

---

## **Analisis Risiko Kredit Perbankan Menggunakan Algoritma *K-Nearest Neighbor* dan *Nearest Weighted K-Nearest Neighbor***

---

**Dian Tri Wilujeng\*, Mohamat Fatekurohman, and I Made Tirta**

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember

*\*Corresponding author:* diantriwilujeng32@gmail.com

---

**Abstract.** Bank is a business entity that collects public funds in the form of savings and also distributes them to the public in the form of credit or other forms. Credit risk analysis can be done in various ways such as marketing analysis and big data using machine learning. One example of a machine learning algorithm is *K-Nearest Neighbor (KNN)* and the development of the *K-Nearest Neighbor* algorithm is *Neighbor Weighted KNearest Neighbor (NWKNN)*. The *K-Nearest Neighbor (KNN)* algorithm is one of the machine learning methods that can be used to facilitate the classification of complex data. The purpose of this study is to determine the results of the application of the algorithm and the comparison of the use of the *KNN* and *NWKNN* algorithms in banking credit. The results obtained are that *NWKNN* is able to predict credit risk better, especially in classifying potential customers with potential losses compared to *KNN*.

**Keywords:** Machine learning, *KNN*, *NWKNN*

---

### **1. PENDAHULUAN**

Bank adalah badan usaha yang menghimpun dana masyarakat dalam bentuk simpanan dan juga menyalurkannya kepada masyarakat dalam bentuk kredit atau juga bentuk bentuk lainnya dalam rangka untuk meningkatkan taraf hidup rakyat banyak [6]. Masalah seringkali muncul ketika nasabah tidak melakukan tanggung jawabnya maka terjadi proses tersebut yang dinamakan kredit macet atau *bad debt*. Semakin lama jangka waktu peminjaman maka semakin besar peluang terjadinya kredit macet. Oleh karena itu, perlu dilakukan analisis risiko kredit terhadap nasabah.

Analisis risiko kredit dapat dilakukan dengan berbagai cara seperti analisis pemasaran dan *big data* dengan menggunakan *machine learning*. Metode *machine learning* merupakan salah satu metode klasifikasi statistika yang dapat dilakukan untuk menganalisis risiko kredit. Salah satu contoh algoritma *machine learning* yaitu *K-Nearest Neighbor (KNN)* dan pengembangan dari algoritma *K-Nearest Neighbor* yaitu *Neighbor Weighted K-Nearest Neighbor (NWKNN)*. Algoritma *K-Nearest Neighbor (KNN)* merupakan salah satu metode *machine learning* yang dapat digunakan untuk memudahkan klasifikasi data kompleks. *KNN* dilakukan dengan mencari kelompok *k* objek dalam data *training* yang paling dekat dengan objek data baru atau data *testing*

[9]. Algoritma NWKNN merupakan pengembangan dari KNN dengan menggunakan pembobotan pada kelas data.

## 2. TINJAUAN PUSTAKA

### 2.1 Kredit

Kredit ialah suatu penyediaan tagihan dan uang yang bisa disamakan yang berdasarkan dengan kesepakatan atau persetujuan pinjam meminjam antara pihak bank dengan pihak lainnya dan untuk mewajibkan peminjam untuk melunasi hutangnya dengan jumlah bunga, imbalan atau bagi hasilnya dalam jangka waktu yang sudah ditentukan [6].

### 2.2 Seleksi Fitur

Seleksi fitur adalah salah satu teknik data mining yang umum digunakan pada tahapan pre-processing. Teknik ini digunakan untuk mengurangi kompleksitas atribut(variabel) yang akan dikelola pada processing dan analisis. Teknik ini dilakukan untuk mengetahui subset fitur yang paling signifikan dari data set. Dengan cara ini akan membantu mempercepat proses pembelajaran/pemodelan[4].

### 2.3 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma pengklasifikasian sebuah subjek baru yang bersumber pada data *training sample* serta atribut. Nantinya hasil dari sampel uji baru digolongkan sesuai dengan mayoritas dari kategori pada KNN. Algoritma *K-Nearest Neighbor* (KNN) menggunakan klasifikasi ketetanggaan yang digunakan sebagai nilai prediksi dari sampel uji yang baru [7]. Langkah pada algoritma KNN dapat diawali dengan penentuan K, kemudian dihitung nilai kedekatan ketetanggaan antara data *training* dengan data *testing* memakai persamaan *Cosine Similarity* (*CosSim*). Perhitungan persamaan *CosSim* dapat diformulasikan sebagai berikut:

$$CosSim(q, d_j) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^n (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^n (w_{ij})^2} \sqrt{\sum_{i=1}^n (w_{iq})^2}} \quad (1)$$

### 2.4 Neighbor Weighted K-Nearest Neighbor

Metode *Neighbor Weighted K-Nearest Neighbor* adalah algoritma perkembangan pada KNN. Algoritma NWKNN menggunakan pembobotan pada kelas data. Bobot yang akan diberikan pada kelas minoritas lebih besar dan bobot lebih sedikit pada tetangga yang berasal dari kelas yang mayoritas [5].

Perhitungan bobot dapat digunakan sebagai berikut:

$$Weight = \frac{1}{\left( \frac{Num(c_i^d)}{Min\{Num(c_j^d) = 1, \dots, K\}} \right)^{1/exp}} \quad (2)$$

### 2.5 Confusion Matrix

Performa dari suatu model klasifikasi dapat diukur dengan tingkat akurasi berdasarkan *confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. TP dan TN

memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberikan informasi ketika *classifier* salah [4].

## 2.6 Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positives* sebagai garis vertikal [10]. Kurva ROC dihasilkan dengan mengubah ambang batas pada skor kepercayaan; karenanya, setiap ambang hanya menghasilkan satu titik dalam kurva ROC [2].

## 2.7 Area Under the Curve

AUC digunakan untuk menghitung area dibawah kurva ROC. Nilai AUC dapat dihitung dengan menambahkan area trapesium dari ukuran AUC. Nilai AUC berada diantara 0 dan 1 dan semakin mendekati angka 1 maka semakin baik model dalam mengklasifikasi data [1]. Nilai AUC dapat dibagi menjadi beberapa kelompok yaitu [3].

1. 0,90 – 1,00 = *Excellent Classification*
2. 0,80 – 0,90 = *Good Classification*
3. 0,70 – 0,80 = *Fair Classification*
4. 0,60 – 0,70 = *Poor Classification*
5. 0,50 – 0,60 = *Failure*

## 2.8 Normalisasi Data

Metode min-maks menggunakan nilai minimum dan maksimum untuk mengkonversi data secara linier. Misalkan A merupakan atribut bertipe numerik, maka suatu nilai  $x_i$  dapat dinormalisasi menjadi nilai baru  $x_i^1$  yang berada dalam rentang [ $minbaru_A$ ,  $maksbaru_A$ ] dengan formula:

$$x_i^1 = \frac{x_i - min_A}{maks_A - min_A} (maksbaru_A - minbaru_A) + minbaru_A \quad (3)$$

## 3. METODE PENELITIAN

Jenis data yang digunakan pada penelitian ini adalah data sekunder oleh pihak bank yang telah dikumpulkan. Dataset terdiri dari 8 variabel *independent* yaitu jenis kelamin, plafond, rate, jangka waktu, pekerjaan, pendapatan, jaminan, dan riwayat pinjaman. Data nasabah yang telah dikumpulkan akan dikelompokkan menjadi status kredit lancar dan macet. Dataset terdiri dari 73% data berstatus lancar atau sebanyak 400 data dan 27% data yang berstatus macet atau sebanyak 148 data. Pemecahan data dibagi dua yaitu data *training* data data *testing*. Terdapat tiga perbandingan data setiap model yaitu 70%:30%, 80%:20%, dan 90%:10%. Sebanyak 384 data *training* dan 164 data *testing* untuk perbandingan 70%:30%, sebanyak 439 data *training* dan 109 data *testing* untuk perbandingan 80%:20% dan sebanyak 494 data *training* dan 54 data *testing* untuk perbandingan 90%:10%.

Adapun tahapan metode yang akan dilakukan untuk penyelesaian penelitian ini sebagai berikut :

1. Pengumpulan data dilakukan dengan teknik *observasi* dengan cara mengumpulkan informasi secara langsung ke lokasi oleh pihak BRI Unit Asembagus Situbondo. Data yang diperoleh sebanyak lebih dari 548 data. Data yang diperoleh akan dibagi menjadi 2 kelompok data yaitu *data training* dan *data testing*.

2. *Preprocessing data* terdiri dari dua tahapan yaitu data *cleaning* untuk membersihkan data yang tidak lengkap dan pendefinisian variabel.
3. Normalisasi Data.
4. Membagi data dua yaitu data *training* dan data *testing*. Pemecahan data dilakukan dengan beberapa perbandingan 70%:30%, 80%:20%, 90%:10%.
5. Penerapan algoritma dengan *software R packages* yang digunakan ialah *class*, KNN, *caret*, dan ROCR. *Packages class* diterapkan pada model KNN, KNN digunakan untuk penerapan terhadap NWKNN, *caret* berfungsi sebagai perhitungan pada *confusion matrix* serta *packages ROCR* berfungsi untuk perhitungan kurva ROC dan AUC.
6. Evaluasi hasil tahap ini dengan beberapa tahap koreksi yaitu dengan *confusion matrix* dengan TPR, TNR, FPR, FNR.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. *Preprocessing Data*

Pengecekan data perlu dilakukan agar program dapat diproses. Data yang digunakan untuk membangun model tidak boleh mengandung baris dan kolom kosong sehingga tahap data *cleaning* perlu dilakukan untuk mengetahui kelengkapan data. Hasil pengecekan data bahwa dataset lengkap sehingga dapat dilanjutkan ke tahap berikutnya. Total variabel secara keseluruhan menjadi 14 kolom. Dataset yang bertambah dengan nilai yang beragam dapat mempersulit proses klasifikasi sehingga perlu dilakukan normalisasi data. Proses normalisasi data menggunakan metode *min-maks* bertujuan untuk mengubah data pada masing-masing variabel agar berada pada rentang yang sama. Data yang telah dinormalisasi akan digunakan untuk tahap selanjutnya.

Berikut merupakan contoh 5 data yang telah diubah menjadi tipe data numerik.

Tabel 1 Dataset setelah Normalisasi

JK_P	Plafond	Rate	Jangka.waktu	PP	Pendapatan	J_SHM	RP_L	SK_L
0	0,012096774	0,251515152	0	1	0,026845638	1	1	1
0	0,02016129	0,045454545	0,5	1	0,035234899	0	1	1
0	0,02016129	0,309366391	0,5	1	0,040268456	0	1	1
0	0,012096774	0,309366391	0,25	1	0,020134228	0	1	1
1	0,004032258	0,015426997	0,25	1	0,013422819	1	1	1

##### 4.2. Implementasi Program

*Package* yang digunakan para klasifikasi model knn ini menggunakan *package class*. Klasifikasi pada model ini yang pertama perlu dilakukan yaitu penentuan nilai parameter *k*. Proses penentuan nilai parameter *k* tidak memiliki acuan khusus, sehingga dapat menentukan sendiri nilai *k* sesuai dengan permasalahan. Nilai *k* yang digunakan pada model ini yaitu 15. Nilai akurasi untuk perbandingan 70:30 yaitu 95,34% dan nilai AUC 0,9420. Perbandingan yang kedua yaitu 80:20 mendapatkan nilai akurasi 93,58% dan nilai AUC 0,8917. Terdapat perbandingan ketiga yaitu 90:10 mendapatkan nilai akurasi sebesar 87,04% dan nilai AUC 0,65. Nilai AUC terbaik ditunjukkan pada perbandingan 70:30 dengan kategori *excellent classification*. Proses klasifikasi ini tidak jauh berbeda dengan model klasifikasi yang pertama. Proses yang tidak jauh berbeda ini menghasilkan nilai yang relatif hampir sama. Parameter yang digunakan juga sama yaitu nilai *k* = 15. nilai yang diperoleh dengan menggunakan model ini yaitu perbandingan

pertama mendapat nilai akurasi 95,34% dan nilai AUC 0,9492. Hasil yang diperoleh dari perbandingan kedua mendapatkan nilai akurasi 93,58% dan nilai AUC 0,8917. Hasil yang diperoleh dari perbandingan ketiga yaitu nilai akurasi 92,59% dan nilai AUC 0,8. Nilai AUC terbaik ditunjukkan pada perbandingan 70:30 dengan selisih nilai sedikit lebih tinggi dari model KNN yaitu 0,9534 dengan kategori AUC *excellent classification*.

### 4.3. Perbandingan Model

Penentuan hasil terbaik dilihat dari nilai AUC dan TPR (proporsi kelas macet yang benar macet) tertinggi dan nilai FNR (eror kelas macet) terendah. Hal tersebut karena perbankan membutuhkan nilai keakuratan pada kelas macet yang baik untuk menghindari kerugian dari kredit macet. Hasil prediksi yang diperoleh dapat dibandingkan dengan nilai aktualnya kemudian *confusion matrix*.

Tabel 2. Evaluasi Model KNN untuk Prediksi Data *Testing*

	ACC	TPR	TNR	FPR	FNR	AUC
70:30	<b>0,9573</b>	<b>0,9091</b>	0,9750	0,025	<b>0,0909</b>	<b>0,9420</b>
80:20	0,9358	0,8077	<b>0,9759</b>	<b>0,0241</b>	0,1923	0,8917
90:10	0,8704	0,3	1,000	0	0,9	0,65

Tabel 2. menunjukkan hasil prediksi pada model KNN. Nilai akurasi ketiga perbandingan yaitu 0,9573, 0,9358 dan 0,8704 artinya semua perbandingan memiliki nilai akurasi yang baik diatas 0.8. Perbandingan terbaik berdasarkan nilai akurasi yang diperoleh yaitu pada perbandingan 70:30 dengan memperoleh nilai akurasi 0,9573. Nilai TPR (*sensitivity*) masing-masing perbandingan yaitu 0,9091, 0,8077 dan 0,3. Nilai *sensitivity* terbaik ditunjukkan pada perbandingan 70:30 yaitu 0,9091. Nilai TNR terbaik ditunjukkan pada perbandingan 80:20 yaitu 0,9759. Nilai FPR terendah diperoleh pada perbandingan 80:20 yaitu 0,0241. Nilai FNR terendah ditunjukkan pada perbandingan 70:30 yaitu 0,0909. Nilai AUC terbaik dari ketiga perbandingan ditunjukkan pada perbandingan 70:30 yaitu 0,9420.

Tabel 3. Evaluasi Model NWKNN untuk Prediksi Data *Testing*

	ACC	TPR	TNR	FPR	FNR	AUC
70:30	<b>0,9573</b>	<b>0,9318</b>	0,9667	0,0333	<b>0,0682</b>	<b>0,9492</b>
80:20	0,9358	0,8077	<b>0,9759</b>	<b>0,0214</b>	0,1923	0,8917
90:10	0,9259	0,6000	1,000	0	0,4000	0,8

Tabel 3. menunjukkan hasil prediksi pada model NWKNN. Nilai akurasi ketiga perbandingan yaitu 0,9573, 0,9358 dan 0,9259 artinya semua perbandingan memiliki nilai akurasi yang baik diatas 0.9. Perbandingan terbaik berdasarkan nilai akurasi yang diperoleh yaitu pada perbandingan 70:30 dengan memperoleh nilai akurasi 0,9634. Nilai TPR (*sensitivity*) masing-masing perbandingan yaitu 0,9318, 0,8077 dan 0,6. Nilai *sensitivity* terbaik ditunjukkan pada perbandingan 70:30 yaitu 0,9318. Nilai TNR terbaik ditunjukkan pada perbandingan 80:20 yaitu 0,9667. Nilai FPR terendah diperoleh pada perbandingan 90:10 yaitu 0. Nilai FNR terendah ditunjukkan pada perbandingan 70:30 yaitu 0,0682. Nilai AUC terbaik dari ketiga perbandingan ditunjukkan pada perbandingan 70:30 yaitu 0,9492. Oleh karena itu, dari indikator-indikator pengukur performa model secara keseluruhan hasil klasifikasi terbaik ditunjukkan pada perbandingan 70:30.

Berdasarkan dari kedua model algoritma yang digunakan memiliki nilai yang relatif sama atau memiliki selisih yang sangat kecil. Pada model KNN dibandingkan dengan ketiga percobaan, perbandingan 70:30 yang memiliki nilai akurasi paling tinggi yaitu 95,73% dan nilai AUC

0,9420. Model kedua variasi dari model KNN sendiri yaitu NWKNN menunjukkan nilai yang sama dengan model KNN dengan nilai paling tinggi tunjukkan pada perbandingan 70:30 dengan memperoleh nilai akurasi sebesar 95,73% dan nilai AUC 0,9492. Berdasarkan evaluasi model dalam ditunjukkan model NWKNN memperoleh nilai akurasi dan nilai AUC yang baik yaitu diatas 8 artinya mampu mengklasifikasi data secara baik.

Penerapan algoritma KNN dan NWKNN dalam penelitian ini secara keseluruhan mampu mengenali nasabah lancar maupun macet secara baik. Beberapa faktor perlu diperhatikan pihak bank dalam memberikan kredit, salah satu diantaranya yaitu riwayat pinjaman calon nasabah. Riwayat pinjaman calon nasabah dalam penelitian ini memiliki kontribusi paling tinggi dimana riwayat pinjaman yang berstatus lancar berpeluang besar untuk melakukan kredit pada pinjaman berikutnya. Sebaliknya dengan riwayat pinjaman berstatus menunggak berpeluang lebih besar macet untuk pinjaman selanjutnya. Sedemikian sehingga riwayat pinjaman perlu diperhatikan oleh pihak bank dalam keputusan memberikan kredit tersebut kepada calon nasabah.

## 5. KESIMPULAN

Hasil yang diperoleh dari penerapan algoritma model KNN dan NWKNN dalam memprediksi risiko kredit adalah baik, hal itu ditunjukkan oleh nilai akurasi, *sensitivity*, *specificity*, nilai AUC yang tinggi dan nilai FPR dan FNR yang cukup rendah. Keakuratan model NWKNN dan KNN dalam memprediksi risiko kredit dapat dilihat dari nilai akurasi kedua model lebih dari 80% artinya kedua model memberikan hasil yang akurat. Berbeda dengan model NWKNN secara keseluruhan berdasarkan indikator-indikator evaluasi model menunjukkan hasil yang baik dari ketiga perbandingan. Hasil terbaik model NWKNN ditunjukkan pada perbandingan 70:30 dengan nilai akurasi 95,73% dan nilai AUC 94,92%. Hasil yang diperoleh dari kedua model menunjukkan bahwa model NWKNN mampu memprediksi risiko kredit lebih baik terutama dalam mengklasifikasikan calon nasabah yang berpotensi macet.

## DAFTAR PUSTAKA

- [1] Bradley, A. P, "The Use of the Area Under the Curve in The Evaluation of Machine Learning Algorithms", Pattern Recognition: Department of Electrical and Computer Engineering, vol. 30, no. 7, pp. 1145-1159, 1997.
- [2] Fawcett, T, "An Introduction to ROC Analysis", Pattern Recognition Letter: Institute for the Study of Learning and Expertise, vol. 27, no. 8, pp. 861-874, 2006.
- [3] Gorunescu, F, Data Mining Concepts Model and Techniques, Intelligent Systems Reference Library: Springer-Verlag Berlin Heidelberg, 2011.
- [4] Han, J., M. Kamber, dan J. Pei, Data mining: Concepts and Techniques, USA: Elsevier, 2012.
- [5] Indriati dan A. Ridok, "Sentiment Analysis for Review Mobile Application Using Neighbor Weighted K-Nearest Neighbor(NWKNN)", Journal of Enviromental Engineering & Sustainable Technology, vol. 2, no. 19, pp. 23-32, 2016.
- [6] Kementerian Sekretariat Negara, Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 Tentang Perubahan atas Undang Undang Nomor 7 Tahun 1992 tentang Perbankan, Jakarta: Kementerian Sekretariat Negara RI.

- [7] Krisandi, N., Helmi, dan B. Prihandoso, “Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. MINAMAS Kecamatan Parindu”, Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster), vol. 2, no. 5, pp. 33-38, 2013.
- [8] Nugroho, A., Kusriani dan R. Arief, “Sistem Pendukung Keputusan Kredit Usaha Rakyat PT. Bank Rakyat Indonesia Unit Kaluangkrik Magelang”, Citec Journal, vol. 2, no. 1, pp. 1053-1068, 2015.
- [9] Utami, S., M. A. Mukid dan Sugito, “Klasifikasi Kinerja Perusahaan di Indonesia dengan Menggunakan Metode Weighted K Nearest Neighbor (Studi Kasus: 436 Perusahaan Yang Terdaftar Di Bursa Efek Indonesia Tahun 2015)”, Jurnal Gaussian, vol. 6, no. 2, pp. 181-191, 2017.
- [10] Vercellis, Carlo, *Business Intelligence: Data mining and Optimization for Decision Making*, Chichester: John Wiley and Sons, 2009.