

PENERAPAN DATA MINING DENGAN METODE NAÏVE BAYES UNTUK MEMPREDIKSI CALON TERJAMIN POTENSIAL (Studi Kasus PT XYZ)

Nurfitria Khoirunnisa¹, Imam Ma'ruf Nugroho², Yusuf Muhyidin

Sistem Informasi¹, Teknik Informatika^{2,3}

Manajemen Informatika¹, Teknik Informatika^{2,3}

Politeknik Negeri Subang¹, Sekolah Tinggi Teknologi Wastukencana^{2,3}

nurfitria@polsub.ac.id , imam.ma@stt-wastukencana.ac.id ,

yusufmuhyidin@stt-wastukencana.ac.id

Received: December 4, 2021. **Revised:** January 6, 2022. **Accepted:** January 10, 2022. **Issue Period:** Vol.6 No.1 (2022), Page 149-161

Abstrak: Sekarang ini bisnis penjaminan kredit di Indonesia semakin berkembang, hal tersebut dikarenakan tingginya minat masyarakat dalam melakukan kredit. PT XYZ adalah salah satu perusahaan BUMN yang bergerak di bidang Penjaminan Kredit di Indonesia. Kendala yang saat ini dihadapi oleh PT XYZ adalah belum adanya sistem yang dapat membantu memprediksi kriteria calon terjamin potensial. Dengan adanya prediksi tersebut, membantu perusahaan mengambil keputusan untuk memberikan jaminan kepada calon terjamin potensial. Data mining merupakan teknologi baru yang berguna untuk membantu perusahaan menemukan informasi yang sangat penting dari gudang data perusahaan yang sebelumnya dianggap tidak berguna. Data ini diolah dengan metode Klasifikasi yang selanjutnya dianalisis menggunakan teknik *Naive Bayes*. Analisis ini menggunakan UML (*Unified Modeling Language*) sebagai tools. Hasil yang diharapkan dari penelitian ini adalah analisa sistem usulan yang dapat membantu perusahaan untuk mengambil keputusan penjaminan.

Kata kunci: Data Mining, Klasifikasi, *Naive Bayes*, Potensial, Prediksi.

Abstract: Now this credit guarantee business in Indonesia is growing, it is because the high public interest in doing credit. PT XYZ is one of the state-owned company engaged in the Credit Guarantee in Indonesia. Constraints currently faced by the PT XYZ is the absence of a system that can help predict potential candidate criteria guaranteed. Given these predictions, helping the company took the decision to provide a guarantee to secure potential candidates. Data mining is a new and useful technologies to help companies find the most important information from enterprise data warehouses that were previously considered useless. This data is processed by the method of classification is further analyzed using *Naive Bayes*. This analysis using UML (*Unified Modeling Language*) as tools. The expected outcome of this research is the analysis of the proposed system that can help companies to make decisions underwriting.

Keywords: Classification, Data Mining, *Naive Bayes*, Potential, Prediction,

I. PENDAHULUAN

Sistem komputerisasi telah membantu perusahaan dalam mengatur segala urusan. Komputer merupakan salah satu alat yang diciptakan untuk menerobos keterbatasan dalam hal kecepatan, ketelitian, ingatan, tidak mengenal lelah dan mampu mengolah informasi dalam jumlah besar. Perkembangan teknologi informasi menyebabkan banyak orang dapat memperoleh data dengan sangat mudah. Data yang ada semakin lama akan terakumulasi semakin banyak. Namun sangat disayangkan akumulasi dari data tersebut tidak



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

dimanfaatkan secara optimal. Data mining adalah salah satu solusi untuk dapat memanfaatkan data secara optimal.

PT XYZ adalah perusahaan yang bergerak di bidang penjaminan kredit. Dimana saat ini kredit telah menjadi salah satu hal penunjang kehidupan masyarakat Indonesia. PT XYZ menjadi pihak ketiga atau penjamin atas pinjaman terjamin. Dengan adanya PT XYZ dapat membantu terjamin dan penerima jaminan agar sama-sama tidak menerima kerugian. Penjaminan yang dilakukan PT XYZ adalah mengcover sementara kredit terjamin yang macet, sehingga operasional bank tetap berjalan.

Dalam hal ini, prediksi minat penjaminan calon terjamin potensial sangat penting bagi PT XYZ, dimana dengan adanya prediksi calon terjamin potensial, PT XYZ dapat mengambil keputusan dalam menentukan strategi yang tepat. Teknologi data mining hadir sebagai solusi untuk mengatasi hal tersebut.

Data mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan^[1]. Aplikasi *data mining* ini memanfaatkan data masukan berupa data terjamin, data bank, pendapatan terjamin, dan besar kredit. Data tersebut akan diolah dengan metode klasifikasi dan dianalisis dengan menggunakan metode naïve bayes. Setelah melewati proses tersebut, maka akan didapatnya pola-pola untuk pengambilan keputusan. Sistem *data mining* dengan metode naïve bayes ini dapat digunakan untuk memprediksi calon terjamin potensial yang sesuai dengan kriteria.

II. METODE DAN MATERI

2.1. Penjaminan

Penjaminan merupakan salah satu kontra garansi atas kemungkinan terjadinya risiko yang harus ditanggung pihak bank.

Penjaminan Konvensional adalah kegiatan pemberian jaminan oleh Penjamin atas pemenuhan kewajiban finansial Terjamin kepada Penerima Jaminan.^[2]

Penjaminan Syariah adalah kegiatan pemberian jaminan oleh Penjamin atas pemenuhan kewajiban finansial Terjamin kepada Penerima Jaminan berdasarkan Prinsip Syariah. Dalam pandangan Islam jaminan merupakan ketentuan yang disyariatkan. Dalam ketentuan hukum Islam, sistem jaminan atau kafalah terdapat pada Fatwa Dewan Syariah Nasional Majelis Ulama Indonesia No. 11/DSN-MUI/IV/2000 tentang Kafalah.^[3]

2.2. Data Mining

Data mining merupakan suatu istilah yang biasa digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* merupakan proses yang menggunakan teknik *statistic*, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai *database* besar. *Data mining* merupakan kegiatan mengumpulkan dan memakai data histori untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining dapat dipakai untuk memperbaiki pengambilan keputusan dimasa depan..^[4]

Secara sederhana, data mining adalah kegiatan menemukan pengetahuan dari sejumlah data yang sangat besar. Data tersimpan dalam bentuk basis data, data warehouse, dan penyimpanan informasi lainnya.

Dengan memanfaatkan *data mining*, pengetahuan menarik atau informasi tingkat tinggi dapat diperoleh dari *database*. Pengetahuan ini dapat digunakan oleh pemakai untuk mendukung keputusan, proses kontrol, manajemen informasi dan pemrosesan *query*.^[4]

2.3. Klasifikasi

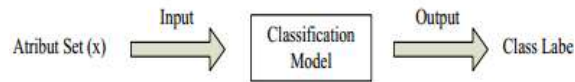


DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Klasifikasi adalah salah satu teknik menambang data yang bertujuan untuk membagi suatu benda, aktivitas, atau transaksi ke dalam bagiannya masing-masing. Klasifikasi merupakan pembagian data ke dalam wadahnya masing-masing dimana wadah ini telah kita tentukan sebelumnya.

Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis.^[5]



Gambar 2 Blok Diagram Model Klasifikasi^[5]

2.4. Metode Naïve Bayes

Metode Naïve Bayes ialah model dengan perhitungan peluang kemunculan suatu nilai pada tiap-tiap atribut. Metode *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris, *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema *Bayes* adalah :^[6]

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \dots\dots\dots (1)$$

Keterangan :
 X : Data dengan class yang belum diketahui
 H : Hipotesis data X merupakan suatu class spesifik
 $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
 $P(H)$: Probabilitas hipotesis H (prior probability)
 $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$: Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis. Karena itu, metode *naive bayes* di atas perlu disesuaikan sebagai berikut :

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad \dots\dots\dots (2)$$

Rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior \times likelihood}{evidence}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* nantinya akan dibandingkan dengan nilai – nilai *posterior* kelas lainnya untuk menentukan dimana kelas yang sesuai untuk suatu sampel yang nantinya akan diklasifikasikan. Lebih lanjut rumus *Naive Bayes* dilakukan dengan menjabarkan $P(F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C) P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2).P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad \dots\dots\dots (3)$$

Hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, dengan semakin banyaknya nilai yang ada, maka hampir mustahil untuk dilakukan analisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa setiap petunjuk



($F_1, F_2 \dots F_n$) saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka akan berlaku suatu persamaan berikut :

$$P(P_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \dots\dots\dots (4)$$

Dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi lebih sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran ($C|F_1, \dots, F_n$) dapat disederhanakan menjadi :

$$(C|F_1, \dots, F_n) = (C)P(F_1|C) P(F_2|C) P(F_3|C) \dots\dots\dots (5)$$

$$= P(C) \prod_{i=1}^n P(F_i|C)$$

Persamaan diatas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

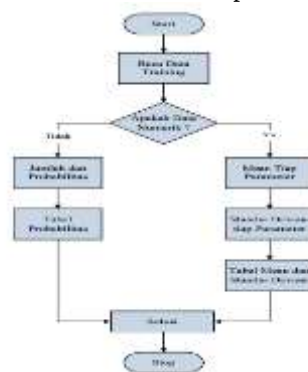
$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots (6)$$

Keterangan :

- P : Peluang
- X_i : Atribut ke i
- x_i : Nilai atribut ke i
- Y : Kelas yang dicari
- y_j : Sub kelas Y yang dicari
- μ : Mean, menyatakan rata – rata dari seluruh atribut
- σ : Deviasi standar, menyatakan varian dari seluruh atribut

Adapun alur dari metode *Naive Bayes* adalah sebagai berikut :

1. Baca data training.
2. Hitung Jumlah dan probabilitas, apabila data numerik maka:
 - a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas. ^[6]



Gambar 3 Skema *Naive Bayes*^[6]

2.6. Metode Pengumpulan Data

Metode pengumpulan data, merupakan kerangka yang digunakan untuk mengumpulkan data.

Tahap-tahap yang penulis lakukan dalam pengumpulan data dengan sebagai berikut :

1. *Research Sampling*
2. Identifikasi Variabel
3. Pengumpulan data



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- Merancang Kuisisioner
- Analisa Data

Rumus untuk menghitung validitas kuesioner menurut Arikunto^[7], adalah sebagai berikut:

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}}$$

Keterangan:

- r_{xy} = koefisien korelasi *item – total (bivariate person)*
 x = skor *item*
 y = skor total
 n = banyaknya subjek

III. PEMBAHASA DAN HASIL

3.1. Hasil Kuesioner

Dari kuesioner yang disebarkan, kemudian dikumpulkan kembali dan hasil dari kuesioner tersebut disajikan dalam bentuk tabel hasil.

Tabel 1 Hasil Kuesioner X₁

Responden	Pengolahan Data Terjamin					
	1	2	3	4	5	6
1	1	1	1	1	0	0
2	1	0	0	1	0	0
3	1	0	1	1	0	1
4	1	0	0	1	0	0
5	0	0	1	0	1	0
6	1	1	1	1	1	0
7	0	1	0	1	0	1
8	1	0	1	1	1	0
9	1	0	1	1	0	1
10	1	1	0	1	1	1
11	1	0	1	1	0	0
12	1	0	1	1	0	1
13	1	1	0	0	1	0
14	0	1	0	0	1	0
15	1	0	0	1	1	1
16	1	0	1	1	0	0
17	1	0	1	0	1	1
18	1	1	1	1	1	1
19	1	1	1	1	1	1
20	1	0	1	1	1	1
21	1	1	1	1	0	0
22	1	1	1	1	1	0
23	1	1	1	1	1	0
24	1	1	1	1	0	0
25	1	1	1	1	0	0
26	0	0	0	1	1	0
27	1	1	1	0	0	1
28	0	0	0	1	1	0
29	0	0	0	1	1	0
30	1	1	0	1	0	0
total skor(x)	24	15	19	25	16	11

Tabel 2 Hasil Kuesioner X₂



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Responden	Pengolahan Data Calon Terjamin						
	1	2	3	4	5	6	7
1	0	1	1	0	1	0	1
2	1	0	0	0	1	1	1
3	1	0	1	0	1	1	1
4	1	0	0	0	1	1	0
5	0	1	1	0	1	0	1
6	1	0	1	0	1	1	1
7	0	1	0	1	1	1	1
8	1	0	1	1	0	0	1
9	1	1	1	1	1	0	1
10	1	0	1	1	1	0	1
11	1	1	0	0	0	0	1
12	1	1	1	1	1	0	1
13	1	0	1	0	0	0	1
14	0	1	0	1	0	1	1
15	1	1	0	0	1	1	0
16	1	1	0	0	1	1	0
17	0	1	0	0	1	0	1
18	0	1	1	1	1	0	1
19	0	1	1	1	1	0	1
20	0	1	1	1	1	0	1
21	0	1	1	1	1	0	1
22	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1
24	0	1	1	0	1	1	1
25	0	1	1	1	1	0	1
26	0	1	0	1	1	1	1
27	0	1	0	0	0	1	1
28	0	1	1	1	0	0	1
29	0	1	1	1	0	0	1
30	0	1	1	0	0	0	1
total skor(x)	14	23	20	16	22	13	27

3.2. Perhitungan Hasil Kuesioner

Dari data hasil yang didapatkan dari kuesioner dan rumus yang digunakan maka ditemukan lah nilai berikut:

Tabel 3 Hasil Perhitungan Nilai untuk Data Terjamin (X_1)

Ukur	Item (X_1)					
	1	2	3	4	5	6
N	30	30	30	30	30	30
$\sum x$	22	13	18	21	13	10
$\sum y$	97	97	97	97	97	97
$\sum xy$	90	56	76	86	56	45
$\sum y^2$	409	409	409	409	409	409
$(\sum x)^2$	484	169	324	441	169	100
$(\sum y)^2$	9409	9409	9409	9409	9409	9409
Korelasi r_{xy}	0.629	0.447	0.599	0.250	0.187	0.503

Tabel 4 Hasil Perhitungan Nilai untuk Data Calon Terjamin (X_2)



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Ukur	Item (X ₂)						
	1	2	3	4	5	6	7
N	30	30	30	30	30	30	30
$\sum x$	14	23	20	16	22	13	27
$\sum y$	116	116	116	116	116	116	116
$\sum xy$	66	106	97	82	107	61	124
$\sum y^2$	568	568	568	568	568	568	568
$(\sum x)^2$	196	529	400	256	484	169	729
$(\sum y)^2$	13456	13456	13456	13456	13456	13456	13456
Korelasi r_{xy}	0.184	0.181	0.455	0.614	0.554	0.155	0.255

Dari perhitungan yang telah dilakukan maka R Hitung Dari masing – masing pertanyaan adalah sebagai berikut :

Tabel 5 r Hitung untuk kuesioner X1

No	Pernyataan	r Hitung	Hubungan r hitung dan r tabel = 0.361	Validasi Item
		(r_{xy})		
1	Apakah anda mengelola data terjamin secara manual ?	0.629	Lebih dari	Valid
2	Apakah anda kesulitan mengelola data terjamin?	0.447	Lebih dari	Valid
3	Apakah sering terjadi Human Error dalam pengelolaan data ?	0.599	Lebih dari	Valid
4	Apakah data terjamin yang anda gunakan valid ?	0.250	Lebih dari	Valid
5	Apakah ada sistem atau software yang digunakan untuk membantu mengelola data terjamin?	0.187	Kurang dari	Tidak Valid
6	Apakah sistem pengelolaan data terjamin saat ini sudah cukup efisien ?	0.503	Kurang dari	Tidak Valid

Tabel 6 r Hitung untuk kuesioner X1

No	Pernyataan	r Hitung	Hubungan r hitung dan r tabel = 0.361	Validasi Item
		(r_{xy})		
1	Apakah anda menggunakan form khusus untuk mengelola data calon terjamin ?	0.184	Kurang dari	Tidak
2	Apakah anda mengelola data calon terjamin secara manual ?	0.181	Kurang dari	Tidak
3	Apakah anda kesulitan mengelola data calon terjamin?	0.455	Lebih dari	Valid
4	Apakah sering terjadi Human Error dalam perediksi calon terjamin ?	0.614	Lebih dari	Valid
5	Apakah ada sistem atau software yang digunakan untuk membantu memprediksi calon terjamin?	0.554	Lebih dari	Valid
6	Apakah prosedur prediksi calon terjamin saat ini sudah cukup efisien?	0.155	Kurang dari	Tidak
7	Apakah dibutuhkan Sistem yang dapat membantu memprediksi calon terjamin?	0.475	Lebih dari	Valid

3.3. Kesimpulan Kuesioner

Dari hasil kuesioner X₁, dan X₂ yang telah di validasi, bahwa mengembangkan Sistem menggunakan *data mining* untuk Memprediksi Calon Terjamin Potensial ini merupakan langkah yang tepat dan sangat membantu. Dimana PT XYZ sudah memiliki form khusus untuk penilaian calon. Dengan bantuan data mining, diharapkan dapat membantu dalam proses memprediksi calon terjamin potensial.



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

3.4. Implementasi Data Mining Naïve Bayes

Pada metode ini, semua atribut akan ikut andil dalam pengambilan keputusan. Apabila diberikan k atribut yang saling bebas (*independence*), nilai probabilitas dapat diberikan sebagai berikut:

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times \dots \times P(x_k | C)$$

[6]

Tahapan awal proses perhitungan Naïve Bayes adalah dengan melakukan pengambilan data training dari data terjamin. Adapun variabel penentu yang digunakan dalam mengklasifikasikan data terjamin yaitu :

1. Tempat Tinggal
Merupakan variabel tempat tinggal terjamin yang dikelompokkan dalam dua kategori yaitu rumah dan kontrak.
2. Jumlah Keluarga
Merupakan variabel jumlah anggota keluarga terjamin yang dikelompokkan dalam 3 kategori yaitu sedikit, sedang, dan banyak. Sedikit apabila jumlah keluarganya < 3 orang, sedang apabila jumlah keluarganya 3-5 orang, dan banyak apabila jumlah keluarganya > 5 orang.
3. Usaha
Merupakan variabel jenis usaha terjamin yang dikelompokkan dalam 3 kategori yaitu mikro, kecil, dan menengah. Dapat dikategorikan mikro apabila perusahaan masih bersifat tradisional dengan kekayaan bersih maksimal 50jt dan penghasilan tahunan maksimal 300jt. Sedangkan kategori kecil apabila perusahaan memiliki kekayaan bersih diatas 50jt sampai 500jt dan memiliki penghasilan tahunan diatas 300jt sampai 2,5miliar. Sedangkan kategori menengah apabila perusahaan memiliki kekayaan bersih diatas 500jt sampai 10miliar dan memiliki penghasilan diatas 2,5miliar hingga 30miliar.
4. Penghasilan
Merupakan variabel penghasilan per bulan yang dikelompokkan dalam 3 kategori yaitu kecil, sedang, dan banyak. Kategori kecil apabila penghasilan perbulan < Rp 3.000.000,-. Sedangkan kategori sedang apabila penghasilan perbulan diantara Rp 3.000.000,- s.d 10.000.000,-. Dan dikatakan banyak apabila penghasilan mencapai > Rp 10.000.000,-.
5. Agunan
Merupakan variabel agunan yang dikelompokkan dalam 3 kategori yaitu kecil, sedang, dan banyak. Dapat dikategorikan kecil apabila kekayaan bersih < 50jt. Sedangkan kategori sedang apabila perusahaan memiliki kekayaan bersih diatas 50jt sampai 500jt. Sedangkan kategori banyak apabila perusahaan memiliki kekayaan bersih diatas 500jt sampai 10miliar.
6. Pinjaman
Merupakan variabel pinjaman yang dikelompokkan dalam 3 kategori yaitu kecil, sedang, dan banyak. Dapat dikategorikan kecil apabila pinjaman sebesar < 25jt. Sedangkan kategori sedang apabila perusahaan memiliki pinjaman diatas 25jt sampai 250jt. Sedangkan kategori banyak apabila perusahaan memiliki pinjaman diatas 250jt.

Langkah-langkah yang dilakukan untuk mengklasifikasi/ memprediksi data ialah :

1. Menghitung jumlah class/ label
 - a. $P(Y=\text{potensial}) = \frac{12}{20}$ “jumlah data potensial pada data terjamin dibagi dengan jumlah seluruh data”
 - b. $P(Y=\text{tidak}) = \frac{8}{20}$ “jumlah data tidak potensial pada data terjamin dibagi dengan jumlah seluruh data”
2. Menghitung jumlah kasus yang sama dengan class yang sama.



Tabel 7 Tabel Kasus

Tinggal	Keluarga	Usaha	Penghasilan	Agunan	Usia	Pinjaman	Penjaminan
kontrak	sedang	kecil	Sedang	kecil	muda	sedang	?
rumah	sedang	kecil	sedang	banyak	tua	sedang	?
kontrak	sedikit	mikro	banyak	sedang	tua	kecil	?
kontrak	banyak	kecil	sedang	sedang	muda	sedang	?
kontrak	sedikit	menengah	sedang	kecil	muda	sedang	?
rumah	sedang	menengah	sedang	sedang	muda	sedang	?
kontrak	sedikit	kecil	sedang	sedang	muda	kecil	?
rumah	banyak	kecil	sedang	sedang	tua	sedang	?
rumah	sedang	kecil	kecil	sedang	tua	kecil	?
kontrak	sedikit	mikro	sedang	kecil	muda	kecil	?

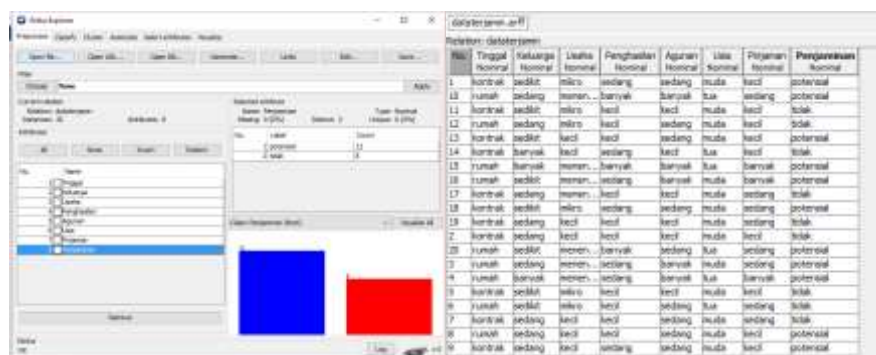
Tabel 8 Perbandingan Probabilitas

Kasus	Probabilitas Potensial	Probabilitas Tidak	Prediksi
Kasus 1	0	0.0049	Tidak
Kasus 2	0.0043	0	Potensial
Kasus 3	0.00055	0	Potensial
Kasus 4	0.0012	0.00047	Potensial
Kasus 5	0	0.00035	Tidak
Kasus 6	0.0067	0.00009	Potensial
Kasus 7	0.0029	0.00118	Potensial
Kasus 8	0.0012	0.00007	Potensial
Kasus 9	0.00077	0.0034	Tidak
Kasus 10	0	0.00081	Tidak

3.5. Implementasi Data Mining dengan Software WEKA



Gambar 5 Tools WEKA



No	Tinggal	Keluarga	Usaha	Penghasilan	Agunan	Usia	Pinjaman	Pengawasan
1	kontrak	sedikit	mikro	sedang	sedang	muda	kecil	potensial
2	rumah	sedang	kecil	banyak	banyak	tua	sedang	potensial
3	kontrak	sedikit	mikro	kecil	sedang	muda	kecil	tidak
4	rumah	sedang	mikro	kecil	sedang	muda	kecil	tidak
5	kontrak	sedikit	kecil	sedang	sedang	muda	kecil	potensial
6	kontrak	banyak	kecil	sedang	kecil	tua	kecil	tidak
7	rumah	sedikit	menengah	sedang	banyak	tua	banyak	potensial
8	rumah	sedikit	menengah	sedang	banyak	muda	banyak	potensial
9	kontrak	sedang	menengah	kecil	kecil	muda	sedang	tidak
10	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	potensial
11	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
12	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
13	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
14	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
15	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
16	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
17	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
18	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
19	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak
20	kontrak	sedikit	mikro	sedang	kecil	muda	sedang	tidak



Gambar 6 Pembuatan file data untuk dianalisis

Setelah dataterjamin.arff terbentuk, selanjutnya dilakukan analisa terhadap data tersebut. Pada bagian ini peranan Data Mining sebagai ilmu digunakan untuk membantu mencari informasi tersembunyi yang ada di dalam data. Teknik Klasifikasi digunakan untuk dapat menentukan apakah calon terjamin ini potensial atau tidak. Di sini dilakukan pengelompokkan data berdasarkan kriteria terakhir. Metode yang dipilih pada Teknik Klasifikasi ini adalah Naïve Bayes. Dengan kriteria terakhir bernama Penjaminan.

```
TIME TAKEN TO BUILD MODEL: 0 seconds

=== Evaluation on training set ===
=== Summary ===

Currently Classified Instances      17      85      4
Incorrectly Classified Instances    3      15      4
Kappa statistic                    0.7059
Mean absolute error                 0.2051
Root mean squared error             0.3405
Relative absolute error             41.4018 %
Root relative squared error         68.4477 %
Total Number of Instances          20

=== Detailed Accuracy By Class ===

      TP Rate    FP Rate    Precision    Recall    F-Measure    ROC Area    Class
-----
      1      0.727      0          1      0.727      0.842      0.948      potensial
      2      1      0.273      0.75      1      0.857      0.948      tidak
Weighted Avg.:    0.95      0.125      0.958      0.95      0.949      0.948

=== Confusion Matrix ===
a b   c-- classified as
0 0 : a = potensial
0 1 : b = tidak
```

Gambar 7 Hasil Klasifikasi data terjamin dengan metode Naïve Bayer pada software Weka bagian III

Dari hasil klasifikasi data terjamin diatas, dapat diperoleh hasil :

- Akurasi yang diperoleh adalah 85% dengan jumlah *Correctly Classified Instances* sebanyak 17.
- Jumlah *Incorrectly Classified Error* sebanyak 3 atau berjumlah 15%
- Kappa statistic 0.7059 menunjukkan nilai konsisten mendekati 1. Hal ini menandakan adanya korelasi antara penilaian data satu sama lainnya.
- Mean Absolute Error* atau nilai tengah kesalahan absolut adalah rata-rata mutlak dari kesalahan meramal, tanpa menghiraukan tanda positif maupun negatif. Data diatas menunjukkan mean absolute error sebesar 0.2051, hal ini menunjukkan rata-rata kesalahan prediksi kecil.
- Mean Squared Error* merupakan metode alternatif untuk mengevaluasi teknik peramalan masing-masing kesalahan (selisih data aktual terhadap data peramalan) dikuadratkan, kemudian dijumlahkan dan dibagi dengan jumlah data. Data diatas menunjukkan *Mean Squared Error* sebesar 0.3405, hal ini menunjukkan kesalahan prediksi kecil.
- Relative Absolute Error* menunjukkan presentase dari *Mean Absolute Error*. Data diatas menunjukkan nilai presentase sebesar 41,4018%
- Root Relative Squared Error* menunjukkan presentase dari *Root Squared Error*. Data diatas menunjukkan presentase sebesar 68,4477%

juara.arff

Relation: dataterjamin_predicted

No.	Tinggal Nominal	Keluarga Nominal	Usaha Nominal	Penghasilan Nominal	Agunan Nominal	Usia Nominal	Pinjaman Nominal	predictedPenjaminan Nominal	Penjaminan Nominal
1	kontrak	sedang	kecil	sedang	kecil	muda	sedang	tidak	
2	rumah	sedang	kecil	sedang	banyak	tua	sedang	potensial	
3	kontrak	sedikit	mikro	banyak	sedang	tua	kecil	potensial	
4	kontrak	banyak	kecil	sedang	sedang	muda	sedang	potensial	
5	kontrak	sedikit	menen...	sedang	kecil	muda	sedang	tidak	
6	rumah	sedang	menen...	sedang	sedang	muda	sedang	potensial	
7	kontrak	sedikit	kecil	sedang	sedang	muda	kecil	potensial	
8	rumah	banyak	kecil	sedang	sedang	tua	sedang	potensial	
9	rumah	sedang	kecil	kecil	sedang	tua	kecil	tidak	
10	kontrak	sedikit	mikro	sedang	kecil	muda	kecil	tidak	

Gambar 7 Hasil Prediksi



3.6. Membandingkan Hasil Prediksi

Tabel 9 Perbandingan Hasil Prediksi

Kasus	Prediksi Manual	Prediksi dengan Software
Kasus 1	Tidak	Tidak
Kasus 2	Potensial	Potensial
Kasus 3	Potensial	Potensial
Kasus 4	Potensial	Potensial
Kasus 5	Tidak	Tidak
Kasus 6	Potensial	Potensial
Kasus 7	Potensial	Potensial
Kasus 8	Potensial	Potensial
Kasus 9	Tidak	Tidak
Kasus 10	Tidak	Tidak

Dari perbandingan table diatas dapat dilihat bahwa perhitungan data mining dengan metode Naïve Bayes baik dilakukan manual maupun menggunakan sistem WEKA memberikan hasil yang sama.

3.7. Manfaat Penggunaan Data Mining dalam Prediksi Penjaminan Potensial

1. Menghemat Waktu Proses Pengecekan Calon
Dengan digunakannya proses Data Mining menggunakan metode Naïve Bayes, waktu yang digunakan untuk pengecekan calon terjamin menjadi lebih singkat. Biasanya untuk melakukan pengecekan >10 calon terjamin memerlukan waktu selama 1 minggu, setelah pengaplikasian data mining ini waktu pengecekan dapat dihemat menjadi satu hari.
2. Menghemat Biaya Proses Pengecekan Calon
Sebelumnya biaya yang dikeluarkan untuk pengecekan satu calon terjamin mencapai > Rp.250.000,- setelah pengaplikasian data mining ini, biaya dapat dihemat sebesar 50% - 70% . Hal ini dikarenakan setelah pengaplikasian data mining, pihak PT XYZ tidak perlu lagi melakukan double survey.
3. Mengurangi Kesalahan Penjaminan
Sebelumnya penentuan penjaminan dilakukan berdasarkan hasil survey bagian penjaminan. Dimana hasil analisa setiap orang berbeda-beda, hal itu dikarenakan setiap orang kadang memiliki kriteria tersendiri yang mengacu pada kriteria yang telah ditetapkan, selain itu juga karena adanya factor eksternal dari orang tersebut. Hal ini menyebabkan seringnya terjadi kesalahan penjaminan. Dengan penerapan sistem ini, penentuan kriteria lebih jelas, sehingga proses analisis data lebih mudah. Selain itu juga system ini maka mempermudah pihak PT XYZ dalam memperkirakan calon terjamin yang akan dijamin, sehingga dapat mengurangi kesalahan penjaminan

IV. KESIMPULAN

Berdasarkan hasil pembahasan maka penulis dapat mengambil beberapa kesimpulan antara lain :

1. Sistem klasifikasi data terjamin dengan menggunakan data mining ini dapat menghemat waktu pengecekan calon terjamin.
2. Sistem klasifikasi data terjamin dengan menggunakan data mining ini dapat menghemat biaya pengecekan calon terjamin.
3. Dengan adanya sistem ini maka mempermudah pihak PT XYZ dalam memperkirakan calon terjamin yang akan dijamin, sehingga dapat mengurangi kesalahan penjaminan.
4. Sistem ini digunakan untuk menampilkan informasi klasifikasi potensial atau tidak calon terjamin dengan menggunakan metode Naive Bayes. Dimana metode Naive Bayes ini di dukung oleh ilmu Probabilistik dan ilmu statistika, sehingga dapat digunakan sebagai data petunjuk untuk mendukung keputusan pengklasifikasian.



REFERENSI

- [1] McLeod, Jr.R. dan G.P. Schell. 2007. Management Information System. 10th ed. Pearson Education, Inc. Ali Akbar Yulianto dan Afia R. Fitriati (penterjemah). 2008. Sistem Informasi Manajemen. Edisi 10. Nina Setyaningsih (editor). Salemba Empat. Jakarta
- [2] KEPDIR No. 50-DIR-XII 2005 Sisdur Penjaminan Kredit Jamkrindo
- [3] Turban, E. 2005. Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1. Penerbit Andi. Yogyakarta.
- [4] Rakhmadi, Rezki S. 2019. Konsep dan Penerapan Sistem Jaminan Pada Lembaga Keuangan Syariah. Madani Syari'ah , Vol.2 : 1-16.
- [5] Susanto, Sani dan Suryadi, Dedy. 2010. Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data. Penerbit Andi. Yogyakarta.
- [6] Bustami. Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah Asuransi. TECHSI : Jurnal Penelitian Teknik Informatika.
- [7] Arikunto, S. 2010. Prosedur Penelitian Suatu Pendekatan Praktek. Rineka Cipta. Jakarta. Penerapan Algoritma Naïve Bayes untuk Mengklasifikasi Data Nasabah Asuransi. TECHSI : Jurnal Penelitian Teknik Informatika.
- [8] Muhammad Hendriawan, Thomas Budiman, Verdi Yasin, Asih Septia Rini (2021) PENGEMBANGAN APLIKASI E-COMMERCE DI PT. PUTRA SUMBER ABADI MENGGUNAKAN FLUTTER, "JISICOM (Journal of Information System, Informatics, and Computing)", E-ISSN: 2597-3673 (Online) , P-ISSN: 2579-5201 (Print), Vol. 4, No.1, (20 Juni 2021), Page 69-88, DOI: <https://doi.org/10.52362/jisicom.v5i1.371> URL Published : <http://journal.stmikjayakarta.ac.id/index.php/jisicom/article/view/371>
- [9] Kemas Hasyim Azhari, Thomas Budiman, Rachmawaty Haroen, Verdi Yasin (2021) ANALISIS DAN RANCANGAN MANAJEMEN PROSES BISNIS UNTUK LAYANAN PELANGGAN DI PT. PGAS TELEKOMUNIKASI NUSANTARA, "JISICOM (Journal of Information System, Informatics, and Computing)", E-ISSN: 2597-3673 (Online) , P-ISSN: 2579-5201 (Print), Vol. 4, No.1, (20 Juni 2021), Page 48-68, DOI: <https://doi.org/10.52362/jisicom.v5i1.381> URL Published : <http://journal.stmikjayakarta.ac.id/index.php/jisicom/article/view/381>
- [10] Adi Mardian, Thomas Budiman, Rachmawaty Haroen; Verdi Yasin (2021), Perancangan Aplikasi Pemantauan Kinerja Karyawan Berbasis Android di PT. Salestrade Corp. Indonesia, "Jurnal Manajemen Informatika Jayakarta", E-ISSN : 2797-0930 (Online), P-ISSN : 2746-5985 (Print), Volume 1, Nomor 3, Juli 2021, halaman 169-185, DOI: 10.52362/jmijayakarta.v1i3.481 , URL Publikasi: <http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/481>
- [11] Nandang Mulyana, Agus Sulistyanto, Verdi Yasin (2021), PERANCANGAN SISTEM INFORMASI PENGELOLAAN ASET IT BERBASIS WEB PADA PT MANDIRI AXA GENERAL INSURANCE, "Jurnal Manajemen Informatika Jayakarta", E-ISSN : 2797-0930 (Online), P-ISSN : 2746-5985 (Print), Volume 1, Nomor 3, Juli 2021, halaman 243-257, DOI: 10.52362/jmijayakarta.v1i3.498 , URL Publikasi: <http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/498>
- [12] Maulia Usnaini, Verdi Yasin, Anton Zulkarnain Sianipar (2021), PERANCANGAN SISTEM INFORMASI INVENTARISASI ASET BERBASIS WEB MENGGUNAKAN METODE WATERFALL, "Jurnal Manajemen Informatika Jayakarta", E-ISSN : 2797-0930 (Online), P-ISSN : 2746-5985 (Print) Volume 1, Nomor 1, Februari 2021, halaman 36-55, DOI: 10.52362/jmijayakarta.v1i1.415, URL Publikasi: <http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/415>
- [13] Putri Setiani, Ifan Junaedi, Anton Zulkarnain Sianipar, Verdi Yasin (2021), PERANCANGAN SISTEM INFORMASI PELAYANAN PENDUDUK BERBASIS WEBSITE DI RW 010 KELURAHAN KEAGUNGAN KECAMATAN TAMANSARI - Jakarta Barat. "Jurnal Manajemen Informatika Jayakarta", E-ISSN : 2797-0930 (Online), P-ISSN : 2746-5985 (Print) Volume 1, Nomor 1, Februari 2021, halaman 20-35, DOI: 10.52362/jmijayakarta.v1i1.414 , URL Publikasi: <http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/414>
- [14] Benni Triyono, Sri Purwanti, Verdi Yasin (2017) REKAYASA PERANGKAT LUNAK SISTEM INFORMASI PENGIRIMAN DAN PENERIMAAN SURAT ATAU PAKET BERBASIS WEB (Studi Kasus : PT. Jaya Trade Indonesia), "Journal of Information System, Applied, Management, Accounting and



DOI: 10.52362/jisamar.v6i1.684

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- Research”, e-ISSN: 2598-8719, p-ISSN: 2598-8700.Vol.1 No.1 (30 Desember 2017) p46-53
<http://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/12>
- [15] Julinda Maya Paramudita, Verdi Yasin (2019) PERANCANGAN APLIKASI SISTEM PENYEWAAN ALAT BERAT (studi kasus: PT. Jaya Alam Sarana Jakarta) , “Journal of Information System, Applied, Management, Accounting and Research”, e-ISSN: 2598-8719, p-ISSN: 2598-8700.Vol.3 No.1 (20 Februari 2019) p23-29 <http://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/73>
- [16] Murvan Awaludin, Verdi Yasin (2020) APPLICATION OF ORIENTED FAST AND ROTATED BRIEF (ORB) AND BRUTEFORCE HAMMING IN LIBRARY OPENCV FOR CLASSIFICATION OF PLANTS, “Journal of Information System, Applied, Management, Accounting and Research”, e-ISSN: 2598-8719, p-ISSN: 2598-8700.Vol.4 No.3 (14 Agustus 2020) p51-59
<http://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/247>
- [17] Ifan Junaedi, Dimas Abdillah, Verdi Yasin (2020) ANALISIS PERANCANGAN DAN PEMBANGUNAN APLIKASI BUSINESS INTELLIGENCE PENERIMAAN NEGARA BUKAN PAJAK KEMENTERIAN KEUANGAN RI, “ Journal of Information System, Applied, Management, Accounting and Research”, e-ISSN: 2598-8719, p-ISSN: 2598-8700.Vol.4 No.3 (14 Agustus 2020) p88-101 <http://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/249>
- [18] Verdi Yasin (2012) Rekayasa Perangkat Lunak Berorientasi Objek”. Jakarta: Mitra Wacana Media,
- [19] Anis Rohmadi, Verdi Yasin (2020) DESAIN DAN PENERAPAN WEBSITE TATA KELOLA PERCETAKAN PADA CV APICDESIGN KREASINDO JAKARTA DENGAN METODE PROTOTYPING, “Journal of Information System, Informatics and Computing”. E-ISSN: 2597-3673 (Online), P-ISSN: 2579-5201 (Print) Vol. 4 No.1, June 22, 2020. Pp.70-85.
<http://journal.stmikjayakarta.ac.id/index.php/jisicom/article/view/210>
- [20] Anggeri S. Nurjaman, Verdi Yasin (2020) KONSEP DESAIN APLIKASI SISTEM MANAJEMEN KEPEGAWAIAN BERBASIS WEB PADA PT. BINTANG KOMUNIKASI UTAMA (Application design concept of web-based staffing management system at PT Bintang Komunikasi Utama). “JISICOM (Journal of Information System, Informatics, and Computing)”, E-ISSN: 2597-3673 (Online) , P-ISSN: 2579-5201 (Print), Vol. 4 No.2, (28 Desember 2020), Page 143-174. DOI: <https://doi.org/10.52362/jisicom.v4i2.363> .
 URL Published : <http://journal.stmikjayakarta.ac.id/index.php/jisicom/article/view/363>
- [21] Ito Riris Immasari, Verdi Yasin (2019) PENGGUNAAN METODE ANALYTIC HIERARCHY PROCESS UNTUK MENGANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PEMILIHAN CALON LEGISLATIF DI DPRD II KOTA TANGERANG. “JISICOM (Journal of Information System, Informatics, and Computing)”, E-ISSN: 2597-3673 (Online) , P-ISSN: 2579-5201 (Print), Vol. 3 No.2, (10 Desember 2019), Page 53-58 URL Published :
<http://journal.stmikjayakarta.ac.id/index.php/jisicom/article/view/139>

