

KOMPARASI KLASIFIKASI PENENTUAN *CUSTOMER* KREDIT MENGUNAKAN ALGORITMA C4.5 DAN KNN PADA PT CITRA SEMESTA ENERGY

Yosep Nuryaman

Program Studi Ilmu Komputer
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
<http://www.nusamandiri.ac.id>
yosepnuryaman@gmail.com



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract—PT. Citra Semesta Energy one of industry agent fuel provide credit product for old or new customer. Credit Product had caused some customers pay late. We use C4.5 and KNN algorithm to classification that customer. And the best result in this research is C4.5 algorithm with accuracy value 84% +/- and with AUC value 0,686+/-0,686. And then compared with KNN algorithm with accuracy value 84.86%+/-6.20 and AUC value 0.692 +/-0.121.

Intisari—PT. Citra Semesta Energy sebagai agen bbm industri menyediakan fasilitas kredit bagi customer lama maupun baru. Fasilitas kredit tersebut ternyata menimbulkan beberapa customer melakukan telat bayar. Peneliti mencoba menggunakan algoritma C4.5 dan KNN untuk mengklasifikasikan customer tersebut. Dan hasil pengklasifikasian terbaik dalam penelitian ini yaitu menggunakan algoritma C4.5 dengan hasil accuracy 84% +/-4,26 % serta nilai grafik AUC 0,686+/-0,686. Dan kemudian dibandingkan dengan algoritma KNN dengan nilai accuracy 84.86% +/-4,26%, serta nilai AUC 0,692+/-0,121.

Kata Kunci: klasifikasi, algoritma C4.5, KNN.

PENDAHULUAN

Permasalahan persaingan perusahaan keagenan bahan bakar minyak (BBM) industri PT. Pertamina Patra Niaga belakangan ini semakin bervariasi. Bukan hanya dari faktor internal seperti persaingan dalam hal mendapatkan pelanggan baru antar sesama agen BBM PT. Pertamina Patra Niaga dan keterlambatan pengiriman karena antrian pengisian (Marshel, Erwan, & Kadarini, 2013), (Hardiyana, 2016) namun juga berasal dari faktor eksternal (Sani &

Uhti, 2014) , (HRD, 2015) seperti semakin banyaknya Agen BBM dari minyak swasta (minyak non pertamina) yang hadir ditengah persaingan.

Atas hal tersebut, PT Citra Semesta Energy yang memiliki salah satu lini bisnis sebagai salah satu agen bbm industri PT. Pertamina Patra Niaga pada tahun 2015 mulai memberikan fasilitas kredit bukan hanya kepada beberapa customer lama, namun kepada beberapa customer baru.

Namun setelah lebih kurang 1 tahun fasilitas kredit ini berjalan, ternyata ditemukan beberapa customer kredit yang melakukan pembayaran tidak tepat waktu (Hadijah, 2016). Bukan hanya terlambat 1 hari atau 2 hari, namun ada yang terlambat hingga 1 bulan bahkan 1 tahun (HRD, 2015)

“Algoritma C4.5 adalah bagian dari algoritma untuk klasifikasi dalam pembelajaran machine learning dan data mining. C4.5 merupakan algoritma yang cocok digunakan untuk masalah klasifikasi pada machine learning dan data mining” (Aprilia, Baskoro, Ambarwati, & Wicaksana, 2013) dalam hal ini klasifikasi data yang dimaksud adalah dalam menentukan nasabah yang mengajukan kredit dengan tanpa berpotensi dan berpotensi gagal bayar.

Atas dasar hal tersebut untuk mengatasi permasalahan di PT Citra Semesta Energy, penulis mencoba menganalisis data yang ada menggunakan algoritma C4.5 yang kemudian dibandingkan dengan perhitungan menggunakan algoritma KNN untuk memberikan rekomendasi terbaik dalam pemberian kredit baik untuk customer lama maupun baru.

BAHAN DAN METODE

Sebagai bahan penelitian, penulis telah mengumpulkan data yang diperlukan dengan

metode penelitian yang penulis gunakan adalah sebagai berikut :

A. Pengumpulan Data

Penulis melakukan observasi langsung mulai tanggal 01 Mei - 31 Mei 2017, untuk mengumpulkan data daftar customer aktif dan tidak aktif selama tahun 2015 serta data transaksi penjualan selama tahun 2015 yang akan digunakan sebagai dasar penelitian. Rincian data yang terkumpul adalah pada table 1 berikut :

Tabel 1. Sampel

Bulan	Jumlah Transaksi
Januari	22
Februari	21
Maret	25
April	28
Mei	23
Juni	19
Juli	16
Agustus	28
September	27
Oktober	29
November	32
Desember	19
Jumlah	289

Sumber : (HRD, 2015)

B. Pengeolahan data awal

Pada tahap ini penulis melakukan *preprocessing* yaitu menghilangkan data-data yang tidak memiliki atribut yang lengkap. Selanjutnya yaitu menyamakan setiap atribut yang ada kedalam bentuk nominal.

C. Metode yang Digunakan

Metode yang digunakan dalam penelitian yaitu menggunakan CRISP-DM (Cross-Industry Standard Proses for Data Mining), yang terdiri dari 6 tahap yaitu :

1. Business/Research Understanding Phase

Fasilitas kredit awalnya dibuat sebagai salah satu cara untuk bersaing dengan perusahaan lainnya. Namun permasalahan muncul ketika beberapa customer terlambat dalam melakukan pembayaran. Penulis menggunakan metode Algoritma C4.5 untuk mengklasifikasikan customer yang mendapatkan fasilitas credit dengan tepat

2. Data Understanding Phase

Atribut-atribut yang menjadi parameter terlihat pada tabel 2 yaitu

Table 2. Atribut dan Nilai Katagori

No	Atribut	Nilai
1	Customer Lama/Baru	Lama

No	Atribut	Nilai Baru
2	Bidang	Industri Konstruksi Lainnya
3	Pengambilan	≤ 8 KL >8 KL
4	Tanpa/Dengan Broker	Tanpa Broker

Sumber : (Nuryaman, 2017)

3. Data Preparation Phase

Data yang diperoleh untuk penelitian ini sebanyak 289 transaksi penjualan, baik berasal dari customer aktif maupun customer tidak aktif. Setelah dilakukan *preprocessing* maka dari sebanyak 289 data transaksi hanya ada 231 data yang dapat digunakan dalam penelitian ini.

4. Modeling Phase

Tahap ini juga disebut tahap learning karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan keputusan. Pada penelitian ini, pembuatan model menggunakan algoritma C4.5 dan algoritma KNN.

5. Evaluation Phase

Pada tahap ini dilakukan pengujian model untuk mendapatkan informasi model yang akurat menggunakan confusion matrix.

6. Deployment Phase

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model ke proses penentuan customer kredit.

HASIL DAN PEMBAHASAN

Data yang sudah disiapkan kemudian diolah menggunakan model yang digunakan, dalam hal ini model yang digunakan adalah perhitungan menggunakan algoritma C4.5 dan algoritma KNN.

Perhitungan Algoritma C4.5

Langkah untuk menentukan pohon keputusan dengan menggunakan algoritma C4.5 dengan memakai data training sebanyak 231 data, yaitu :

a. Menyiapkan data training sebanyak 231 data yang digunakan dalam penelitian ini. Data training biasanya diambil dari data primer yang sebelumnya sudah dikelompokkan kedalam kelas-kelas tertentu.

b. Hitung nilai entropy dan gain

Setelah dilakukan perhitungan entropy dan gain didapat hasil nilai entropy seperti hitungan pada table 3 berikut:

Tabel 3. Nilai entropy dan gain untuk menentukan node awal

	Pelang	(s1)	(s2)	Entrop	Gain
Total	231	39	192	0.6550	
lamb					0.014
Lam	180	25	155	0.5813	
Baru	51	14	37	0.8478	
Jenis					0.018
Indu	175	28	147	0.6343	
Kons	45	11	34	0.8023	
Lain	11	0	11	0	
Peng					0.013
≤ 8KL	219	34	185	0.6228	
>8KL	12	5	7	0.9798	
Brok					0.023
Tan	35	12	23	0.9275	
Brok	196	27	169	0.5783	

Sumber : (Nuryaman, 2017)

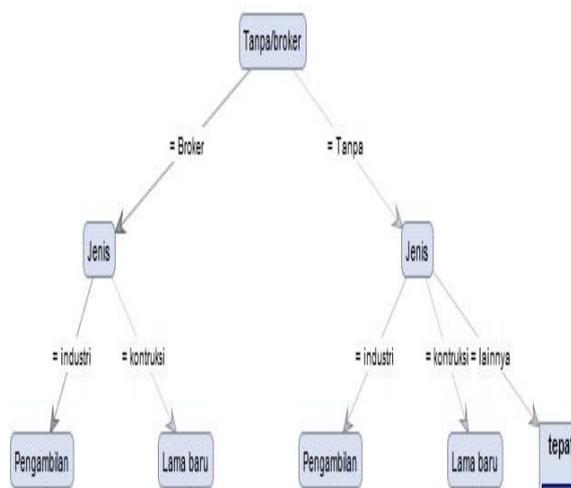
Dari tabel 3 dapat dilihat nilai gain tertinggi ada pada atribut tanpa/broker yakni 0,02379 sehingga dapat disimpulkan sebagai akar dari pohon keputusan atau node awal. Kemudian dilakukan kembali perhitungan nilai entropy dan gain untuk menentukan node 1.1, nilai yang dihitung berdasarkan atribut tanpa/broker dengan nilai tanpa dan broker. Perhitungan nilai entropy dan gain dapat dilihat dibawah ini.

Tabel 4. Nilai entropy dan gain untuk menentukan node 1.1

	Pelan	(s1)	(s2)	Entropi	Gain
Broker "tanpa"					
	35	12	23	0.927527	
Baru					0.00781
Lam	20	6	14	0.881291	
Baru	15	6	9	0.970951	
Jenis					0.37601
Ind	14	10	4	0.863121	
Kon	10	2	8	0.721928	
Lain	11	0	11	0	
pengam					0.25572
≤ 8KL	30	7	23	0.783777	
>8KL	5	5	0	0	

Sumber : (Nuryaman, 2017)

Setelah dilakukan perhitungan entropy dan gain hingga node terakhir maka akan menghasilkan pohon keputusan sebagai berikut :



Sumber : Nuryaman (2017)

Gambar 1. Pohon keputusan perhitungan manual

Pohon keputusan gambar 1, yang diperoleh dari perhitungan manual sama dengan pohon keputusan yang dihasilkan oleh rapidminer.

Dalam penelitian ini dilakukan evaluasi menggunakan confusion matrix dan menghasilkan nilai accuracy 84%, precision 55%, dan recall 26 %. dari data training dapat di evaluasi menggunakan rapid miner. Dan hasil evaluasi data yang ada adalah sebagai berikut

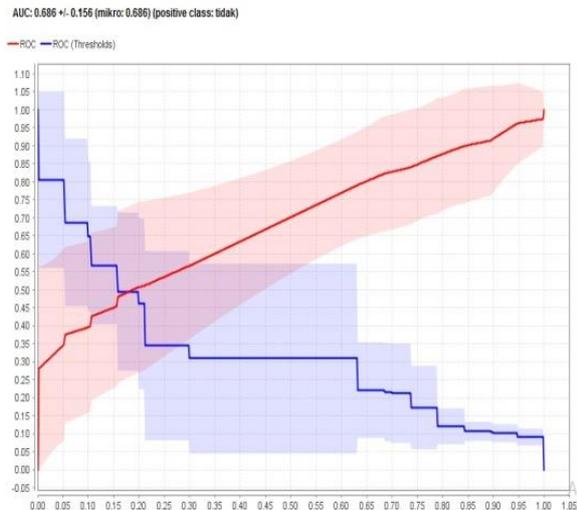
accuracy: 84.00% +/- 4.26% (mikro: 83.98%)

	true tepat	true tidak	class precision
pred. tepat	184	29	86.38%
pred. tidak	8	10	55.56%
class recall	95.83%	25.64%	

Sumber : Nuryaman (2017)

Gambar 2 Confusion Matrix (accuracy) data training

Sedangkan untuk untuk validasi dapat menggunakan curva ROC. Dan hasil validasi penelitian ini adalah



Sumber : (Nuryaman, 2017)
Gambar 4. Grafik ROC

Model yang dihasilkan dengan metode C4.5 terlihat pada nilai accuracy, peceision dan recall. Untuk klasifikasi nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu,2011)

- a) 0.90-1.00 = klasifikasi sangat baik
- b) 0.80-0.90 = klasifikasi baik
- c) 0.70-0.80 = klasifikasi cukup
- d) 0.60-0.70 = klasifikasi buruk
- e) 0.50-0.60 = klasifikasi salah

Perhitungan Menggunakan KNN

Untuk menghitung setiap jarak dari atribut yang ada, maka dilakukan pembobotan untuk setiap atribut yang ada.

Tabel 5. Pembobotan Atribut dan Sub Atribut

Atribut	BA	Sub Atribut	BSA
Lama/Baru	1	Lama	1,0
		Baru	0,5
Industri	2	Industri	3,0
		Konstruksi	1,5
		Lainnya	1
Pengambilan	3	≤ 8 KL	1,0
		> 8KL	0,5
Broker	4	Melalui Broker	1

Sumber : (Nuryaman, 2017)

Selanjutnya dilakukan penginputan bobot kedalam data training seperti tampilan pada tabel berikut.

Tabel 6. Pembobotan Data Training

No	Cust	Ind	Peng	Broker	label	
1	SURYA	1	3	1	1	tepat
2	MSP	0.5	1.5	1	1	tidak
3	CS2	0.5	3	0.5	0.5	tidak
4	ASE	1	1.5	1	1	tepat
5	CS2	1	3	1	1	tepat
6	CS2	0.5	3	0.5	0.5	tidak
7	SURYA	1	3	1	1	tepat
8	MSP	0.5	0.5	1	1	tidak
9	BCA	0.5	1	1	0.5	tepat
10	CS2	0.5	3	0.5	0.5	tidak

Sumber : (Nuryaman, 2017)

Untuk pengujian dapat ditarik data testing dan dihitung jarak dari setiap data training sesuai rumus yang ada.

Tabel 7 Perhitungan Jarak dan Rank

No	Parameter	Jarak	Label	Rank
1	AG-SUR	0,00	Tepat	1
2	AG-MSP	1,58	Tidak	8
3	AG-CS2	0,87	Tidak	4
4	AG-ASE	1,50	Tepat	7
5	AG-CS2	0,00	Tepat	2
6	AG-CS2	0,87	Tidak	6
7	AG-SUR	0,00	Tepat	3
8	AG-MSP	2,55	Tidak	10
9	AG-BCA	2,12	Tepat	9
10	AG-CS2	0,87	Tidak	5

Sumber : (Nuryaman, 2017)

Dari hasil diatas apabila ditentukan K=5 maka 3 data menyatakan tepat dan 2 data menyatakan tidak tepat, maka dapat ditarik kesimpulan bahwa apabila customer tersebut membeli akan membayar dengan tepat waktu.

Dan sebagai evaluasi untuk menentukan k terbaik dilakukan dengan metode confusion matrix dan curva AUC yang tampak seperti tabel dan gambar dibawah ini.

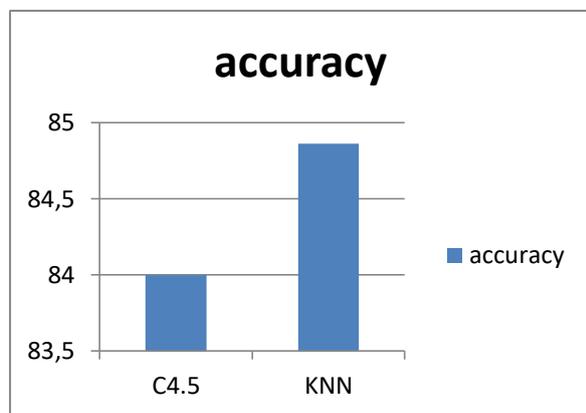
Tabel 8. Percobaan K terbaik

K	Accuraccy	AUC
K1	84.86%+/-6.20	0.500
K3	83.55%+/-6.38	0.673
K5	84.86%+/-6.20	0.692
K7	83.99%+/-7.01	0.692
K9	82.68%+/-6.15	0.675
K11	80.54%+/-5.11	0.658

Sumber : (Nuryaman, 2017)

Perbandingan Algoritma C4.5 dan KNN

Setelah semua metode dilakukan, maka hal selanjutnya adalah membandingkan antara hasil pengolahan menggunakan algoritma C4.5 dan KNN seperti yang terangkum pada gambar dibawah ini.



Sumber : (Nuryaman, 2017)
Gambar 4 Perbandingan Accuracy

<http://www.sttkd.ac.id/jurnal/index.php/JGH/article/view/30>

HRD, D. (2015). *Data Penjualan 2015*. Jakarta.

Marshel, S., Erwan, K., & Kadarini, S. N. (2013). Analisis Antrian SPBU (64 78118)(Studi Kasus Jalan Hasanudin Pontianak). *Jurnal Mahasiswa Teknk Sipil Universitas Tanjungpura*, 2(2), 1-10.

Nuryaman, Y. (2017). *Laporan Akhir Penelitian "Komparasi Klasifikasi Penentuan Customer Kredit Menggunakan Algoritma C4.5 Dan Knn Pada Pt Citra Semesta Energy."* Jakarta.

Sani, & Uchti. (2014). KONTRIBUSI KECEPATAN LAYANAN DAN TARIF TERHADAP KEPUASAN KONSUMEN PT.JNE KONTER PERAK TIMUR. *Jurnal Pendidikan Tata Niaga (JPTN)*, 2(1). Retrieved from <http://jurnalmahasiswa.unesa.ac.id/index.php/jptn/article/view/7069>

KESIMPULAN

Dari hasil percobaan yang sudah dilakukan dengan hasil metode klasifikasi menggunakan algoritma C4.5 memberikan hasil akurasi 84% dan AUC 0.686 sedangkan KNN dengan akurasi 84.86% dan AUC 0.692 maka dapat disimpulkan algoritma KNN mampu memberikan nilai evaluasi sedikit lebih baik dibandingkan algoritma C4.5 dalam menentukan klasifikasi customer kredit.

REFERENSI

Aprilia, D., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Miner*. Jakarta: Gramedia Pustaka Utama.

Hadijah, S. (2016). Penggolongan Kualitas Kredit dan Cara Menghindari Kredit Macet - Cermati. Retrieved August 13, 2018, from <https://www.cermati.com/artikel/penggolongan-kualitas-kredit-dan-cara-menghindari-kredit-macet>

Hardiyana, H. (2016). PENANGANAN PENUMPANG YANG MENGALAMI PENUNDAAN KEBERANGKATAN AKIBAT RETURN TO APRON PADA PENERBANGAN ROYAL BRUNEI AIRLINES DI PT. GAPURA ANGKASA CABANG BANDARA INTERNASIONAL JUANDA-SURABAYA. *Jurnal Ground Handling*, 1(1). Retrieved from

