

## KLASIFIKASI PENENTUAN *CUSTOMER* KREDIT MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Afri Yudha<sup>1</sup>, Yosep Nuryaman<sup>2</sup>, Budi Prasetya<sup>3</sup>

<sup>1,3</sup>Dosen Program Studi Teknologi Informasi Universitas Darma Persada

<sup>2</sup>Dosen Program Studi Teknologi Informasi Universitas Bina Sarana Informatika

email : ibnugazali@gmail.com

### ABSTRAK

*Sebagai salah satu supplier suku cadang kendaraan untuk perusahaan-perusahaan, PT Denco mencoba memberikan pelayanan terbaik untuk para pelanggannya baik dari segi kualitas maupun sistem pembayaran kredit yang memudahkan mereka. Namun dikarenakan adanya pandemic covid 19, banyak pelanggan terkena imbas langsung yang mengakibatkan pembayaran mereka menjadi terganggu. Sebagai langkah untuk menjaga stabilitas keuangan, maka diperlukan suatu cara untuk menentukan pelanggan-pelanggan yang ada untuk dipertahankan sebagai pelanggan kredit. KNN merupakan salah satu algoritma yang tepat untuk melakukan klasifikasi. Oleh sebab itu penulis mencoba mengklasifikasi pelanggan-pelanggan tersebut dan menghasilkan nilai accuracy 84.86% +/-4,26%, serta nilai AUC 0,692 +/- 0,121*

**Kata Kunci:** kredit, data mining, , knn.

### 1. PENDAHULUAN

#### 1.1. Latar Belakang

Pandemi Covid 19 yang telah berlangsung dari awal tahun 2020 telah membuat berbagai lini bisnis menghadapi masalah yang serius. Bahkan diantaranya harus melakukan penutupan kegiatan usaha secara permanen.

Salah satu perusahaan yang terkena imbasnya yaitu PT Deco yang bergerak di bidang penyediaan suku cadang kendaraan. Banyak pelanggan-pelanggan yang mengalami keterlambatan pembayaran dikarenakan pandemic ini. "Algoritma KNN baik digunakan dalam klasifikasi nasabah kredit koperasi, sehingga Koperasi dapat mencegah resiko kredit nasabah mereka"[4].

Oleh sebab itu untuk mengatasi permasalahan PT Deco, maka penulis mencoba menganalisa data yang ada menggunakan algoritma knn untuk memberikan rekomendasi terbaik dalam menentukan customer kredit.

#### 1.2. Rumusan Masalah

Berdasarkan latar belakang yang penulis jabarkan diatas maka yang menjadi rumusan masalah penelitian ini sebagai berikut yaitu, bagaimana cara mengklasifikasi dan mengetahui akurasi penentuan customer kredit menggunakan algoritma k-nearest neighbor

#### 1.3. Batasan Masalah

Untuk membatasi penelitian ini agar tidak terlalu luas, maka penulis hanya membatasi penelitian pada mencari akurasi dan klasifikasi penentuan klasifikasi penentuan customer kredit menggunakan algoritma k-nearest neighbor.

#### 1.4. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui cara cara mengklasifikasi dan mengetahui akurasi penentuan customer kredit menggunakan algoritma k-nearest neighbor

## 2. LANDASAN TEORI

### 2.1. Bahan Dan Metode

Sebagai bahan penelitian, penulis telah mengumpulkan data yang diperlukan dengan metode penelitian yang penulis gunakan adalah sebagai berikut :

#### A. Observasi

Penulis melakukan observasi langsung mulai tanggal 01 Januari-31 Januari 2021, untuk mengumpulkan data daftar customer selama pandemi.

#### B. Wawancara

Wawancara dilakukan kepada bagian penjualan terkait data-data yang sudah terkumpul dalam proses observasi.

#### C. Studi Pustaka

Studi Pustaka dilakukan untuk memperkuat penelitian. Dengan membaca beberapa jurnal ilmiah, buku-buku yang di baca di perpustakaan, artikel di internet yang berhubungan dengan tema penulisan penelitian

### 2.2. Studi Literatur

#### 2.2.1. Data Mining

“Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstarksi pengetahuan (knowledge) secara otomatis.[1] “

Menambahkan definisi lain dari data mining “ merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sah (Sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database).[1]”

#### 2.2.2. KDD (*Knowledge discovery in database*)

Istilah data mining dan KDD seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar[5]. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu hapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut :

1. *Data Selection*
2. *Pre-processing/ Cleaning*
3. *Transformation*
4. *Data mining*
5. *Interpretation/ Evalution*

#### 2.2.2. Klasifikasi

“Klasifikasi adalah teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan.[1]”

### 2.2.3. KNN

Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat [7].

Teknik *K-Nearest Neighbor* dengan melakukan langkah-langkah yaitu mulai input : Data training, label data training, k, data testing.

- a. Untuk semua data testing, hitung jaraknya ke setiap data training.
- b. Tentukan k data training yang jaraknya paling dekat dengan data.
- c. Testing
- d. Periksa label data dari k data ini.
- e. Tentukan label yang frekuensinya paling banyak.
- f. Masukkan data testing ke kelas dengan frekuensi paling banyak
- g. Berhenti.

Label untuk semua data testing didapat.

Untuk menghitung jarak antara dua titik x dan y bisa digunakan jarak Euclidean sebagai berikut :

$$d(X_1, Y_2) = \sum_1 \frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \quad (1)$$

Dimana:  $X_1, 1=1,2$ , adalah atribut katagori, dan  $n_{ij}, n_1$  mewakili frekuensi yang sesuai.

### 2.2.4. Rapid Miner

Rapid Miner merupakan “perangkat lunak yang bersifat terbuka (open source). Rapid Miner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi Rapid Miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik”[1].

Rapid Miner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari Universit of Dortmund[5]. Rapid Miner menyediakan GUI (Graphical User Interface) untuk merancang sebuah pipeline analisis. GUI ini akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analisis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh Rapid Miner untuk menjalankan analisis secara otomatis[5].

## 3. STUDI PERENCANAAN

### 3.1. Populasi

Populasi adalah “keseluruhan gejala atau satuan yang ingin diteliti”[6]. Populasi dalam penelitian ini merupakan data customer aktif maupun tidak aktif serta penjualan selama tahun 2020.

### 3.2. Sampel Penelitian

Sampel merupakan “bagian dari populasi yang ingin diteliti. Oleh karena itu, sampel harus dilihat sebagai suatu pandangan pendugaan terhadap populasi dan bukan populasi itu sendiri”[6]. Penentuan sampel dalam penelitian ini dilakukan dengan teknik acak (stratified random sampling). Rumus dari penarikan teknik acak terlapis (stratified random sampling).

Tabel 1. Sampel Penelitian

Bulan	Jumlah Transaksi
-------	------------------

Januari	22
Februari	21
Maret	25
April	28
Mei	23
Juni	19
Juli	16
Agustus	28
September	27
Oktober	29
November	32
Desember	19
<b>Jumlah</b>	<b>289</b>

Sumber : PT DECO (2021) [2]

### 3.3. Metode Analisis Data

Dalam penelitian ini menggunakan model CRISP-DM (Cross-Industry Standard Proses for Data Mining), yang terdiri dari 6 tahap yaitu :

#### 1. Business/Research Understanding Phase

Salah satu perusahaan yang terkena imbasnya yaitu PT Deco yang bergerak di bidang penyediaan suku cadang kendaraan. Banyak pelanggan-pelanggan yang mengalami keterlambatan pembayaran dikarenakan pandemic ini. Oleh sebab itu diperlukan suatu cara untuk menentukan mana customer yang mampu melakukan pembayaran dengan tepat waktu dan mana yang tidak. Penulis menggunakan metode Algoritma KNN untuk mengklasifikasikan customer yang mendapatkan fasilitas credit dengan tepat

#### 2. Data Understanding Phase

Atribut-atribut yang menjadi parameter terlihat pada tabel 2. yaitu

Tabel 2. Atribut dan Nilai Katagori

No	Atribut	Nilai
1	Jenis Customer	Lama
		Baru
2	Bidang Kegiatan	Industri
		Konstruksi
		Lainnya
3	Klasifikasi Usaha	Besar
		Menengah
4	Tanpa/Dengan Broker	Tanpa
		Broker

Sumber : Pengolahan data (2021)

#### 3. Data Preparation Phase

Data yang digunakan dalam penelitian ini adalah sebanyak 289 transaksi setelah dilakukan preprocessing. Kemudian dari data sebanyak 289 data dibagi menjadi data training sebanyak 231 dan data testing diambil sebanyak 58 data, Seperti yang pada tabel 1, dengan rumus dibawah ini, sb :

- Rumus Data Training (Data Populasi \* 80%)  
=  $289 \times 80\% = 231$  data
- Rumus Data Testing ( Data Populasi \* 20%)  
=  $289 \times 20\% = 58$  data

#### 4. Modeling Phase

Tahap ini juga disebut tahap learning karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan keputusan. Pada penelitian ini, pembuatan model menggunakan algoritma KNN.

#### 5. Evaluation Phase

Pada tahap ini dilakukan pengujian model untuk mendapatkan informasi model yang akurat.

#### 6. Deployment Phase

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini juga diterapkan model ke proses penentuan customer kredit

### 4. PEMBAHASAN

Data yang sudah disiapkan kemudian diolah menggunakan model yang digunakan, dalam hal ini model yang digunakan adalah perhitungan menggunakan algoritma KNN. Untuk menghitung setiap jarak dari atribut yang ada, maka dilakukan pembobotan untuk setiap atribut yang ada.

Tabel 3. Pembobotan Atribut

Atribut	BA	Kelas	BSA
Jenis Customer	1	Lama	1,0
		Baru	0.5
Bidang kegiatan	2	Industri	3,0
		Konstruksi	1.5
		Lainnya	1
Klasifikasi Usaha	3	Besar	1,0
		Menengah	0.5
Broker	4	Melalui Broker	1

Sumber : Pengolahan data (2021)

Selanjutnya dilakukan penginputan bobot kedalam data training seperti tampilan pada tabel berikut.

Tabel 4. Pembobotan Data Training

No	Cust		Ind	Peng	Broker	Tepa
1	ADHI	1	3	1	1	tepat
2	MUS	0.5	1.5	1	1	tidak
3	UPA	0.5	3	0.5	0.5	tidak
4	HK	1	1.5	1	1	tepat
5	SAE	1	3	1	1	tepat
6	SURYA	0.5	3	0.5	0.5	tidak
7	CSE	1	3	1	1	tepat
8	BCA	0.5	0.5	1	1	tidak
9	CIPTA	0.5	1	1	0.5	tepat
10	BSJ	0.5	3	0.5	0.5	tidak

Sumber : Pengolahan data (2021)

Untuk pengujian dilakukan dengan menarik data testing dan menghitung jarak dari data training yang digunakan sesuai dengan rumus yang ada.

Tabel 5. Perhitungan Jarak dan Rank

No	Parameter	Jarak	Label	Rank
1	AG-ADHI	0,00	Tepat	1
2	AG-MUS	1,58	Tidak	8
3	AG-UPA	0,87	Tidak	4
4	AG-HK	1,50	Tepat	7
5	AG-SAE	0,00	Tepat	2
6	AG-SURYA	0,87	Tidak	6
7	AG-CSE	0,00	Tepat	3
8	AG-BCA	2,55	Tidak	10
9	AG-CIPTA	2,12	Tepat	9
10	AG-BSJ	0,87	Tidak	5

Sumber : Pengolahan data (2021)

Dari hasil diatas apabila ditentukan  $K=5$  maka 3 data menyatakan tepat dan 2 data menyatakan tidak tepat, maka dapat ditarik kesimpulan bahwa apabila customer tersebut membeli akan membayar dengan tepat waktu.

Dan sebagai evaluasi untuk menentukan k terbaik dilakukan dengan metode confusion matrix dan curva AUC yang tampak seperti tabel dan gambar dibawah ini.

Tabel 6. Percobaan K terbaik

K	Accuraccy	AUC
K1	84.86%+/-6.20	0.500
K3	83.55%+/-6.38	0.673
K5	84.86%+/-6.20	0.692
K7	83.99%+/-7.01	0.692
K9	82.68%+/-6.15	0.675
K11	80.54%+/-5.11	0.658

Sumber : Pengolahan data (2021)

Maka berdasarkan percobaan yang terangkum dalam tabel 6 maka k terbaik adalah k-5. Untuk hasil lengkap percobaan k-5 dapat dilihat pada gambar 1 dan gambar 2.

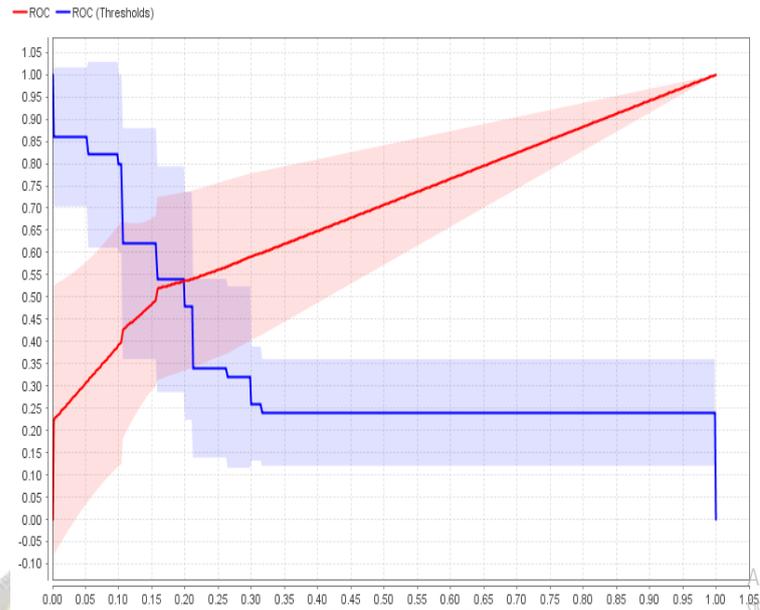
Selain itu dilakukan juga perhitungan menggunakan rapid miner dan menghasilkan nilai akurasi sebesar 84,86% Serta nilai AUC 0.692.

accuracy: 84.86% +/- 6.79% (mikro: 84.85%)

	true tepat	true tidak	class precision
pred. tepat	182	25	87.92%
pred. tidak	10	14	58.33%
class recall	94.79%	35.90%	

Gambar 1. Confusion Matrix (accuracy)

Sumber : Pengolahan data (2021)



Gambar 2. Curva AUC  
Sumber : Pengolahan data (2021)

Untuk klasifikasi nilai AUC dapat dibagi menjadi beberapa kelompok[3]

- a) 0.90-1.00 = klasifikasi sangat baik
- b) 0.80-0.90 = klasifikasi baik
- c) 0.70-0.80 = klasifikasi cukup
- d) 0.60-0.70 = klasifikasi buruk

## 5. KESIMPULAN

Dari hasil percobaan yang sudah dilakukan dengan Algoritma KNN mampu menghasilkan klasifikasi customer kredit dengan akurasi sebesar 84.86% dan AUC 0.692, sehingga dapat disimpulkan customer yang terdapat di data latih dapat menerima penentuan kredit.

## DAFTAR PUSTAKA

1. Aprilia, Dennis, Donny Aji Baskoro, Lia Ambarwati dan I Wayan Simri Wicaksana, 2013, *Belajar Data Mining dengan Rapid Miner*, Gramedia Pustaka Utama, Jakarta
2. Divisi Penjualan, 2021, *Data Penjualan 2020*, PT DECO.
3. Gorunescu, Florin, 2011, *Data mining :concepts, models, and techniques*, Springer, Verlag Berlin Heidelberg
4. Haditsah Annur dan Moh, Efendi Lasulika, 2019, *Klasifikasi Nasabah Kredit Koperasi Menggunakan Algoritma K-Nearest Neighbor*.ISSN : 2460-4801. Jurnal Informatika Upgris Vol.5, No.2 Desember 2019.
5. Kusri dan Emha Taufiq Luthfi, 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta
6. Prasetyo, Bambang, Lina Miftahul Jannah, 2011, *Metode Penelitian Kuantitatif: Teori dan Aplikasi*, PT Raja Grafindo Persada, Jakarta
7. Santoso, B, 2007, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis (1 ed.)*, Graha Ilmu, Yogyakarta