**ANALISIS PENERAPAN NAÏVE BAYES UNTUK MEMPREDIKSI RESIKO KREDIT ANGGOTA KOPERASI KELUARGA GURU**

**Weko Susanto1, Luthfi Indriyani2**

1Universitas Bina Sarana Informatika

Wko.76ers@yahoo.com

2Universitas Bina Sarana Informatika

Luthfi.lfy@bsi.ac.id

***Abstrak***

*Pada laporan pertanggungjawaban terdapat penurunan pemberian (piutang) kepada anggota koperasi sebesar 17.319.802.163, pada tahun 2015 pemberian piutang 127.866.969.180 menjadi 110.547.167.017 pada tahun 2016 di seluruh wilayah dan salah satu faktor penurunan kinerja disebabkan kredit bermasalah atau tidak tertagih sehingga modal utama mengendap pada anggota yang tidak tertagih di Koperasi Keluarga Guru Jakarta pada piutang 2015 dan 2016. Kredit adalah peminjaman uang dengan pembayaran secara tidak langsung atau membayar dengan secara berkala, dengan batas jumlah pinjaman tertentu yang diizinkan oleh pihak Koperasi. Kredit merupakan sumber utama penghasilan dalam sebuah koperasi simpan pinjam. Karena dalam proses pemberian kredit tidak jarang terjadi kredit macet, di mana akan terjadi tunggakan kredit dalam masa angsuran. Data mining merupakan teknik mengolalah data dengan jumlah yang besar untuk mendapatkan informasi yang berharga untuk mengambil keputusan yang penting. Dalam penelitian ini studi kasus yang dilakukan pada data anggota Koperasi Keluarga Guru Jakarta dengan menggunakan model Naive Bayes (NBC). Dari hasil penelitian ini untuk menentukan prediksi kelayakan kredit lancar atau kredit macet, dan evaluasi performance naive bayes. Untuk hasil penelitian algoritma naive bayes ini, dapat di accuracy yaitu 84,00% masuk klasifikasi baik, algoritma naive bayes mampu menangani data range jauh berbeda.*

**Kata kunci:** *data mining, klasifikasi, algoritma naive bayes.*

***Abstract***

*In accountability report impairment provision (receivable) to members of the cooperative amounted to 17,319,802,163 from giving receivables 127 866 969 180 in 2015 to 110 547 167 017 in 2016 to the whole region and one of the factors due to non-performing loans decreased performance or uncollected so major capital settles on members who are not collectible in Family Cooperative Teacher Jakarta in 2015 and 2016. loans receivable are borrowing money with indirect payment or pay a regular basis, with a limit of a certain loan amount allowed by the Cooperative. Credit is the main source of income in a savings and loan cooperative. Because in the process of granting credit is not uncommon credit crunch, where there will arrear credit in the installment period. Data mining is a technique of managing large amounts of data to get valuable information to make important decisions. In this research case study carried out on the data members of the Cooperative Family Teachers Jakarta using Naive Bayes models (NBC). From the results of this study to determine the prediction of credit worthiness or bad credit, and evaluation of naive bayes performance. To research Naive Bayes algorithm, can be in accuracy is 84,00%, good clasificasion, even naive Bayes algorithm is able to handle Missing or incomplete data.*

*Keywords: data mining, debtors, naive bayes algorithm, classification.*

**1. Pendahuluan**

Dalam kelesuan ekonomi global dan regional banyak menyebabkan daya beli masyarakat melemah yang berdampak buruk bagi roda usaha ekonomi disegala bidang termasuk dalam kegiatan simpan pinjam di Koperasi Keluarga Guru Jakarta (KKGJ). Banyak peminjam di KKGJ juga mengalami kesulitan dalam pengembalian simpan pinjam sehingga mengalami dampak buruk banyak tagihan yang tidak tertagih. Keuntungan menurun sehingga sisa hasil usaha menjadi berkurang.

Pada laporan pertanggungjawaban tahun buku 2017 terdapat penurunan pemberian (piutang) kepada anggota koperasi sebesar 17.319.802.163 dari pemberian piutang 127.866.969.180 pada tahun 2015 menjadi 110.547.167.017 pada tahun 2016 pada keseluruhan wilayah dan salah satu faktor penurunan kinerja disebabkan kredit bermasalah atau tidak tertagih sehingga modal utama mengendap pada anggota yang tidak tertagih.

Pemberian piutang didasarkan pada perhitungan manual pemotongan penghasilan sebesar 30-50% sehingga penghasilan yang dapat dibawa pulang (take home pay) sebesar 50-70%, dengan besaran yang dibawa pulang kemungkinan terjadi kredit macet sangat kecil namun kenyataan ada saja anggota yang tidak lancar saat pembayaran angsuran dan menjadi problem di KKGJ.

Dari fenomena penurunan pendapatan dari kredit macet maka harus segera dicari solusi dalam memecahakan masalah, dengan mengumpulkan data yang ada untuk dijadikan informasi sehingga menghasilkan pengetahuan dan mampu mengklasifikasi dalam menentukan faktor utama terjadinya kredit bermasalah, sehingga akan mendapat pengetahuan baru dari data yang ditambang untuk mengambil keputusan dalam menemukan jalan keluar dari masalah yang ada.

Untuk menambang data, permasalahannya ada pada kelengkapan informasi data historis dan terlalu banyak atau besar data sehingga menyebabkan analisa data secara manual menjadi tidak memungkinkan. Solusi yang dapat dilakukan adalah dengan menggunakan teknik data mining sebagai serangkaian proses untuk

menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar dan tidak lengkap. Data mining berkaitan dengan ilmu-ilmu lain seperti Database System, Data Warehousing, Statistic, Machine Learning, Information Retrievel, dan Komputasi tingkat tinggi.

Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya memerlukan standart yang dapat dipertangungjawabkan hasilnya yang sesuai dengan standar pengolahan data mnjadi informasi dan informasi menjadi pola atau pengetahuan baru salah satu prosedur adalah CRISP-DM (Cross- Industry Standard Process for Data Mining).

Bentuk klasifikasi yang dapat digunakan dalam menentukan kelas resiko kredit sebagai bahan pertimbangan untuk memutuskan anggota sebagai calon debitur bermaslah atau tidak adalah dengan mengunakan metode Naive Bayes atau yang sering disebut sebagai Naive Bayes Classifier (NB C). Naïve Bayes adalah suatu pendekatan probabilitas untuk menghasilkan pengklasifikasian dimana digabungkannya probabilitas kata dengan probabilitas kategori untuk menentukan kemungkinan kategori bagi suatu obyek (Jiawei Han dan Micheline Kamber, 2001). Algoritma NBC bertujuan untuk melakukan klasifikasi data pada kelas tertentu NBC merupakan algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat) (Zhang, 2004). Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Eko Prasetyo, 2012) Dalam penelitian ini, data yang digunakan adalah data peminjam pada Koperasi Keluarga Guru Jakarta Data tersebut berisi data debitur beserta kriteria yang digunakan dalam menentukan kelayakan debitur.

Data mining disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Bustami, 2014).

Dikutip dari buku Introduction to Data Mining, Naïve Bayes mengasumsikan bahwa nilai dari sebuah input atribut pada kelas yang diberikan tidak tergantung dengan nilai dari atribut yang lain. Naive Bayes mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberi label kelas y (Pang-Ning Tan, et al, 2006).

Menurut Gorunescu proses klasifikasi didasarkan pada empat komponen mendasar, yaitu:

1. Kelas (Class)

Variabel dependen dari model, merupakan variabel kategorikal yang merepresentasikan “label” pada objek setelah klasifikasinya. Contoh kelas semacam ini adalah: adanya kelas penyakit jantung, loyalitas pelanggan, kelas bintang (galaksi), kelas gempa bumi (badai), dll.

1. Prediktor (Predictor)

Variabel independen dari model, direpresentasikan oleh karakteristik (atribut) dari data yang akan diklasifikasikan dan berdasarkan klasifikasi yang telah dibuat. Contoh prediktor tersebut adalah : merokok, konsumsi alkohol, tekanan darah, frekuensi pembelian, status perkawinan, karakteristik (satelit) gambar, catatan geologi yang spesifik, kecepatan dan arah angin, musim, lokasi terjadinya fenomena, dll.

1. Pelatihan dataset (Training dataset)

Kumpulan data yang berisi nilai-nilai dari kedua komponen sebelumnya dan digunakan untuk melatih model dalam mengenali kelas yang cocok/sesuai, berdasarkan prediktor yang tersedia. Contoh set tersebut adalah: kelompok pasien yang diuji pada serangan jantung, kelompok pelanggan supermarket (diselidiki oleh intern dengan jajak pendapat), database yang berisi gambar untuk monitoring teleskopik dan pelacakan objek astronomi, database badai, database penelitian gempa.

1. Dataset Pengujian (Testing Dataset)

Berisi data baru yang akan diklasifikasikan oleh (classifier) model yang telah dibangun di atas sehingga akurasi klasifikasi (model performance) dapat dievaluasi.

Menurut Claudia Clarentia Ciptohartono, 2013. Dari hasil penelitian membuktikan bahwa algoritma Naïve Bayes dapat diterapkan untuk menilai kelayakan kredit pada BCA Finance Jakarta. Dan pengolahan data awal merupakan tahapan yang sangat mempengaruhi hasil akurasi yang baik sehingga akurasi akhir yang dihasilkan termasuk kategori Excellent. Penilaian kelayakan kredit pada BCA Finance Jakarta pada data awal dengan pre-processing menghasilkan akurasi sebesar 85,57%.

**2. Metode Penelitian**

Adapun tahapan yang dilakukan pada rancang bangun sistem pendukung penentuan kelas resiko kredit adalah sebagai berikut:

1. Pengumpulan data dan informasi

Data yang digunakan dalam menunjang penelitian ini adalah diperoleh dari database koperasi untuk wilayah Jakarta Pusat dan Utara yang meminjam kredit. Data tersebut berjumlah 450 record yang terdiri dari 9 atribut yaitu: A1: Jenis Kelamin, A2: Kecamatan, A3: Usia, A4: Gaji, A5: Jumlah Permohonan, A6: Lama, A7: Pokok Hutang, A8: Bunga, A9 Angsuran.

 Sedangkan untuk Obyek penelitian yang dituju adalah calon debitur yang akan diklasifikasikan dalam kelas resiko kredit lancar atau macet, dengan melakukan proses klasifikasi data peminjam yang berisiko kredit lancar dan peminjam yang berisiko kredit macet dengan mengenali pola data yang telah ada sebelumnya.

1. Pengolahan data dan informasi

Pada penelitian ini akan digunakan sebanyak 450 data record histories peminjam, semua data tersebut digunakan untuk training sebanyak 450 data dan 10 testing data.

Data yang digunakan dalam penelitian ini mempunyai 9 attribut yang semua attribut kategori, serta 1 attribut sebagai variabel output atau kelas resiko kredit. Ke-9 attribut yang akan digunakan dalam membangun sistem klasifikasi dalam menentukan kelas resiko kredit

1. Tahapan Analisis dan pengolahan data dengan metode Naïve Bayes

Gambar 2.1 Tahapan Analisis dan pengolahan data dengan metode Naïve Bayes

**3. Hasil dan Pembahasan**

Data yang digunakan dalam menunjang penelitian ini adalah diperoleh dari database petugas Jakarta Pusat dan Utara yang menangani perkreditan kepada anggota. Data tersebut berjumlah 450 record yang terdiri dari 9 atribut yaitu: A1: Jenis Kelamin, A2: Kecamatan, A3: Usia, A4: Gaji, A5: Jumlah Permohonan, A6: Lama, A7: Pokok Hutang, A8: Bunga, A9: Angsuran. Dengan Obyek penelitian yang dituju adalah calon debitur yang akan diklasifikasikan dalam kelas resiko kredit lancar atau macet.

**3.1. Pengolahan Data dan Informasi**

Pada penelitian ini digunakan sebanyak 450 data record histories debitur, data tersebut digunakan untuk training sebanyak 450 data dan 10 testing sebanyak data. Data training adalah data yang digunakan untuk membuat model klasifikasi. Sedangkan data testing, adalah data yang nantinya untuk mengukur tingkat keakuratan sistem yang telah dibuat.

Data yang digunakan dalam penelitian ini mempunyai 9 atribut, serta 1 attribut sebagai variabel output atau kelas resiko kredit.

**3.2. Analisis Mengunakan Metode Naïve Bayes**

**3.2.1. Langkah Penghitungan Naïve Bayes**

1. Pendefinisian kelas P(Ci|X)

Ci merupakan kelas klasifikasi dimana terdapat C1 yang menunjukkan kelas resiko kredit lancar dan C2 yang menunjukkan kelas resiko kredit macet. X merupakan data testing yang terdiri dari vektor atas semua atribut penentu kreditur terbaik yaitu XJeniskelamin, XKecamatan, XGaji, XPinjaman, XLama, XPokok, XBunga, XAnggs.

.

1. Menghitung Probabilitas Prior P(Ci)

Dari 450 data training yang digunakan, diketahui kelas C1 sebanyak 322 data dan kelas C2 sebanyak 128 data. Penghitungan Probabilitas prior untuk kemungkinan kelas kreditur lancar P(C1) berdasarkan persamaan (2.2) adalah sebagai berikut:

$$P\left(C\_{1}\right)= \frac{322}{450}=0.716$$

Sedangkan penghitungan probabilitas prior untuk kemungkinan kelas kreditur macet P(C2) berdasarkan persamaan (2.2) adalah sebagai berikut:

$$P\left(C\_{2}\right)= \frac{128}{450}=0.284$$

1. Penghitungan Probabilitas linkehood X bersyarat C (P(X|Ci).

Penghitungan probabilitas linkehood dilakukan pada data training sejumlah 450 data dengan menggunakan X yang merupakan vektor untuk atribut penentuan resiko kreditur yaitu: XJeniskelamin, XKecamatan, XGaji, XPinjaman, XLama, XPokok, XBunga, XAnggs, sehingga (X|Ci) yang dijabarkan menjadi P(XJeniskelamin, XKecamatan, XGaji, XPinjaman, XLama, XPokok, XBunga, XAnggs | Ci) dan untuk setiap X dihitung kemungkinannya terhadap Ci berdasarkan persamaan 2.3.

Table 3.1 Data Jenis Kelamin

|  |  |  |  |
| --- | --- | --- | --- |
| Jenis Kelamin | Lancar | Macet | Jumlah |
| Perempuan | 209 | 87 | 296 |
| Lelaki | 113 | 41 | 154 |
| Jumlah | 322 | 128 | 450 |

$$P\left(C\_{lancar}\right)=\frac{209}{322}=0.649$$

$$P\left(C\_{Macet}\right)=\frac{87}{128}=0.680 $$

1. Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel.

Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan menggunakan aturan perkalian dengan mengacu pada tabel 4.1 sebagai berikut:

$$P\left(X\_{Perempuan}\right)=\frac{450}{296}=0.658$$

$$P\left(X\_{Lelaki}\right)=\frac{450}{154}=0.342$$

Untuk detail hasil penghitungan setiap atribut dapat dilihat pada tabel 4.2 berikut:

Tabel .3.2. Hasil Penghitungan probabilitas posterior X bersyarat C (atribut kategorial).

| Atribut | Kategori Atribut | P(X|Ci) *Prior* | P(X) *Evidence* | P(Ci)*Likehood*  |
| --- | --- | --- | --- | --- |
| Lancar | Macet | Lancar | Macet | Lancar | Macet |
| Jenis Kelamin | p | 0,649 | 0,680 | 0,658 | 0,658 | 0,716 | 0,284 |
| L | 0,351 | 0,320 | 0,342 | 0,342 | 0,716 | 0,284 |
| Kecamatan | Cempaka putih | 0,090 | 0,125 | 0,100 | 0,100 | 0,716 | 0,284 |
| Cilincing | 0,006 | 0,008 | 0,007 | 0,007 | 0,716 | 0,284 |
| Dikdas DKI Jakarta | 0,009 | 0,031 | 0,016 | 0,016 | 0,716 | 0,284 |
| Dikdas Jakarta Pusat | 0,012 | 0,039 | 0,020 | 0,020 | 0,716 | 0,284 |
| Gambir | 0,081 | 0,055 | 0,073 | 0,073 | 0,716 | 0,284 |
| Johar Baru | 0,112 | 0,078 | 0,102 | 0,102 | 0,716 | 0,284 |
| Kelapa Gading | 0,012 | 0,008 | 0,011 | 0,011 | 0,716 | 0,284 |
| Kemayoran | 0,199 | 0,219 | 0,204 | 0,204 | 0,716 | 0,284 |
| Kep. Seribu Utara | 0,003 | 0,000 | 0,002 | 0,002 | 0,716 | 0,284 |
| Koja | 0,003 | 0,016 | 0,007 | 0,007 | 0,716 | 0,284 |
| Menteng | 0,081 | 0,039 | 0,069 | 0,069 | 0,716 | 0,284 |
| Pademangan | 0,012 | 0,000 | 0,009 | 0,009 | 0,716 | 0,284 |
| Penjaringan | 0,012 | 0,000 | 0,009 | 0,009 | 0,716 | 0,284 |
| Sawah besar | 0,028 | 0,070 | 0,040 | 0,040 | 0,716 | 0,284 |
| Senen | 0,168 | 0,102 | 0,149 | 0,149 | 0,716 | 0,284 |
| Tanah Abang | 0,155 | 0,195 | 0,167 | 0,167 | 0,716 | 0,284 |
| Tanjung Priok | 0,012 | 0,000 | 0,009 | 0,009 | 0,716 | 0,284 |
| Tambora | 0,000 | 0,008 | 0,002 | 0,002 | 0,716 | 0,284 |
| Kep. Seribu Selatan | 0,003 | 0,008 | 0,004 | 0,004 | 0,716 | 0,284 |
| Gaji | Gaji1 | 0,031 | 0,094 | 0,049 | 0,049 | 0,716 | 0,284 |
| Gaji2 | 0,158 | 0,141 | 0,153 | 0,153 | 0,716 | 0,284 |
| Gaji3 | 0,180 | 0,344 | 0,227 | 0,227 | 0,716 | 0,284 |
| Gaji4 | 0,599 | 0,398 | 0,542 | 0,542 | 0,716 | 0,284 |
| Gaji5 | 0,028 | 0,023 | 0,027 | 0,027 | 0,716 | 0,284 |
| Gaji6 | 0,003 | 0,000 | 0,002 | 0,002 | 0,716 | 0,284 |
| Pinjaman | Pinjaman1 | 0,177 | 0,008 | 0,129 | 0,129 | 0,716 | 0,284 |
| Pinjaman2 | 0,280 | 0,141 | 0,240 | 0,240 | 0,716 | 0,284 |
| Pinjaman3 | 0,149 | 0,125 | 0,142 | 0,142 | 0,716 | 0,284 |
| Pinjaman4 | 0,134 | 0,109 | 0,127 | 0,127 | 0,716 | 0,284 |
| Pinjaman5 | 0,127 | 0,094 | 0,118 | 0,118 | 0,716 | 0,284 |
| Pinjaman6 | 0,099 | 0,180 | 0,122 | 0,122 | 0,716 | 0,284 |
| Pinjaman7 | 0,025 | 0,125 | 0,053 | 0,053 | 0,716 | 0,284 |
| Pinjaman8 | 0,009 | 0,102 | 0,036 | 0,036 | 0,716 | 0,284 |
| Pinjaman9 | 0,000 | 0,016 | 0,004 | 0,004 | 0,716 | 0,284 |
| Pinjaman10 | 0,000 | 0,102 | 0,029 | 0,029 | 0,716 | 0,284 |
| Lama | Lama1 | 0,152 | 0,180 | 0,160 | 0,160 | 0,716 | 0,284 |
| Lama2 | 0,323 | 0,211 | 0,291 | 0,291 | 0,716 | 0,284 |
| Lama3 | 0,196 | 0,211 | 0,200 | 0,200 | 0,716 | 0,284 |
| Lama4 | 0,102 | 0,117 | 0,107 | 0,107 | 0,716 | 0,284 |
| Lama5 | 0,211 | 0,266 | 0,227 | 0,227 | 0,716 | 0,284 |
| Lama6 | 0,016 | 0,016 | 0,016 | 0,016 | 0,716 | 0,284 |
| Pokok | Pokok1 | 0,689 | 0,156 | 0,538 | 0,538 | 0,716 | 0,284 |
| Pokok2 | 0,292 | 0,570 | 0,371 | 0,371 | 0,716 | 0,284 |
| Pokok3 | 0,019 | 0,195 | 0,069 | 0,069 | 0,716 | 0,284 |
| Pokok4 | 0,000 | 0,055 | 0,016 | 0,016 | 0,716 | 0,284 |
| Pokok5 | 0,000 | 0,016 | 0,004 | 0,004 | 0,716 | 0,284 |
| Pokok6 | 0,000 | 0,008 | 0,002 | 0,002 | 0,716 | 0,284 |
| Bunga | Bunga1 | 0,457 | 0,148 | 0,369 | 0,369 | 0,716 | 0,284 |
| Bunga2 | 0,283 | 0,234 | 0,269 | 0,269 | 0,716 | 0,284 |
| Bunga3 | 0,211 | 0,250 | 0,222 | 0,222 | 0,716 | 0,284 |
| Bunga4 | 0,050 | 0,234 | 0,102 | 0,102 | 0,716 | 0,284 |
| Bunga5 | 0,000 | 0,125 | 0,036 | 0,036 | 0,716 | 0,284 |
| Bunga6 | 0,000 | 0,008 | 0,002 | 0,002 | 0,716 | 0,284 |
| Usia | Usia1 | 0,006 | 0,016 | 0,009 | 0,009 | 0,716 | 0,284 |
| Usia2 | 0,093 | 0,109 | 0,098 | 0,098 | 0,716 | 0,284 |
| Usia3 | 0,236 | 0,242 | 0,238 | 0,238 | 0,716 | 0,284 |
| Usia4 | 0,665 | 0,633 | 0,656 | 0,656 | 0,716 | 0,284 |
| Anggs | Anggs1 | 0,284 | 0,484 | 0,362 | 0,362 | 0,716 | 0,284 |
| Anggs2 | 0,284 | 0,488 | 0,469 | 0,469 | 0,716 | 0,284 |
| Anggs3 | 0,284 | 0,028 | 0,138 | 0,138 | 0,716 | 0,284 |
| Anggs4 | 0,284 | 0,000 | 0,027 | 0,027 | 0,716 | 0,284 |
| Anggs5 | 0,284 | 0,000 | 0,002 | 0,002 | 0,716 | 0,284 |
| Anggs6 | 0,000 | 0,008 | 0,002 | 0,002 | 0,716 | 0,284 |

**3.2.2. Menghitung Probabilitas Posterior**

Tahapan selanjutnya adalah menggunakan Probabilitas Prior untuk menentukan class terhadap temuan kasus baru, dengan cara terlebih dahulu menghitung Probabilitas Posteriornya, hal tersebut dilakukan apabila ditemukan kasus baru dalam pengolahan data. Berikut tabel probabilitas posterior untuk menghitung kasus baru yang ditemukan:

Tabel 3.3. Penghitungan Probabilitas Posterior

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Atribut | Data test | Kelas | P(X|Ci) | P(X) |
|  |  |  | Lancar | Macet | Lancar | Macet |
| Total | 450 | 0,716 | 0,284 |  |  |
| Je-kel | P | P | 0,649 | 0,680 | 0,658 | 0,658 |
| Kec. | Senen | Senen | 0,168 | 0,102 | 0,149 | 0,149 |
| Usia | 48 | Usia3 | 0,236 | 0,242 | 0,238 | 0,238 |
| Gaji  | 3.500.000  | Gaji4 | 0,599 | 0,398 | 0,542 | 0,542 |
| Pinj. | 10.000.000  | Pinjm.1 | 0,177 | 0,008 | 0,129 | 0,129 |
| lama | 24  | Lama2 | 0,323 | 0,211 | 0,291 | 0,291 |
| Pokok | 416.667  | Pokok2 | 0,292 | 0,570 | 0,371 | 0,371 |
| Bunga | 150.000  | Bunga2 | 0,283 | 0,234 | 0,269 | 0,269 |
| Anggs | 566.667  | Anggs1 | 0,484 | 0,055 | 0,362 | 0,362 |
| P(X|Ci) | 0,0001491 | 0,0000003 |  |  |
| P(X) |  |  | 0,00007202 | 0,00007202 |

**3.2.3. Tingkat Akurasi dengan 9 Atribut**

Pada evaluasi ini dilakukan proses komparasi antara hasil dari sistem dengan data asli. Perbandingan dari kedua hasil tersebut akan dijadikan acuan dalam menghitung tingkat akurasi sistem. Detail analisa hasil akurasi pada evaluasi ini dengan menggunakan 9 atribut dapat dilihat pada Tabel 3.4 berikut:

Tabel 3.4. Perbandingan Hasil Implementasi Sistem dengan Data Asli dengan perhitungan manual excel dan rapid miner

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Data debitur | Data Fakta | Naive bayes | Keterangan |
| 1 | Kreditur - 1 | Lancar | Macet | Tidak Sesuai |
| 2 | Kreditur - 2 | Macet | Macet | Sesuai |
| 3 | Kreditur - 3 | Macet | Macet | Sesuai |
| 4 | Kreditur - 4 | Lancar | Lancar | Sesuai |
| 5 | Kreditur - 5 | Lancar | Lancar | Sesuai |
| 6 | Kreditur - 6 | Lancar | Macet | Tidak Sesuai |
| 7 | Kreditur - 7 | Lancar | Lancar | Sesuai |
| 8 | Kreditur - 8 | Macet | Macet | Tidak Sesuai |
| 9 | Kreditur - 9 | Lancar | Lancar | Sesuai |
| 10 | Kreditur - 10 | Lancar | Lancar | Sesuai |
| 11 | Kreditur - 11 | Macet | Lancar | Tidak Sesuai |
| 12 | Kreditur - 12 | Lancar | Macet | Tidak Sesuai |
| 13 | Kreditur - 13 | Lancar | Lancar | Sesuai |
| 14 | Kreditur - 14 | Macet | Macet | Sesuai |
| 15 | Kreditur - 15 | Lancar | Lancar | Sesuai |
| 16 | Kreditur - 16 | Lancar | Lancar | Sesuai |
| 17 | Kreditur - 17 | Lancar | Lancar | Sesuai |
| 18 | Kreditur - 18 | Macet | Macet | Sesuai |
| 19 | Kreditur - 19 | Lancar | Lancar | Sesuai |
| 20 | Kreditur - 20 | Macet | Macet | Sesuai |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| 440 | Kreditur - 440 | Lancar | Lancar | Sesuai |
| 441 | Kreditur - 441 | Lancar | Lancar | Sesuai |
| 442 | Kreditur - 442 | Lancar | Lancar | Sesuai |
| 443 | Kreditur - 443 | Lancar | Lancar | Sesuai |
| 444 | Kreditur - 444 | Lancar | Macet | Tidak Sesuai |
| 445 | Kreditur - 445 | Macet | Macet | Sesuai |
| 446 | Kreditur - 446 | Macet | Macet | Sesuai |
| 447 | Kreditur - 447 | Lancar | Lancar | Sesuai |
| 448 | Kreditur - 448 | Macet | Lancar | Tidak Sesuai |
| 449 | Kreditur - 449 | Macet | Macet | Sesuai |
| 450 | Kreditur - 450 | Lancar | Lancar | Sesuai |

Table 3.5 Hasil mengunakan Algoritma Naive Bayes

| Uraian | Ket |
| --- | --- |
| Agsran 1 |  |  |  |  |  | Lancar |
| Agsran 2 | Gaji2 | Menteng |  |  |  | Lancar |
| Kep. Seribu |  |  |  | Lancar |
| Senen | Pjman4 |  |  | Lancar |
| Gaji3 | Pokok1 | Gambir |  |  | Lancar |
|  | Menteng  |  |  | Lancar |
|  | Senen |  |  | Lancar |
|  | Tnh Abang |  |  | Lancar |
|  | Tj. Priok |  |  | Lancar |
|  | Kmyoran | Jk=L |  | Lancar |
| Pokok2 | Pjman1 |  |  | Lancar |
|  | Pjman3 | Lama3 | Dikdas DKI | Lancar |
|  | Pjman4 |  | Senen | Lancar |
|  | Pjman5 |  |  | Lancar |
| Gaji4 |  |  |  |  | Lancar |
| Gaji5 |  |  |  |  | Lancar |
| Gaji6 |  |  |  |  | Lancar |
| Agsran3 | Gaji5 |  |  |  |  | Lancar |

Perbandingan hasil klasifikasi data testing mengunakan algoritma naive bayes dengan data asli menggunakan 9 atribut (A1, A2 A3,..., A9), didapatkan data yang diklasifikasikan prediksi lancar sebanyak 322 data yang diklasifikasikan pengahasilan yang dibawah pulang sebesar 50 – 70% dari gaji dengan data asli. Ternyata 289 kreditur sesuai lancar dan 39 diprediksi lancar ternyata macet. Begitu juga dengan data macet sebanyak 128 data di perediksi macet ternyata lancar dan 89 sesuai dengan data asli dengan demikian dapat dihitung tingkat keakurasianya.

**3.2.4. Confussion Matrix**

Pada data asli tahap selanjutnya adalah pengukuran kinerja klasifikasi dengan menggunakan Confusion Matrix berdasarkan nilai akurasi, sensitivitas, spesifisitas dan AUC yang ditunjukkan oleh Tabel 3.6. hasil penghitungan dalam laporan tabel sebagai berikut:

Tabel 3.6. Model Confussion Matrix algoritma Naïve Bayes Classifier

|  |  |  |  |
| --- | --- | --- | --- |
| Accuracy: | 84,00% |  |  |
|  | True Lancar | True Macet | Class Presicion |
| Pred. Lancar | 289 | 39 | 88,11% |
| Pred. Macet | 33 | 89 | 72,95% |
| Class Recall | 89,75% | 69,53% |  |

Jumlah True Positive (TP) adalah 289 record diklasifikasikan sebagai kredit LANCAR dan False Positive (FN) sebanyak 39 record diklasifikasikan sebagai kredit LANCAR tetapi masuk klasifikasi kredit MACET. Berikutnya 33 record untuk False Negative (FN) diklasifikasikan sebagai kredit MACET, dan 89 record True Negative (TN) diklasifikasikan sebagai kredit MACET tetapi masuk klasifikasi kredit LANCAR. Berdasarkan tabel 4.4 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algortima Naïve Bayes Classifier adalah sebesar 84,00% dan dapat dihitung untuk mencari nilai accuracy, sensitivity, specificity, PPV, dan NPV pada persamaan dibawah ini:

$$Accuracy =\frac{TP + TN}{TP+ TN+FP+FN}= \frac{\left(289+89\right)}{\left(289+89+33+39\right)}$$

$$=0,8400$$

$$Sensitivity= Recall=TP\_{rate}=\frac{TP}{TP+FN }= \frac{289}{\left(289+39\right)}=0,8811$$

$$Specificity=TN\_{rate}=\frac{TN}{TN+FP }= \frac{89}{(89+33)}=0,7295$$

$$PPV =\frac{TP}{TP+ FP}= \frac{289}{\left(289 +33\right)}=0,8975$$

$$NPV =\frac{TN}{TN+ FN}= \frac{39}{\left(39 +33\right)}=0,6953$$

$$AUC =\frac{TP\_{rate}+TN\_{rate}}{2}= \frac{0,8811+0,7295}{2}=0,8053$$

Hasil dari perhitungan persamaan diatas terlihat pada Tabel 3.7 dibawah ini:

Tabel 3.7 PPV, NPV

|  |  |
| --- | --- |
| Parameter | Nilai (%) |
| *Accuracy*  | 84,00%  |
| *PPV* | 89,75%  |
| *NPV* | 69,53% |
| *Recall/ TPrate* | 88,11% |
| Spesifitas/ *TNr*ate | 72,95% |
| A U C | 80,53% |

**3.2.5. Kurva ROC**

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC. Data peminjam KKGJ. menghasilkan Kurva ROC pada gambar 4 mengekspesikan confusion matrix dari tabel 4.4 Garis horizontal adalah false positives dan garis vertikal true positives. Menghasilkan nilai AUC (Area Under Curve) sebesar 0,898 dengan nilai akurasi klasifikasi Baik (good classification) pada dataset peminjam KKGJ.



Gambar 3.1. Kurva ROC dengan Metode Naive Bayes Data Peminjam KKGJ

**4. Kesimpulan**

Sistem klasifikasi data nasabah ini digunakan untuk menampilkan informasi klasifikasi lancar atau tidak lancarnya calon nasabah dalam membayar ciclan bulanan pada koperasi Koperasi Keluarga Guru Jakarta dengan menggunakan algoritma Naive Bayes.

Aplikasi data mining dengan algoritma naïve bayes ini dapat mempermudah pihak koperasi untuk melakukan evaluasi penentuan kelayakan pemohon dengan cepat.

Pengujian pada data rekapitulasi peminjam yang berjumlah 450 data dari Koperasi Keluarga Guru Jakarta dengan proses mining algoritma naïve bayes menghasilkan tingkat akurasi 84,00%, dimana dalam pengujian model data, keseluruhan data set digunakan sebagai data training. Penentuan data training dapat mempengaruhi hasil pengujian, karena pola dari data training tersebut akan dijadikan sebagai rule untuk menentukan kelas pada data testing. Sehingga besar atau kecil tingkat akurasi yang di dapat dipengaruhi oleh penentuan data training.

**Referensi**

Bustami, (2014). "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *TECHSI*, vol. III, pp. 11-14.

Ciptohartono, Claudia Clarentia .(2013). Algoritma Klasifikasi Naïve Bayes Untuk Menilai Kelayakan Kredit.

Han, Jia wei., dan Kamber, Micheline, (2001), Data Mining: Concepts and Techniques, Elsevier, San Diego.

H. Zhang and S. Sheng, (2004). "Learning Weighted Naive Bayes with Accurate Ranking," in *Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM’04)*, Canada,

Pang-ning Tan. 2006. Introduction To Dataminning. Pearson Education. Michigan State University

Prasetyo, E. (2012). Data Mining Konsep dan Aplikasi Menggunakan Matlab. Yogyakarta. Andi