# PERBANDINGAN ALGORITMA C4.5 DAN NAIVE BAYES UNTUK PREDIKSI POLA PEMBAYARAN ANGSURAN KREDIT MOTOR

**(Studi Kasus : PT SUMMIT OTO FINANCE)**

[1]*Siska Agustin Rahayu (1410651055)*

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

## ABSTRAK

Tunggakan pembayaran kredit merupakan salah satu permasalahan bagi perusahaan pembiayaan yang menimbulkan kecil atau lambatnya perputaran modal kerja, sehingga semakin kecil pula kemampuan perusahaan dalam meningkatkan volume pemberian kredit. Perusahaan PT Summit Oto Finance juga memiliki permasalahan tentang pola pembayaran angsuran salah satunya adalah adanya kredit yang tidak lancar. Penelitian ini dilakukan untuk mencari pola pembayaran yang lancar dan tidak pada PT Summit Oto Finance, dengan melakukan klasifikasi perbandingan algoritma C4.5 dan Naive Bayes untuk prediksi pola pembayaran angsuran kredit motor pada *customer* perusahaan. Penelitian ini menggunakan total dataset sebanyak 300 data. Data tersebut terbagi menjadi 240 data (80%) sebagai data latih dan uji. Selanjutnya, sisa datanya digunakan sebagai data validasi sebesar 60 data (20%), dengan parameter validasinya menggunakan k-fold cross validation. Dari hasil penelitian ini akurasi tertinggi pada metode C4.5 sebesar 96,66% dan presisi 100%, didapat pada skenario ke 8 uji ke 4. Sedangkan akurasi tertinggi pada metode naive bayes sebesar 93,33% dan presisi 100%, didapat pada skenario ke 8 uji ke 6. Model terbaik yang dihasilkan oleh metode C4.5, selanjutnya diuji ulang dengan data validasi dan menghasilkan nilai akurasi sebesar 76,66% serta presisi sebesar 86,8%.

## PENDAHULUAN

### Latar Belakang

Perusahaan PT Summit Oto Finance awalnya bernama PT Summit Sinar Mas Finance, hasil kerjasama usaha antara PT Sinar Mas Multiartha dan Sumitomo Corporation, Jepang. Awalnya PT Summit Sinar Mas Finance memfokuskan aktivitas usaha pada sewa guna usaha. Namun di tahun 2003 PT Summit Sinar Mas Finance mengubah aktivitas usahanya menjadi perusahaan pembiayaan kendaraan bermotor secara kredit, sekaligus mengganti namanya menjadi PT Summit Oto Finance. Permasalahan yang sering dihadapi adalah sejumlah konsumen yang pembayarannya memiliki tunggakan. Tunggakan pembayaran kredit merupakan salah satu permasalahan bagi perusahaan pembiayaan yang bisa menimbulkan kecil atau lambatnya perputaran modal kerja, sehingga semakin kecil pula kemampuan

perusahaan dalam meningkatkan volume pemberian kredit. Sebuah perusahaan pembiayaan pastilah mempunyai data yang begitu besar, banyak yang belum menyadari bahwa dari pengolahan data tersebut dapat memberikan informasi berupa klasifikasi data lancar atau tidaknya pembayaran cicilan kredit. Klasifikasi merupakan salah satu metode data mining yang dapat digunakan untuk mengolah data sehingga mendapatkan pengetahuan, hasil atau kesimpulan yang dapat membantu memecahkan masalah. Dalam penelitian ini metode klasifikasi alternatif belum diketahui untuk sebuah rujukan dalam mengklasifikasi customer sehingga menentukan metode alernatif juga menjadi pokok permasalahan dalam penelitian ini khususnya peneliti.

**Tujuan Penelitian**

Tujuan penelitian pada tugas akhir ini adalah :

1. Membandingkan nilai Presisi dari hasil perhitungan Algoritma C4.5 dan Naïve Bayes
2. Membandingkan nilai Akurasi dari hasil perhitungan Algoritma C4.5 dan Naïve Bayes

**Manfaat Penelitian**

Manfaat yang diperoleh dengan tercapainya tujuan penelitian ini diantaranya yaitu:

1. Dapat menentukan algoritma terbaik atau dengan nilai akurasi dan presisi tertinggi dari algoritma C4.5 dan Naive Bayes.
2. Sebagai alternatif algoritma untuk menentukan pola pembayaran angsuran kredit motor yang lancar dan tidak lancar.
3. Dapat digunakan menjadi sarana dalam proses pengambilan keputusan yang berguna bagi perusahaan di masa yang akan datang.

**METODE PENELITIAN**

**Implementasi & Pengujian**

Pada implementasi ini bertujuan untuk pengerjaan dan penerapan pada data yang telah diperoleh, dengan menggunakan algoritma C4.5 dan *Naive Bayes*. Untuk menghitung nilai agar di dapatkan hasil yang di inginkan maka atribut yang digunakan yaitu umur, status, penghasilan, angsuran, jangka waktu dan label. Dan pada pengujian ini menerapkan metode C4.5 dan *Naive Bayes*, dengan evaluasi model klasifikasi untuk memprediksi obyek yang benar dan salah menggunakan perhitungan *Confusion Matrix*. Parameter validasinya menggunakan *k-fold cross validation*, dengan nilai akhir implementasi dan pengujian data yaitu nilai perbandingan dan

akurasi metode C4.5 dan *Naive Bayes* untuk menentukan pola konsumen kredit.

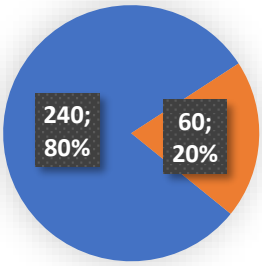
**Perhitungan Akurasi dan Presisi**

Tahap ini adalah tahap evaluasi dari hasil yang diperoleh dari konsep algoritma C4.5 dan algoritma *Naive Bayes*. Hasil perhitungan sementara dari 20 data *Training* adalah sebagai berikut :

1. Algoritma C4.5
   a. *Akurasi* $= \frac{TP + TN}{TP+TN+FP+FN} = \frac{18+2}{18+2+0+0} = 1 * 100 = 100\%$
   b. *Presisi* $= \frac{TP}{TP+FP} = \frac{18}{18+0} = 1 * 100 = 100\%$
2. Algoritma Naive Bayes
   c. *Akurasi* $= \frac{TP + TN}{TP+TN+FP+FN} = \frac{18+2}{18+2+0+0} = 1 * 100 = 100\%$
   d. *Presisi* $= \frac{TP}{TP+FP} = \frac{18}{18+0} = 1 * 100 = 100\%$

**HASIL DAN PEMBAHASAN**

Dari total 300 data akan dipartisi menjadi tiga bagian yaitu data uji, data latih dan data validasi. Data uji dan data data latih akan masuk pada skenario uji menggunakan K-Fod Cross Validation sedangkan sisanya akan dijadikan data validasi. Berikut rincian dari partisi data

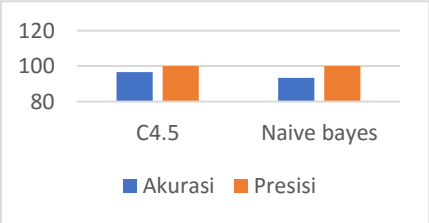



Partisi data Dari total 300 data akan dijadikan data validasi 20% atau dengan jumlah 60 data. 240 data akan dijadikan data uji dan latih. Menggunakan skenario K-Fold Cross Validation dengan nilai k = 2, 4, 5, 6, 8 dan 10.

Dalam k = 2, dari data 240 dapat dipartisi yaitu 120 data uji dan 120 data latih dengan dua kali pergantian data. Dalam k = 4, dari 240 data dapat dipartisi yaitu 60 data uji dan 180 data latih dengan empat kali pergantian data. Dalam k = 5, dari data 240 dapat dipartisi yaitu 48 data uji dan 192 data latih dengan lima kali pergantian data. Dalam k = 6, dari 240 data dapat dipartisi yaitu 40 data uji dan 200 data latih dengan enam kali pergantian data. Dalam k = 8, dari 240 data dapat dipartisi yaitu 30 data uji dan 210 data latih dengan delapan kali pergantian data. Dalam k = 10, dari 240 data dapat dipartisi yaitu 24 data uji dan 216 data latih dengan sepuluh kali pergantian data.

**Analisis**

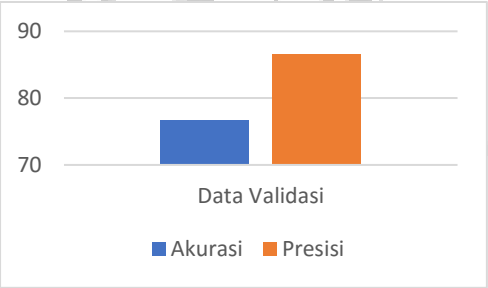
Dari serangkaian uji coba metode naive bayes dan C4.5 dengan skenario K-Fold Cross Validation pada data yang sama didapat hasil akurasi dan presisi tertinggi pada kedua metode tersebut yang tergambar pada Gambar 4.2.



Hasil akurasi dan presisi tertinggi dari naive bayes dan C4.5

Nilai akurasi dan presisi tertinggi pada metode C4.5 didapat pada cross fold validation dengan k = 8 dan pada uji coba keempat. Sedangkan nilai akurasi dan presisi tertinggi pada metode naive bayes didapat pada cross fold validation dengan k = 8 dan pada uji coba keenam. Sehingga dapat disimpulkan bahwa metode C4.5 lebih baik daripada naive bayes dalam kasus ini.

Sehingga data latih yang akan digunakan dalam mengklasifikasikan data validasi dengan jumlah 60 data yang terpartisi awal adalah data latih pada cross fold validation pada k = 8 di uji coba keempat. Berikut adalah hasil uji coba dari data latih k = 8 di uji coba keempat pada data validasi yang tergambar pada grafik 4.3.




Hasil Skenario k = 8 di uji keempat terhadap data validasi

Dari grafik diatas dapat dilihat bahwa nilai akurasi dari uji coba data validasi ini 76.66%, terdapat penurunan 20% nilai akurasi dari hasil uji coba cross fold validation pada k = 8 uji keempat. Dan dari grafik diatas dapat dilihat nilai preisi yang dihasilkan dari uji terhadap data validasi adalah 86.48%, terdapat penurunan 13.51% nilai presisi dari hasil uji coba cross fold validation pada k = 8 uji keempat.

Rangkuman Hasil

| Skenario Uji C4.5 | | | Skenario Uji Naive Bayes | | |
|---|---|---|---|---|---|
| Uji Coba ke | Akurasi | Presisi | Uji Coba ke | Akurasi | Presisi |
| 2.1 | 82.5% | 90.36% | 2.2 | 85.83% | 86.73% |
| 4.2 | 91.66% | 95% | 4.3 | 85% | 88.09% |
| 5.3 | 95.83% | 97.14% | 5.2 | 87.5% | 93.93% |
| 6.3 | 95% | 100% | 6.3 | 90% | 93.10% |
| 8.4 | 96.66% | 100% | 8.6 | 93.33% | 100% |
| 10.5 | 95.83% | 100% | 10.5 | 91.66% | 94.11% |

## KESIMPULAN DAN SARAN

### Kesimpulan

Dari hasil penelitian perbandingan metode C4.5 dan Naive Bayes terhadap klasifikasi data konsumen kredit motor dengan total 300 data dan skenario uji menggunakan k-fold cross validation dalam penelitian ini

menggunakan nilai k sebesar 2, 4, 5, 6, 8 dan 10 adalah :

1. Nilai akurasi terbesar pada metode C4.5 adalah 96.66% dan nilai presisi terbesar adalah 100%. Kedua nilai ini didapatkan pada skenario uji k-fold cross validation pada k = 8 uji keempat.
2. Nilai akurasi terbesar pada metode Naive Bayes adalah 93.33% dan nilai presisi terbesar adalah 100%. Kedua nilai ini didapatkan pada skenario uji k-fold cross validation pada k = 8 uji keenam.
3. Nilai akurasi yang didapatkan dari uji terhadap data validasi adalah 76.66% dan nilai presisi 86.48%. Hal ini mengalami penurunan akurasi sebesar 20% dan penurunan presisi sebesar 13.51%.
4. Metode C4.5 dalam penelitian kasus ini memiliki keunggulan dibandingkan metode Naive Bayes dari segi akurasi.

**Saran**

Penelitian ini jauh dari kesempurnaan. Maka dari itu dibutuhkan kritik dan saran untuk perbaikan serta pengembangan pada penelitian ini kedepannya. Berikut hal-hal yang dapat dikembangkan kedepannya :

1. Penelitian selanjutnya dapat menggunakan metode selain C4.5 dan Naive Bayes untuk mendapatkan nilai akurasi dan presisi yang lebih baik.
2. Jika nilai akurasi dan presisi sudah bisa digunakan, penelitian selanjutnya dapat membuat sistem informasi atau program serta sistem prediksi dari kasus ini.

**DAFTAR PUSTAKA**


Adisaputra, G. (2003). Anggaran Perusahaan (Cetakan Kedua).

Basuki, A. (2006). Metode Bayes. *Institute Teknologi Sepuluh November*.

Ciptohartono, C. C. (2014). Algoritma Klasifikasi Naive Bayes untuk Menilai Kelayakan Kredit. *Fakultas Ilmu Komputer Udinus*.

Han, J., & Kamber, M. (2006). Classification and prediction. *Data mining: Concepts and techniques*, 347-350.

Kusrini, E. T. L. (2009). Data Mining Algorithms. *Publisher ANDI*.

Masripah, S. (2016). Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit. *Bina Insani ICT Journal*, *3*(1), 187-193.

Oktaviana, A. R. (2016). Penerapan Data Mining Klasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank BRI Batang. *Skripsi Universitas Dian Nuswantoro*.

Pratikto, F. R. (2009). Evaluasi Pemohon Kredit Mobil di PT "X" Dengan Menggunakan Teknik Data Mining Decision Tree. *Jurusan Teknik Industri, Universitas Katolik Parahyangan*.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. *Journal of artificial intelligence research*, *4*, 77-90.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, *28*(1), 71-72.

Raviya, K. H., & Gajjar, B. (2013). Performance Evaluation of different data mining classification algorithm using WEKA. *Indian Journal of Research*, *2*(1), 19-21.

Syamsuddin, L. (2002). Manajemen Keuangan Perusahaan, edisi baru, cetakan ketujuh. *Penerbit: Rajawali Pers, Jakarta*.

Tan, A. C., Naiman, D. Q., Xu, L., Winslow, R. L., & Geman, D. (2005). Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics*, *21*(20), 3896-3904.

Witten, I. H., Frank, E., & Hall, M. A. (2011). Introduction to weka. *Data mining: practical machine learning tools and techniques*, *2*, 365-368.