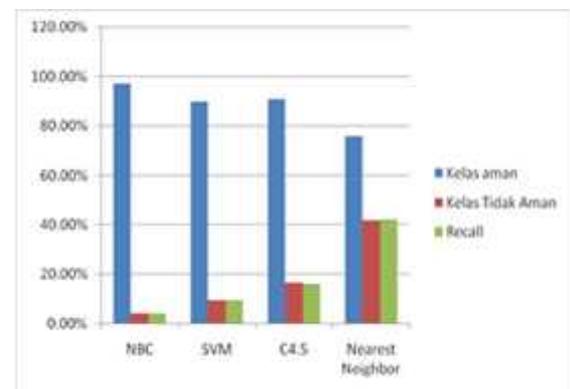


Gambar 7. Grafik Presentase Presisi Cross Validation keempat algoritma

Sedangkan pada Tabel 5 merupakan persentase nilai recall nya, dan juga ditunjukkan dalam bentuk grafik pada Gambar 8.

Tabel 5. Presentase Recall Cross Validation keempat algoritma

Algoritma	Kelas aman	Kelas Tidak Aman	Recall
NBC	97.12 %	4.00 %	4.05 %
SVM	89.78 %	9.33 %	9.35 %
C4.5	90.73 %	16.67 %	16.00 %
Nearest Neighbor	76.04 %	42.00 %	42.30 %



Gambar 8. Grafik Presentase Recall Cross Validation keempat algoritma

Dari Tabel 1 menunjukkan hasil akurasi model dari keempat algoritma bahwa algoritma NBC memiliki nilai presentase akurasi lebih tinggi dibandingkan dengan algoritma yang lain untuk jenis pengujian menggunakan *Cross Validation*. Sedangkan C4.5 memiliki nilai presentase akurasi lebih tinggi dibandingkan algoritma yang lain untuk jenis pengujian menggunakan *Bootstrapping Validation*.

Sementara dari Tabel 2 untuk presentase nilai presisi jenis pengujian *bootstrapping* lebih tinggi algoritma C4.5 dibandingkan dengan algoritma yang lain, sedangkan berdasarkan Tabel 3 nilai *recall*

C4.5 lebih rendah dibandingkan dengan algoritma lainnya.

Pada Tabel 4, untuk presentase nilai presisi jenis pengujian *Cross Validation* lebih tinggi algoritma Nearest Neighbor dibandingkan dengan algoritma yang lain, sedangkan berdasarkan Tabel 5 nilai *recall* NBC lebih rendah dibandingkan dengan algoritma lainnya.

Secara keseluruhan kinerja algoritma C4.5 lebih baik dibandingkan algoritma yang lainnya. Hal ini dapat dibuktikan dengan nilai presisi dan recall yang tetap stabil pada posisinya, sementara NBC bergeser dari posisi teratas ketika dilakukan uji presisi.

Kesimpulan dan Saran

Dari penelitian yang dilakukan dapat diambil kesimpulan :

1. Perbandingan keempat algoritma ini menggunakan dua model pengujian yaitu *Cross Validation* dan *Bootstrapping Validation*
2. Dari kedua model pengujian ini, model yang paling direkomendasikan adalah *Bootstrapping validation* karena pada model ini setiap data yang ada di data sampel mempunyai peluang yang sama untuk menjadi data training dan data testing. Hal ini terlihat dari nilai presisi dan recall yang tetap stabil pada algoritma C4.5 ketika menggunakan model pengujian *boot-strapping validation*.
3. Secara keseluruhan kinerja algoritma C4.5 lebih baik dibandingkan algoritma yang lainnya. Hal ini dapat dibuktikan dengan nilai presisi dan recall yang tetap stabil pada posisinya, sementara NBC bergeser dari posisi teratas ketika dilakukan uji presisi.

Dari penelitian ini, jika akan dikembangkan perlu diperhatikan :

1. Pengukuran kinerja sebuah algoritma data mining dapat dilakukan berdasarkan beberapa kriteria antara lain akurasi, kecepatan komputasi, robustness, skalabilitas dan interpretabilitas. Penelitian ini baru menggunakan dua kriteria yaitu berdasarkan akurasi dan interpretabilitas. Akan lebih baik jika semua kriteria diuji coba agar algoritma yang diteliti lebih teruji kinerjanya.
2. Akurasi sebuah algoritma bisa ditingkatkan dengan menggunakan beberapa teknik antara lain teknik bagging dan boosting. Penelitian ini belum menggunakan kedua teknik tersebut untuk meningkatkan akurasi karena penelitian ini hanya terbatas pada perbandingan algoritma NBC, SVM, C4.5 dan *Nearest Neighbor*
3. Penelitian ini juga menggunakan data sampel yang cukup terbatas yaitu 463 nasabah yang terdiri dari Mudharib level aman dan tidak aman. Dimana perbandingan jumlah level aman dan tidak aman tidak sama yaitu jumlah level aman

lebih banyak sedangkan level tidak aman sangat sedikit sekali. Untuk mengestimasi akurasi sebuah algoritma akan lebih baik jika jumlah data sampel yang digunakan mendekati populasi yang ada dan juga levelnya mempunyai jumlah perbandingan yang relatif sama. Diharapkan pada penelitian selanjutnya, data nasabah yang digunakan lebih banyak dan juga levelnya mempunyai jumlah perbandingan yang relatif sama dibandingkan penelitian ini agar pengklasifikasian data jauh lebih akurat.

Daftar Pustaka

- [1] Muhammad, 2005, Manajemen Pembiayaan BankSyari'ah, UPP AMP YKPN, Yogyakarta
- [2] Cyhe, K.H., Chin, T.W., dan Peng, G.C., 2004, Credit Scoring Using Data Mining Techniques, *Singapore Management Review*
- [3] Bie, R., Fu, Z., Sun, Q., & Chen, C., 2009, A Comparison Study of Bayesian Classifier on Web Pages Classification, *New Generation Computing*, 161-168
- [4] Han, J., & Kamber, M., 2006, *Data Mining Concept and Technique*, Morgan Kaufman Publisher, San Fransisco
- [5] Kusriani & Luthfi, T. E., 2009, Algoritma Data Mining, Andi Offset, Yogyakarta.
- [6] Tan, P. N., Stenbach, M., & Kumar, V., 2006, *Introduction to Data Mining*, Pearson Education, Boston.