

PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR

¹Farida Gultom[✉], ²Tober Simanjuntak

¹Mahasiswa Pascasarjana Magister Teknologi Informasi, STMIK Mikroskil Medan, Indonesia

²Universitas Efarina, Pematangsiantar, Indonesia

Email: faridagultom20@gmail.com

DOI: <https://doi.org/10.46880/jmika.Vol4No2.pp98-102>

ABSTRACT

Application developers and users are the main keys to the market's impact on application development. In developing applications, it is necessary to predict applications in the market accurately, accurate prediction results are very important in showing the correct rating and decision making in the selection of new prospective debtors. Tests conducted in this test use a dataset of credit customers from Bank Danamon. This study predicts the smooth payment of credit by combining the Naïve Bayes and K-Nearest Neighbor methods. Predicting the smooth level of credit payments is done by combining the Naïve Bayes algorithm and K-Nearest Neighbor in order to predict the smooth payment of credit in the future, this can be seen from the prediction results obtained by 80%.

Keyword: *Credit, Credit Risk, K-Nearest Neighbor, Naïve Bayes.*

ABSTRAK

Pengembang dan pengguna aplikasi merupakan kunci utama pada dampak pasar terhadap pengembangan aplikasi. Dalam melakukan pengembangan aplikasi perlu memprediksi aplikasi di *market* dengan akurat, hasil prediksi yang akurat sangat penting dalam menunjukkan *rating* dan pengambilan keputusan yang tepat dalam seleksi calon debitur yang baru. Pengujian yang dilakukan dalam pengujian ini menggunakan dataset nasabah kredit dari Bank Danamon. Penelitian ini melakukan prediksi terhadap kelancaran pembayaran kredit dengan mengkombinasikan metode Naïve Bayes dan *K-Nearest Neighbor*. Prediksi tingkat kelancaran pembayaran kredit ini dilakukan dengan mengkombinasikan algoritma Naïve Bayes dan *K-Nearest Neighbor* agar dapat memprediksi kelancaran pembayaran kredit kedepannya, hal ini dapat dilihat dari hasil prediksi yang diperoleh sebesar 80%.

Kata Kunci: *Kredit, Resiko Kredit, K-Nearest Neighbor, Naïve Bayes.*

PENDAHULUAN

Kredit merupakan penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga (Pemerintah Indonesia, 1998). Pada perkembangan awalnya, kredit memiliki fungsi dalam merangsang sikap saling menolong yang bertujuan untuk mencapai kebutuhan, baik dalam bidang usaha atau pemenuhan kebutuhan sehari-hari. Kredit diharapkan dapat membawa dampak positif secara sosial ekonomis bagi debitur, kreditur, ataupun masyarakat. Pemberian kredit terhadap nasabah oleh pihak bank dapat mengalami berbagai masalah ataupun risiko. Permasalahan atau risiko yang dapat

dialami oleh bank dalam pemberian kredit adalah sikap nasabah yang tidak melakukan pembayaran angsuran secara tepat waktu ataupun menunda pembayaran sampai beberapa bulan pembayaran angsuran yang menyebabkan terjadinya kredit macet (Hasan, 2017).

Sistem penilaian kredit merupakan bagian penting dan integratif dari manajemen risiko perusahaan. Sistem ini memungkinkan pencegahan kerugian utang dengan mengidentifikasi, menganalisis dan memantau risiko kredit pelanggan (Abid, Masmoudi, & Zouari-Ghorbel, The Consumer Loan's Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank, 2018). Oleh karena itu, bank menjadi lebih mungkin untuk secara akurat memprediksi solvabilitas pemohon pinjaman dan

memprediksi tingkat kelancaran pembayaran di masa depan dalam banyak cara, seperti batas kredit dan lama pinjaman yang dialokasikan untuk pelanggan (Abid, Zaghdene, Masmoudi, & Zouari-Ghorbel, 2017). Penelitian ini melakukan pengamatan pada variabel untuk membedakan pembayaran yang lancar (*current*) dan pembayaran yang tidak lancar (*not current*). Parameter yang digunakan pada penelitian ini terdiri dari status account, jangka waktu kredit, sejarah kredit, tujuan kredit, plafon kredit, tingkat angsuran kredit sebagai persen dari pendapatan, status pribadi, jaminan, usia, jumlah kredit di bank yang sedang berjalan, dan jumlah tanggungan. Data ini akan digunakan untuk membangun model penilaian kredit yang memungkinkan prediksi kemungkinan pembayaran lancar (*current*) kredit baru.

Beberapa peneliti menggunakan *machine learning* untuk memprediksi tingkat kelancaran pembayaran kredit untuk membantu pengambilan keputusan pemberian kredit dan menganalisis kemampuan nasabah dalam membayar kredit. Penelitian membandingkan 2 metode yaitu algoritma C4.5 dan Naïve Bayes (Rifai & Aulianita, 2018). Penelitian lainnya (Hasan, 2017) menggunakan Naïve Bayes berbasis *forward selection* untuk memprediksi tingkat kelancaran pembayaran kredit dengan tingkat akurasi 71,91%. Tahun 2016, Lob Abid et al menggunakan Logistic Regression untuk analisis prediksi pembayaran pinjaman dengan menggunakan data bank (Abid, Masmoudi, & Zouari-Ghorbel, The Consumer Loan's Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank, 2018) dengan nilai akurasi 68,49%. Tahun 2018 Admel Husejinovic melakukan penelitian prediksi pembayaran kartu kredit dengan menggunakan beberapa aplikasi machine learning dan membandingkan nilai akurasi dari setiap algoritma, dimana Naïve Bayes nilai akurasi 80,6%, *K-Nearest Neighbor* 79,2%, dan C4.5 79% (Husejinovic, Keco, & Masetic, 2018). Hasil dari penelitian tersebut Naïve Bayes dan *K-Nearest Neighbor* memiliki nilai akurasi yang lebih tinggi dibanding dengan metode lainnya. Pada penelitian lainnya penggunaan *K-Nearest Neighbor* yang dikombinasikan dengan *Modified Particle Swarm Optimization* (Jamaluddin & Siringoringo, 2017). Algoritma-algoritma modifikasi tersebut dimunculkan untuk memperbaiki kelemahan dari data mining yang sudah ada sebelumnya. Kelemahan dari *K-Nearest Neighbor* berada pada proses perhitungan yang dilakukan hampir disetiap

data pada tahap klasifikasi (Sari, Ernawati, & Pranowo, 2015).

Untuk mengatasi masalah tersebut maka diusulkan suatu pendekatan dengan mengkombinasikan algoritma Naïve Bayes dengan *K-Nearest Neighbor* untuk mengatasi kelemahan dari metode *K-Nearest Neighbor*. Algoritma ini diharapkan dapat meningkatkan efektivitas waktu, keakuratan akurasi dan lama waktu proses.

TINJAUAN LITERATUR

Pada Undang-Undang Republik Indonesia Nomor 10 Tahun 1998, kredit diartikan sebagai penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

Penilaian kredit merupakan sebagai seperangkat model keputusan dengan teknik dasar spesifik untuk memberikan dukungan kepada pemberi pinjaman ketika memberikan kredit kepada pelanggan. Ini juga dianggap sebagai sistem pendukung keputusan yang membantu manajer dalam proses pengambilan keputusan keuangan. Dengan cara yang sama, perkembangan pesat dalam industri kredit telah membuat model penilaian kredit memiliki kegunaan yang menonjol karena sangat terkait dengan keputusan evaluasi penerimaan kredit. Oleh karena itu, model-model ini diterapkan untuk mengklasifikasikan aplikasi kredit sebagai *Baccepted* atau *Brejected* (Bank Indonesia, 2009)

Banyak resiko yang terkait dengan pinjaman bank, bagi bank dan bagi mereka yang mendapatkan pinjaman. Analisis resiko dalam pinjaman bank harus dipahami sebagai makna resiko. Resiko kredit adalah dimana nasabah/debitur atau *counterpart* tidak mampu memenuhi kewajiban keuangan sesuai kontrak/kesepakatan yang telah dilakukan.

Naïve Bayes

Algoritma Naïve Bayes adalah sebuah metode untuk klasifikasi yang menggunakan metode probabilitas dan statistika yang dipelopori oleh ilmuwan dari Inggris bernama Thomas Bayes. Algoritma Naive Bayes digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa-masa sebelumnya sehingga dikenal juga sebagai Teorema Bayes. Ciri utama dari Naïve Bayes terdapat pada asumsi yang sangat kuat terhadap independensi

dari masing-masing kondisi/kejadian yang pernah terjadi.

Tahapan-tahapan pada algoritma Naive Bayes adalah:

1. Perhitungan jumlah kelas/label.
2. Perhitungan Jumlah Kasus Per Kelas
3. Perkalian Semua Variable Kelas
4. Perbandingan Hasil Per Kelas

Dasar Algoritma Naive Bayes dengan rumus Bayes yaitu:

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)} \dots\dots\dots(1)$$

Ket:

x : Data dengan kelas yang belum diketahui

H : Hipotesis data merupakan suatu kelas spesifik

$P(H|x)$: Probabilitas hipotesis berdasarkan kondisi (posteriori probability)

$P(H)$: Probabilitas hipotesis (prior probability)

$P(x|H)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas H

Adapun alur dari algoritma Naive Bayes adalah:

1. Membaca data training
2. Menghitung jumlah dan probabilitik, namun apabila data numerik maka:
 - o Cari nilai rata-rata dan standar deviasi dari masing-masing parameter data numerik.
 - o Cari nilai probabilistik dengan cara perhitungan terhadap jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Dapatkan nilai pada tabel rata-rata, standar deviasi dan probabilitas.

K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) ialah algoritma yang dapat melakukan prediksi. Prinsip kerja algoritma ini sangat sederhana yakni dengan menghitung jarak paling dekat. Artinya jika ada masukkan objek yang baru yang tidak dikenali, maka algoritma K-NN akan mencari objek yang terdekat dengan objek yang baru dimasukkan pada basis data, dan kemudian melakukan tindakan yang sama kepada objek yang baru dimasukkan kepada objek yang terdekat (Islam, Wu, Ahmadi, & Sid-Ahmed, 2010). Ada beberapa persamaan untuk mengukur jarak yaitu :

1. *Euclidean Distance*

$$D(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \dots\dots\dots(2)$$

$D(x,y)$ dibaca distance x dan y

2. *Manhattan Distance*

$$D(x, y) = \sum_{i=1}^m |x_i - y_i| \dots\dots\dots(3)$$

Rumus ini mencari jarak hanya menjumlahkan semua selisih dari jarak x_i dan y_i

3. *Minkowsky Distance*

$$D(x, y) = (\sum_{i=1}^m |x_i - y_i|^{1/r}) \dots\dots\dots(4)$$

4. *Chebychev Distance*

$$D(x, y) = \max|x_i - y_i| \dots\dots\dots(5)$$

Algoritma K-Nearest Neighbor:

1. $Z = (x', y')$ adalah data uji dengan data x' dan label kelas y' yang belum diketahui
2. C adalah himpunan label kelas data.
3. Hitung jarak (x') jarak di antara data uji z ke setiap vector data latih, simpan dalam D .
4. Pilih $Dz \in D$, yaitu K tetangga terdekat dari z

METODOLOGI PENELITIAN

Permasalahan yang akan diselesaikan dalam penelitian ini yaitu :

1. Masalah yang pertama yang ingin diselesaikan pada penelitian ini yaitu mencatat semua data yang berhubungan dengan seluruh aspek dari nasabah kredit termasuk data pribadi dari nasabah.
2. Masalah yang kedua yang ingin diselesaikan pada penelitian ini yaitu parameter yang menentukan kelancaran pembayaran kredit dari nasabah.
3. Masalah yang terakhir yaitu memprediksi tingkatkelancaran pembayaran nasabah

Berdasarkan analisis masalah di atas, maka ketiga masalah tersebut dapat mempengaruhi kelancaran pembayaran kredit oleh nasabah.

Preprocessing Data

Data yang digunakan pada tahapan *preprocessing* adalah dataset loan dengan jumlah data sebanyak 363 data dan atribut sebanyak 14 atribut. Dari 14 atribut akan digunakan 10 atribut. Atribut yang digunakan yaitu atribut 3, 4, 6, 8, 9, 10, 11, 12, 13, 14 dan dari 10 atribut tersebut ada 3 data yang berbentuk kategorik yaitu atribut 3, 12, 14. Sampel Dataset dapat dilihat pada Tabel berikut:

Tabel 1. Data Preprocessing

No	1	2	3	4	5	6	7	8	9	10
1	1	550.000.000	75.000.000	750.000.000	15.416.659	3565367418	4	1	60	2
2	2	160.000.000	85.000.000	255.000.000	6.500.000	3503726568	4	2	48	1
3	2	30.000.000	15.000.000	90.000.000	1.300.000	3581004334	21	1	24	1
4	2	60.000.000	30.000.000	165.000.000	2.197.333	3580888653	22	3	30	1
5	2	35.000.000	16.000.000	60.000.000	1.672.223	3565899212	18	1	30	1
6	1	170.000.000	55.000.000	185.000.000	5.043.336	3569028107	26	3	48	1
7	2	95.000.000	45.000.000	150.000.000	4.500.000	3512894357	17	1	42	1
8	1	150.000.000	45.000.000	185.000.000	5.074.996	3556839425	8	1	36	1
9	2	50.000.000	35.000.000	80.000.000	2.288.889	3582730523	10	1	36	1
....
363	3	50.000.000	30.000.000	150.000.000	2.288.889	3590874107	20	1	36	1

HASIL DAN PEMBAHASAN

Setelah aplikasi dibangun, maka selanjutnya adalah melakukan pembuktian penerapan model algoritma Naïve Bayes dan *K-Nearest Neighbor* dalam memprediksi kelancaran pembayaran kredit dengan menampilkan nilai akurasi yang lebih baik.

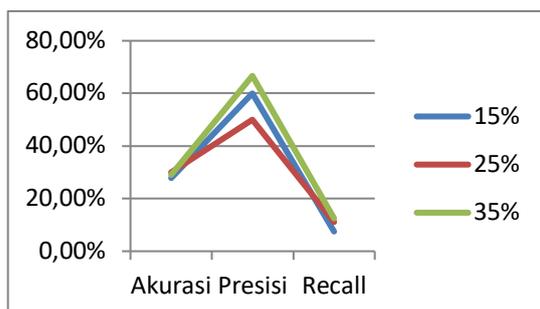
Tabel 2. Data Hasil Testing

Naïve Bayes	Data Testing		
	15%	25%	35%
Akurasi	27,78%	30%	29%
Presisi	60,00%	50%	66,67%
Recall	7,5%	11,11%	12,50%
<i>K-NN dan NB</i>			
Akurasi	79,63%	80%	78,75%
Presisi	80,77%	81,40%	78,84%
Recall	97,67%	97,22%	98,02%

Pada bagian ini, akan dilakukan evaluasi terhadap hasil pengujian yang telah dilakukan. Pengujian pertama dilakukan dengan menggunakan metode Naïve Bayes dengan dataset 15%, 25%, dan 35%. Pengujian akan menggunakan akurasi, dan hasil pengujian dari metode ini dapat dilihat pada table berikut:

Tabel 3. Hasil akurasi Naïve Bayes

Data Testing	Akurasi	Presisi	Recall
15%	27,78%	60,00%	7,50%
25%	30%	50%	11,11%
35%	29%	66,67%	12,50%



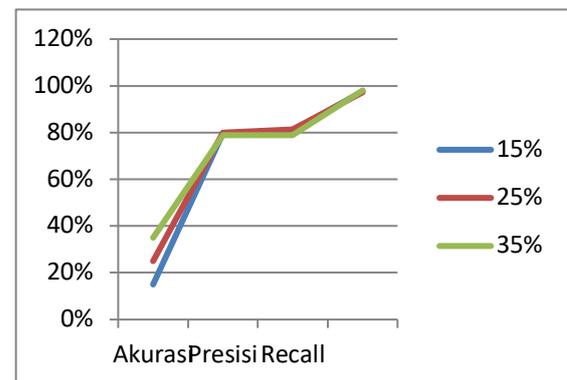
Gambar 1. Grafik Nilai Akurasi dengan menggunakan Naïve Bayes

Hasil akurasi Naïve Bayes kombinasi *K-Nearest Neighbor*

Hasil akurasi dengan kombinasi Naïve Bayes dan *K-Nearest Neighbor* ditunjukkan pada Tabel dan grafik berikut:

Tabel 3. Hasil akurasi Naïve Bayes kombinasi *K-Nearest Neighbor*

Data Testing	Akurasi	Presisi	Recall
15%	79,63%	80,77%	97,67%
25%	80,00%	81,40%	97,22%
35%	78,74%	78,84%	98,02



Gambar 2. Grafik Nilai akurasi kombinasi algoritma Naïve Bayes dan *K-NN*

Dari hasil pengujian diatas dapat dilihat perbandingan menggunakan algoritma Naïve Bayes dan Kombinasi Naïve Bayes dan KNN

Tabel 4. Perbandingan Akurasi

Algoritma	Akurasi	Presisi	Recall
NB	30%	50%	11,11%
Combine NB dan KNN	80%	78%	97,22%

Dari hasil pengujian yg telah dilakukan dapat dilihat hasil pengukuran kinerja model yang terbaik. Pada saat pengujian dilakukan dengan hanya menggunakan satu metode yaitu Naïve Bayes hasil tingkat nilai akurasi masih rendah. Akan tetapi setelah dilakukan pengujian dengan kombinasi 2 metode yaitu Naïve Bayes dan *K-Nearest Neighbor* di dapatkan tingkat nilai akurasi yang lebih baik dengan nilai 80,00%.

KESIMPULAN

Hasil pengujian yang telah dilakukan dengan menggunakan metode Naïve Bayes dan kombinasi Naïve Bayes dengan *K-Nearest Neighbor* maka dapat disimpulkan sebagai berikut:

1. Pada proses pengujian dengan penggunaan metode Naïve Bayes dan pengujian yang dilakukan dengan menggunakan *K-Nearest Neighbor* didapat hasil tingkat akurasi yang sangat rendah.
2. Pengujian yang dilakukan dengan mengkombinasikan algoritma Naïve Bayes dan *K-Nearest Neighbor* dengan pengujian yang dilakukan sebanyak 3 kali, hasil akurasi yang tertinggi yang dihasilkan 80%

DAFTAR PUSTAKA

- Abid, L., Masmoudi, A., & Zouari-Ghorbel, S. (2018). The Consumer Loan's Payment Default Predictive Model: an Application of the Logistic Regression and the Discriminant Analysis in a Tunisian Commercial Bank. *Journal of the Knowledge Economy*, 9 (3), 948-962.
- Abid, L., Zaghdene, S., Masmoudi, A., & Zouari-Ghorbel, S. (2017). Bayesian Network Modeling: A Case Study of Credit Scoring Analysis of Consumer Loans Default Payment. *Asian Economic and Financial Review*, 7 (9), 846-857.
- Bank Indonesia. (2009). Peraturan Bank Indonesia No 11 Tahun 2009. *PBI No. 11/25/PBI/2009*. Jakarta.
- Hasan, M. (2017). Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naïve Bayes Berbasis Forward Selection. *ILKOM Jurnal Ilmiah*, 9 (3), 317-324.
- Husejinovic, A., Keco, D., & Masetic, Z. (2018). Application of Machine Learning Algorithms in Credit Card Default Payment Prediction. *International Journal Of Scientific Research*, 7 (10), 425-426.
- Islam, M. J., Wu, Q. M., Ahmadi, M., & Sid-Ahmed, M. A. (2010). Investigating the Performance of Naïve-Bayes Classifiers and *K-Nearest Neighbor* Classifiers. *Journal of Convergence Information Technology*, 5 (2), 133-137.
- Jamaluddin, & Siringoringo, R. (2017). Improved Fuzzy *K-Nearest Neighbor* Using Modified Particle Swarm Optimization. *Journal of Physics: Conference Series*, 930 (1/012024), 1-7.
- Pemerintah Indonesia. (1998). Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 tentang Perubahan atas Undang-Undang Nomor 7 Tahun 1992 tentang Perbankan. Jakarta: Sekretariat Negara.
- Rifai, A., & Aulianita, R. (2018). Komparasi Algoritma Klasifikasi C4.5 dan Naïve Bayes Berbasis Particle Swarm Optimization Untuk Penentuan Resiko Kredit. *Journal Speed – Sentra Penelitian Engineering dan Edukasi*, 10 (2), 49-55.
- Sari, M. K., Ernawati, & Pranowo. (2015). Kombinasi Metode *K-Nearest Neighbor* Dan Naïve Bayes Untuk Klasifikasi Data. *Seminar Nasional Teknologi dan Multimedia (Semnasteknomedia) 2015*. 3, pp. 32-37. Yogyakarta: Amikom.