

## IMPLEMENTASI ANALISIS DATA KREDIT NASABAH MENGGUNAKAN METODE *K-NEAREST NEIGHBORS*

Fajar Ramadhan Akbar, Sentot Achmadi, Ali Mahmudi  
Teknik Informatika – ITN Malang  
fajarramaakbar23@gmail.com

### ABSTRAK

Dalam unit simpan pinjam pada koperasi unit desa memiliki sistem yang digunakan dalam menganalisis data nasabah dengan pencatatan pada buku, sehingga dalam mengklasifikasi peminjaman baru anggota menghabiskan banyak waktu. Dengan adanya permasalahan tersebut menimbulkan kendala yang dihadapi salah satunya sering terjadi pinjaman anggota yang jatuh tempo dalam pelunasan yang memiliki kasus sama dengan anggota lain.

Pada penelitian ini bagaimana merancang sebuah sistem analisis data kredit nasabah menggunakan metode *k-nearest neighbor* untuk mengklasifikasi data peminjaman baru. Kriteria yang digunakan untuk perhitungan yaitu kriteria pokok pinjaman dan lama angsuran.

Berdasarkan hasil perhitungan kinerja sistem yang telah dilakukan dalam mengklasifikasi peminjaman baru berdasarkan data lama koperasi diperoleh nilai kinerja sistem sebesar 74 %. Dari hasil pengujian performa dengan *confusion matrix* terhadap algoritma *K-Nearest Neighbor* yang menggunakan dua metode jarak yaitu metode *Euclidian Distance* diperoleh akurasi terbesar 79%, nilai *presicion* terbesar yaitu 81% dan *recall* terbesar 92% sedangkan metode *Cosine Similarity* dengan nilai akurasi sebesar 76%, *presicion* sebesar 76% dan *recall* sebesar 90%.

**Kata kunci :** Koperasi Unit Desa, Klasifikasi, KNN, *Euclidian Distance*, *Cosine Similarity*

### 1. PENDAHULUAN

#### 1.1 Latar Belakang

Perekonomian di Indonesia terdiri dari 3 sektor yaitu pemerintah, swasta, dan koperasi. Dari ketiga sektor tersebut, koperasi menduduki tempat yang sentral dalam usaha pembangunan ekonomi, yaitu mempunyai peran penting dalam meningkatkan taraf hidup yang lebih baik untuk mewujudkan masyarakat yang adil dan makmur serta mengembangkan demokrasi ekonomi Indonesia [1].

Dalam unit simpan pinjam pada koperasi unit desa memiliki sistem yang digunakan dalam menganalisis data nasabah dilakukan dengan tradisional yaitu pencatatan pada buku sehingga proses mengklasifikasi peminjaman baru anggota menghabiskan banyak waktu. Dengan adanya permasalahan tersebut menimbulkan kendala yang dihadapi salah satunya sering terjadi pinjaman anggota yang jatuh tempo dalam pelunasan yang memiliki kasus sama dengan anggota lain,

Oleh karena itu, penulis ingin merancang sebuah yang dapat mempermudah dalam klasifikasi data peminjaman anggota baru yang termasuk dalam kategori beresiko tinggi atau rendah. Pada penelitian ini dirancang sebuah aplikasi analisis data kredit nasabah menggunakan metode *k-nearest neighbor* untuk mengklasifikasi data peminjaman baru. Kriteria yang digunakan untuk perhitungan yaitu kriteria pokok pinjaman dan lama angsuran.

Penelitian ini melakukan klasifikasi data yang hasil akhirnya berupa pemberian kategori terhadap data baru yang belum diketahui kategorinya berdasarkan kedekatan dengan data lama. *K-Nearest Neighbor (KNN)* adalah metode yang termasuk kelompok dalam pengklasifikasian data yang sederhana dan mudah untuk pengimplementasian, lebih efektif didata *training* yang lebih besar dan dapat mengklasifikasi data dengan akurat. Sehingga metode KNN sesuai dengan data yang digunakan dalam penelitian ini.

#### 1.2 Rumusan Masalah

Berdasarkan permasalahan yang teridentifikasi di atas, maka dapat dirumuskan beberapa rumusan masalah sebagai berikut:

1. Bagaimana merancang sistem analisis data kredit nasabah berbasis web.
2. Bagaimana mengaplikasikan metode *K-Nearest Neighbor (KNN)* pada sistem.

#### 1.3 Tujuan

Terdapat beberapa tujuan dari pembuatan aplikasi ini sebagai berikut:

1. Membuat sebuah sistem analisis data kredit nasabah berbasis web.
2. Bagaimana mengaplikasikan metode *K-Nearest Neighbor (KNN)* pada sistem.

#### 1.4 Batasan Masalah

Berdasarkan pembuatan aplikasi ini terdapat beberapa batasan dalam pembuatan yaitu sebagai berikut:

1. Penelitian ini dilaksanakan di Koperasi Unit Desa Karangploso.

2. Data yang digunakan adalah data internal koperasi unit simpan pinjam pada koperasi unit desa Karangploso dari Januari 2018 – September 2019.
3. Data set yang digunakan berjumlah 340 data. 330 data untuk data *training* dan 10 data untuk data testing.
4. Bahasa pemrograman yang digunakan dalam proses pembuatan yaitu PHP, *JavaScript*.
5. Metode data *mining* yang digunakan adalah algoritma *K-Nearest Neighbor* dengan *Euclidian Distance* dan algoritma *K-Nearest Neighbor* dengan *Cosine Similarity*.

## 2. TINJAUAN PUSTAKA

### 2.1 Penelitian Terdahulu

Leidiyana, H. pada tahun 2013 melakukan penelitian dengan judul “Penerapan Algoritma *K-Nearest Neighbor* Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor”. Pada penelitiannya peneliti membahas algoritma *k-nearest neighbour (KNN)* yang diterapkan untuk mengklasifikasi resiko kredit kepemilikan kendaraan bermotor. Berdasarkan hasil *testing* pengukuran performa algoritma dengan menggunakan metode *cross validation*, *confusion matrix* dan *kurva ROC* menghasilkan akurasi dan nilai AUC berturut-turut 81,465 dan 0,984, maka dengan nilai AUC tersebut masuk ke dalam kategori sangat baik (*excellent*) [2].

Hendro Marcos pada tahun 2014 dengan penelitian yang berjudul “Implementasi Data *Mining* untuk Klasifikasi Nasabah Kredit Bank “X” Menggunakan *Classification Rule*”, pada penelitian ini membahas tentang penentuan diterima atau ditolak permohonan berdasarkan dari data-data yang sudah. Metode yang diterapkan yaitu metode klasifikasi data *mining* yaitu *classification rule* dan *decision tree*. Berdasarkan hasil akurasi dan model yang terbentuk dari metode-metode klasifikasi maka didapat algoritma C4.5 atau J48 mendapat akurasi terbesar yaitu 89,18% [3].

Ditahun berikutnya Amin, R K, Dkk. pada melakukan penelitian yang berjudul “Implementasi Klasifikasi *Decision Tree* Dengan Algoritma C4.5 Dalam Pengambilan Keputusan Permohonan Kredit Oleh Debitur (Studi Kasus: Bank Pasar Daerah Istimewa Yogyakarta)”. Pada penelitiannya membahas kinerja algoritma C4.5 pada identifikasi kelayakan kredit debitur. Diperoleh nilai *precision* terbesar 78,08% dengan partisi data 90%:10% dan nilai *recall* terbesar adalah 96,4% dengan partisi data sebesar 80%:20% yang membuktikan bahwa algoritma C4.5 memiliki tingkat akurasi yang tinggi dan lebih baik dari ID3 [4].

Penelitian yang telah dilakukan oleh Menarianti, I. pada tahun 2015 dengan judul “Klasifikasi Data *Mining* Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi”. Metode yang digunakan yaitu *logistic regression*, *discriminant analys*, *k-nearest neighbor*, *naive bayes*, *decision tree*, *neural network* dan *support vector machine*. Berdasarkan hasil dari analisis komparasi dengan menggunakan *cross validation*, *confusion matrix*, *ROC curve* dan *T-Test* pada beberapa algoritma klasifikasi data *mining* dapat disimpulkan bahwa algoritma yang paling akurat nilai akurasi adalah algoritma *logistic regression*. Karena memiliki nilai akurasi tertinggi yaitu 87,41% dengan uji *T-test* paling dominan terhadap algoritma lainnya [5].

Fahrudin, F. Dkk. pada tahun 2018 melakukan penelitian berjudul “Sistem Pendukung Keputusan Pemberian Kredit Modal Kerja Menggunakan Metode *Simple Additive Weighting* Pada Bank BPR Kabupaten Cirebon”. Pada penelitiannya membahas tentang pendukung keputusan pemberian kredit nasabah yang cepat dalam penentuannya dan diperoleh hasil bahwa sistem pendukung keputusan dengan metode *Analytical Hierarchy Process* dapat membantu penyeleksian pinjaman kredit sebesar 80% dan telah sesuai prosedur yang diharapkan [6].

### 2.2 Metode *K-Nearest Neighbor (KNN)*

*K-Nearest Neighbor (KNN)* termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok k objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing* [7].

Langkah awal dalam perhitungan dengan metode *K-Nearest Neighbor (KNN)* ialah, terlebih dahulu membagi data menjadi data latih dan data uji. Kemudian dilakukan pencarian nilai jarak dengan perhitungan menggunakan metode jarak, salah satu metodenya yaitu metode *Euclidean Distance* dan *Cosine Similarity*. Dengan mengetahui nilai jarak tersebut maka akan dilakukan tahapan pengklasifikasian dengan metode *K-Nearest Neighbor (KNN)*.

### 2.3 Metode *Euclidian Distance*

*Euclidian Distance* adalah perhitungan jarak dari 2 buah titik dalam *Euclidian Space*. *Euclidian space* pertama kali diperkenalkan oleh Euclid, seorang matematikawan dari Yunani sekitar tahun 300 B.C.E. untuk mempelajari hubungan antara sudut dan jarak. *Euclidian* ini berkaitan dengan Teorema *Phytagoras* dan biasanya diterapkan pada 1, 2 dan 3 dimensi.

Persamaan 1 merupakan rumus untuk menghitung jarak pada metode *K-Nearest Neighbor (KNN)* dengan *Euclidian Distance* :

$$d(P, Q) = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2} \quad (1)$$

Dimana

- D(P,Q) : jarak euclidian
- n : jumlah data latih
- P : inputan data ke -1 dari data uji
- Q : inputan data ke -1 dari data latih

### 2.4 Metode Cosine Similarity

Cosine Similarity adalah ukuran kesamaan yang lebih umum digunakan dalam *information retrieval* dan merupakan ukuran sudut antara dokumen (titik (ax,bx)) dan (titik (ay,by)) [8]. Dengan jarak antara 0 sampai 1, jika hasil mendekati 1 maka jarak perhitungan berdekatan begitu juga sebaliknya.

Persamaan 2 merupakan rumus untuk menghitung jarak pada metode *K-Nearest Neighbor* (KNN) dengan *Cosine Similarity* :

$$\cos(\theta_{QD}) = \frac{\sum_{i=1}^n Q_i D_i}{\sqrt{\sum_{i=1}^n (Q_i)^2} \cdot \sqrt{\sum_{i=1}^n (D_i)^2}} \quad (2)$$

Dimana

- cos(θ<sub>QD</sub>) : kemiripan Q terhadap dokumen D
- Q : data uji
- d : data latih
- n : jumlah data latih

## 3. METODE PENELITIAN

### 3.1 Desain Arsitektur Sistem

Desain arsitektur sistem pada rancangan sistem terdiri dari input, proses, output seperti pada Tabel 1 sebagai berikut :

Tabel 1 Desain Arsitektur Sistem

Input	Proses	Output
1. Data anggota baru 2. Variabel data : • Nama • Alamat • Pokok Pinjaman • Tanggal • Lama Angsuran • Resiko Bermasalah 3. Jumlah nilai k atau nilai ketetangaan.	1. Melakukan perhitungan jarak antara data latih dan data uji. 2. Mengurutkan data berdasarkan jarak 3. Menentukan data yang berketetangaan berdasarkan nilai K. 4. Label mayoritas berdasarkan nilai K.	1. Hasil klasifikasi perhitungan merupakan resiko bermasalah yang terjadi pada calon peminjam berdasarkan kedekatan dengan data lama.

Desain arsitektur sistem Tabel 1 menunjukkan bahwa data masukan digunakan untuk melakukan perhitungan. Inputan berupa data anggota koperasi dengan kriteria seperti pada kolom input point 2 dan menentukan jumlah nilai ketetangaan. Data yang telah diinputkan akan diproses dengan algoritma KNN seperti pada kolom proses. Hasil perhitungan yang diperoleh merupakan output dari sistem yang mengklasifikasi data baru.

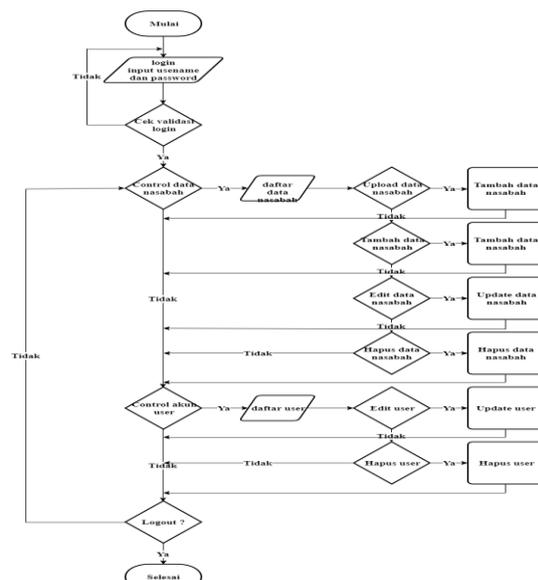
### 3.2 Flowchart Perhitungan KNN



Gambar 1 Flowchart Perhitungan KNN

Gambar 1 menunjukkan alur kerja algoritma KNN. Sistem dimulai dengan menginputkan data baru dan nilai ketetangaan untuk menentukan jarak tetangga. Proses mendapatkan nilai jarak antara data baru dengan data *training* menggunakan metode *Euclidian Distance* dan metode *Cosine Similary*, setelah mendapatkan nilai jarak, nilai tersebut diurutkan berdasarkan masing-masing metode jarak. Dari hasil pengurutan berdasarkan jarak dihasilkan resiko bermasalah untuk data baru dengan label mayoritas pada nilai K.

### 3.3 Flowchart Admin

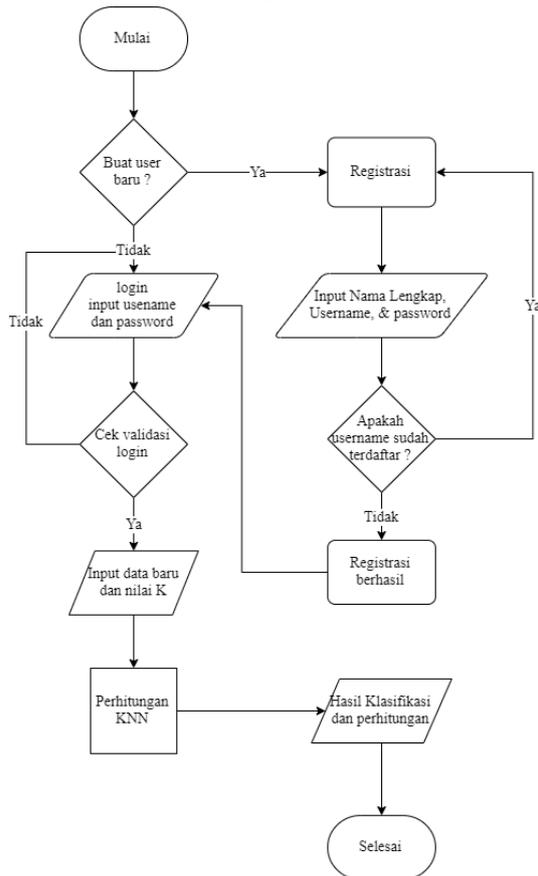


Gambar 2 Flowchart Admin

Gambar 2 merupakan *flowchart* admin yang dimulai dengan menginputkan *username* dan *password*, cek kevalidan apakah inputan login sesuai dengan di database. Admin dapat mengontrol data nasabah dengan mengubah, menambah, menghapus data-data nasabah.

Admin dapat mengontrol pengguna aplikasi dengan melihat daftar user. Admin hanya bisa mengubah dan menghapus akun user, sedangkan menambahkan akun user dapat dilakukan pada halaman registrasi diawal login.

### 3.4 Flowchart Petugas

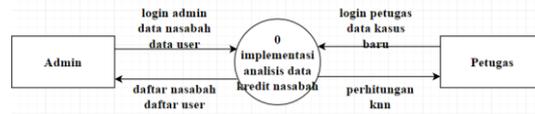


Gambar 3 Flowchart Petugas

Gambar 3 *Flowchart* petugas sistem ini dimulai dengan halaman login. Jika belum memiliki akun *user* maka akan di alihkan ke halaman registrasi akun dengan pemberian hak akses sebagai petugas. Petugas dapat menginputkan data baru untuk mendapatkan hasil klasifikasi dan bisa melihat hasil perhitungan dari metode yang digunakan.

### 3.5 DFD Level 0

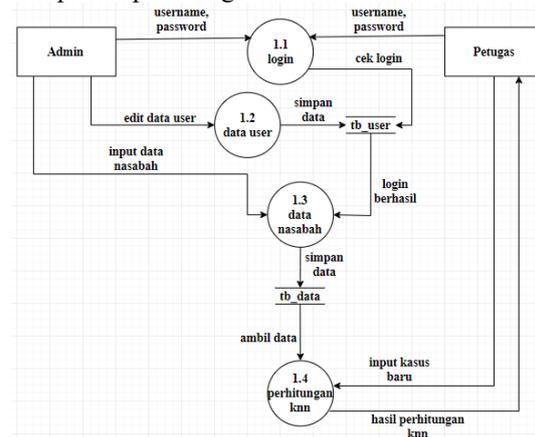
DFD level 0 menggambarkan terdapat 2 pelaku yang terdapat di dalam system yaitu admin dan petugas. Admin bertugas untuk mengelola data nasabah dan data user, dan petugas dapat mengakses sistem perhitungan knn.



Gambar 4 DFD level 0

### 3.6 DFD Level 1

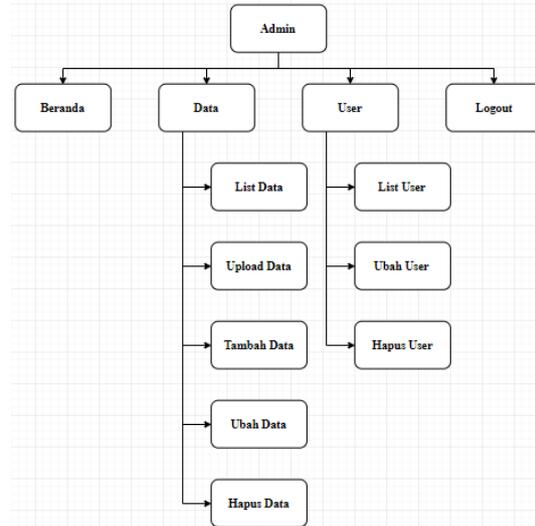
Diagram level 1 di atas menggambarkan bahwa sistem ini terdiri dari proses registrasi, proses *login*, proses data *user*, proses data nasabah dan proses perhitungan knn.



Gambar 5 DFD level 1

### 3.7 Struktur Menu Sistem

#### 1. Struktur Menu Admin

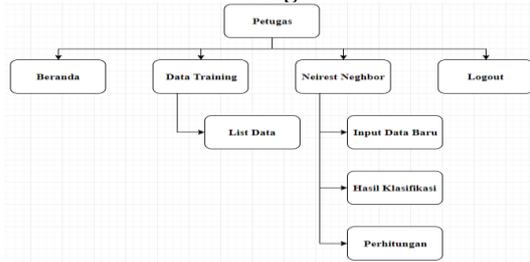


Gambar 6 Tampilan Struktur Menu Admin

Gambar 6 menunjukkan struktur menu-menu utama pada admin seperti beranda yang menampilkan diagram lingkaran jumlah anggota yang per resiko bermasalah. Menu data menampilkan seluruh data nasabah dengan berbagai macam resiko dan memiliki sub menu yaitu upload data, tambah data ubah data dan hapus data.

Pada menu user menampilkan seluruh daftar pengguna sistem yang dapat diedit dan dihapus oleh admin.

## 2. Stuktur Menu Petugas



Gambar 7 Tampilan Struktur Menu Petugas

Gambar 7 menunjukkan menu-menu yang ada pada petugas. Beranda menginformasikan dalam bentuk diagram lingkaran tentang jumlah data anggota berdasarkan resiko bermasalah. Data *training* merupakan hanya menampilkan data-data anggota. *Nearest neighbor* merupakan menu proses pengklasifikasi data baru.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Perhitungan Metode K-Nearest Neighbor (KNN) dengan Eucliden Distance

Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek, berdasarkan k buah data latih yang jaraknya paling dekat dengan objek tersebut. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data latih, dan nilai k harus ganjil dan lebih dari satu. Dekat atau jauhnya jarak data latih yang paling dekat dengan objek yang akan diklasifikasi dapat dihitung dengan menggunakan metode jarak. Berikut perhitungan metode *K-Nearest Neighbor* (KNN) dengan *Eucliden Distance* pada studi kasus Koperasi Unit Desa Karangploso dalam mengklasifikasi resiko bermasalah pada peminjaman anggota koperasi.

Data *training* yang digunakan pada kasus ini berjumlah 330 data *training*. Adapun data *training* yang digunakan dapat dilihat pada Tabel 2. Untuk data baru yang ingin dilakukan perhitungan dapat dilihat pada Tabel 4.

Tabel 2 Data Training

No	Nama	Alamat	Pinjam (ribuan)	Tanggal	Angsuran	Resiko
1	Puji Samsul	Mojosari	1000	30/03/2017	6	Tinggi
2	Raki / Sulaiman	Supiturang	3300	06/07/2018	6	Rendah
3	P Patah	Mangg esari	1000	10/02/2018	3	Tinggi
4	H Moh Yahya	Takera n	1000	10/05/2018	12	Rendah
5	A Sodikin	Supiturang	8000	15/01/2019	3	Tinggi
...	...	...	...	...	...	...
325	Nasir / Nuridin	Bocek	1000	03/08/2018	10	Rendah

Tabel 3 Data Training (Lanjutan)

No	Nama	Alamat	Pinjam (ribuan)	Tanggal	Angsuran	Resiko
326	Jaelani	Bocek	2000	12/09/2018	4	Rendah
327	Muhammad Asyikin	Bocek	5000	09/08/2018	3	Tinggi
328	Sugiarto / Nuriyadi	Karangan	3000	12/04/2018	10	Rendah
329	Suprianto	Karangan	1500	12/06/2017	10	Tinggi
330	Sutikno	Caru	2000	30/08/2018	10	Rendah

Tabel 4 Data Testing

No	Nama	Alamat	Pinjam (ribuan)	Tanggal	Angsuran	Resiko
1	Panawi / Sugianto	Boro	2000	12/07/2018	10	Rendah

**Langkah 1.** Menentukan nilai *parameter* K. Nilai K ditentukan sendiri oleh petugas dengan syarat nilai k berupa angka ganjil dan lebih dari satu. Dalam perhitungan ini petugas menggunakan nilai k = 3.

**Langkah 2.** Menghitung jarak antara data *testing* (Tabel 4) dengan data *training* (Tabel 2) dimana X merupakan kriteria pokok pinjaman (ribuan) dan Y merupakan kriteria lama Angsuran.

Tabel 5 Perhitungan Jarak

No	X	Y	Eucliden Distance (2000, 10)
1	100000	6	$\sqrt{(100000 - 2000)^2 + (6 - 10)^2} = 98000,0008$
2	33000	6	$\sqrt{(33000 - 2000)^2 + (6 - 10)^2} = 31000,00026$
3	10000	3	$\sqrt{(10000 - 2000)^2 + (3 - 10)^2} = 8000,003062$
4	10000	12	$\sqrt{(10000 - 2000)^2 + (12 - 10)^2} = 8000,00025$
5	80000	3	$\sqrt{(80000 - 2000)^2 + (3 - 10)^2} = 78000,00031$
...	...	...	...
325	1000	10	$\sqrt{(1000 - 2000)^2 + (10 - 10)^2} = 1000$
326	2000	4	$\sqrt{(2000 - 2000)^2 + (4 - 10)^2} = 6$
327	5000	3	$\sqrt{(5000 - 2000)^2 + (3 - 10)^2} = 3000,008167$
328	3000	10	$\sqrt{(3000 - 2000)^2 + (10 - 10)^2} = 1000$
329	1500	10	$\sqrt{(1500 - 2000)^2 + (10 - 10)^2} = 500$
330	2000	10	$\sqrt{(2000 - 2000)^2 + (10 - 10)^2} = 0$

**Langkah 3.** Mengurutkan jarak yang terbentuk berdasarkan hasil perhitungan data *testing* dengan

data *training* dari kecil ke besar. Seperti Tabel 6 berikut:

**Tabel 6** Pengurutan Jarak Data

No	X	Y	Euclidian Distance (2000, 10)	Urutan Jarak
1	10000 0	6	98000,0008	326
2	33000	6	31000,00026	309
3	10000	3	8000,003062	252
4	10000	1 2	8000,00025	231
5	80000	3	78000,00031	325
...	...	...	...	...
32 5	1000	1 0	1000	70
32 6	2000	4	6	29
32 7	5000	3	3000,008167	187
32 8	3000	1 0	1000	71
32 9	1500	1 0	500	42
33 0	2000	1 0	0	17

**Langkah 4.** Menentukan jarak terdekat sampai urutan K dan mayoritas label. Hasil ditunjukkan pada Tabel 7:

**Tabel 7** Jarak terdekat dan mayoritas label

No	X	Y	Euclidian Distance (2000,10)	Urutan	Resiko Bermasalah	Termasuk 3-NN
9	20 00	1 0	0	1	Rendah	Ya
2 6	20 00	1 0	0	2	Rendah	Ya
4 3	20 00	1 0	0	3	Tinggi	Ya

Dapat disimpulkan bahwa data *testing* (2000,10) termasuk dalam resiko bermasalah **Rendah**.

**4.2 Perhitungan Metode K-Nearest Neighbor (KNN) dengan Cosine Similarity**

Berikut perhitungan metode *K-Nearest Neighbor (KNN)* dengan *Cosine Similarity* pada studi kasus Koperasi Unit Desa Karangploso dalam mengklasifikasi resiko bermasalah pada peminjaman anggota koperasi.

Data *training* yang digunakan pada kasus ini berjumlah 330 data *training*. Adapun data *training* yang digunakan dapat dilihat pada Tabel 2. Untuk data baru yang ingin dilakukan perhitungan dapat dilihat pada Tabel 4.

**Langkah 1.** Menentukan *parameter K*. Nilai K telah didefinisikan pada awal inputan sama dengan nilai K pada metode *Euclidian Distance*.

**Langkah 2.** Menentukan jarak antara data *testing* (Tabel 2) dengan data *training* (Tabel 4). Dimana X merupakan kriteria pokok pinjaman (ribuan) dan Y merupakan lama angsuran. Berikut hasil perhitungan jarak pada Tabel 8:

**Tabel 8** Perhitungan Jarak

No	X	Y	Cosine Similarity (2000, 10)
1	1000 00	6	$\frac{(100000 * 2000) + (6 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{100000^2 + 6^2}}$ = 0.999987798
2	3300 0	6	$\frac{(33000 * 2000) + (6 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{33000^2 + 6^2}}$ = 0.999988393
3	1000 0	3	$\frac{(10000 * 2000) + (3 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{10000^2 + 3^2}}$ = 0.999988955
4	1000 0	1 2	$\frac{(10000 * 2000) + (12 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{10000^2 + 12^2}}$ = 0.99999278
5	8000 0	3	$\frac{(80000 * 2000) + (3 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{80000^2 + 3^2}}$ = 0.999987687
...	...	...	...
32 5	1000	1 0	$\frac{(1000 * 2000) + (10 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{1000^2 + 10^2}}$ = 0.999987501
32 6	2000	4	$\frac{(2000 * 2000) + (4 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{2000^2 + 4^2}}$ = 0.9999955
32 7	5000	3	$\frac{(5000 * 2000) + (3 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{5000^2 + 3^2}}$ = 0.99999032
32 8	3000	1 0	$\frac{(3000 * 2000) + (10 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{3000^2 + 10^2}}$ = 0.999998611
32 9	1500	1 0	$\frac{(1500 * 2000) + (10 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{1500^2 + 10^2}}$ = 0.999998611
33 0	2000	1 0	$\frac{(2000 * 2000) + (10 * 10)}{\sqrt{2000^2 + 10^2} + \sqrt{100000^2 + 10^2}}$ = 1

**Langkah 3.** Mengurutkan jarak yang terbentuk berdasarkan hasil perhitungan data *testing* dengan data *training* dari besar ke kecil. Berikut urutan jarak pada Tabel 9:

**Tabel 9** Pengurutan Jarak Data

No	X	Y	Cosine Similarity (2000, 10)	Urutan Jarak
1	10000 0	6	0.999987798	301
2	33000	6	0.999988393	284
3	10000	3	0.999988955	255
4	10000	1 2	0.99999278	134
5	80000	3	0.999987687	306
...	...	...	...	...
32 5	1000	1 0	0.999987501	319
32 6	2000	4	0.9999955	76
32 7	5000	3	0.99999032	227
32 8	3000	1 0	0.999998611	50
32 9	1500	1 0	0.999998611	41
33 0	2000	1 0	1	1

**Langkah 4.** Menentukan jarak terdekat sampai urutan K dan mayoritas label. Hasil ditunjukkan pada Tabel 10:

**Tabel 10** Jarak terdekat dan mayoritas label

No	X	Y	Cosine Similarity (2000,10)	Urutan	Resiko	Termasuk 3-NN
9	2000	10	1	1	Rendah	Ya
26	2000	10	1	2	Rendah	Ya
43	2000	10	1	3	Rendah	Ya

Dapat disimpulkan bahwa data *testing* (2000,10) termasuk dalam resiko bermasalah **Rendah**.

### 4.3 Pengujian Algoritma

#### 1. Confusion Matrix

Tabel 11 adalah tabel *confusion matrix* yang dihasilkan menggunakan algoritma KNN dengan pengukuran jarak *Eucliden Distance*. Diketahui 330 data, 211 data rendah diklasifikasi sebagai rendah, 19 data rendah diklasifikasi sebagai tinggi, 51 data tinggi diklasifikasi sebagai rendah, 49 data tinggi diklasifikasi tinggi. Berikut model *confusion matrix* :

**Tabel 11** Model *Confusion Matrix* Metode *Eucliden Distance*

Klasifikasi yang benar	Diklasifikasi sebagai	
	Rendah	Tinggi
Rendah	211	19
Tinggi	51	49

$$Akurasi = \frac{211 + 49}{211 + 49 + 51 + 19} \times 100\% = 79\%$$

$$Presicion = \frac{211}{211 + 51} \times 100\% = 81\%$$

$$recall = \frac{211}{211 + 19} \times 100\% = 92\%$$

Diperoleh perhitungan nilai akurasi sebesar 79%, *presicion* sebesar 81% dan *recall* sebesar 92%.

Tabel 12 adalah tabel *confusion matrix* yang dihasilkan menggunakan algoritma kNN dengan pengukuran jarak *Cosine Similarity*. Diketahui 330 data, 207 data rendah diklasifikasi sebagai rendah, 23 data rendah diklasifikasi sebagai tinggi, 56 data tinggi diklasifikasi sebagai rendah, 44 data tinggi diklasifikasi tinggi. Berikut model *confusion matrix* :

**Tabel 12** Mode *Confusion Matrix* Metode *Cosine Similarity*

Klasifikasi yang benar	Diklasifikasi sebagai	
	Rendah	Tinggi
Rendah	207	23
Tinggi	56	44

$$Akurasi = \frac{207 + 44}{211 + 44 + 56 + 23} \times 100\% = 76\%$$

$$Presicion = \frac{207}{207 + 56} \times 100\% = 79\%$$

$$recall = \frac{207}{207 + 23} \times 100\% = 90\%$$

Diperoleh perhitungan nilai akurasi sebesar 76%, *presicion* sebesar 79% dan *recall* sebesar 90%. Berdasarkan hasil model *confusion matrix* dari kedua metode dapat disimpulkan bahwa perhitungan jarak metode *eucliden distance* pada metode KNN lebih baik dari metode *cosine similarity*.

#### 2. Pengujian sistem

Pengujian sistem dilakukan dengan perbandingan hasil perhitungan manual dan hasil perhitungan dengan sistem yang telah dirancang. Dengan menggunakan 10 data baru. Hasil ditunjukkan pada Tabel 13. Untuk mendapatkan keterangan dilakukan pencocokan antara resiko bermasalah dari hasil perhitungan manual dengan hasil klasifikasi sistem. Apakah resiko bermasalah data baru sama dengan klasifikasi *eucliden* dan apakah resiko bermasalah data baru sama dengan klasifikasi *cosine similarity*.

**Tabel 13** Pengujian Sistem

No	Nama	Pokok Pinjaman (ribuan)	Lama Angsur	Eucliden Distance		Ket	Cosine Similarity		Ket
				manual	program		manual	program	
1	Panawi / Sugianto	2000	10	Rendah	Rendah	T	Rendah	Rendah	T
2	Solihah Umi	5000	6	Rendah	Rendah	T	Rendah	Rendah	T
3	Supakiyah / Puji S	50000	6	Tinggi	Tinggi	T	Tinggi	Tinggi	T
4	Yofi / Miftakul	40000	12	Tinggi	Tinggi	T	Rendah	Rendah	T
5	Malik K	9000	6	Rendah	Rendah	T	Tinggi	Tinggi	T
6	Puji	4500	6	Rendah	Rendah	T	Tinggi	Tinggi	T

Tabel 14 Pengujian Sistem (Lanjutan)

7	Agus Setiawan / Untung	5500	6	Rendah	Rendah	T	Rendah	Rendah	T
8	Lasari / Mistari	6000	6	Rendah	Rendah	T	Tinggi	Tinggi	T
9	P Sareh / Tikno	8000	6	Rendah	Rendah	T	Rendah	Rendah	T
10	Rakit / Sulaiman	25000	6	Rendah	Rendah	T	Rendah	Rendah	T

Keterangan :

T = *True*. Terjadi apabila hasil perhitungan manual sama dengan hasil sistem.

F = *False*. Terjadi apabila hasil perhitungan manual berbeda dengan hasil sistem.

Pengujian sistem terdapat keterangan yang bernilai F. Pada keterangan yang bernilai F disebabkan oleh mayoritas resiko bermasalah setelah pengurutan jarak, jika terdapat nilai jarak terkecil yang sama nilainya maka pengurutan data dilakukan berdasarkan urutan data atau id data terkecil. Sehingga menghasilkan kategori sesuai dengan urutan tersebut.

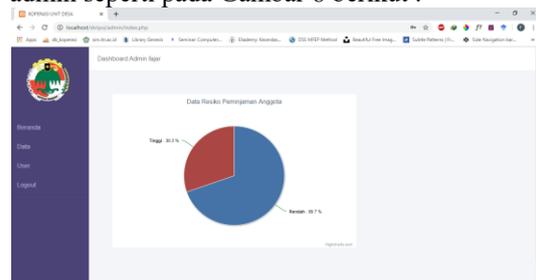
Berdasarkan pengujian sistem yang telah dilakukan dengan 10 data baru dengan nilai k=3, 10 diklasifikasi dengan benar. Maka diperoleh :

$$\begin{aligned}
 & \text{Pengujian sistem} \\
 & = \frac{\text{banyak hasil pengujian benar}}{\text{banyak data training}} \times 100\% \\
 & = \frac{10}{10} \times 100\% = 100\%
 \end{aligned}$$

Hasil yang ditunjukkan dari perhitungan pengujian sistem baik. Maka, sistem yang telah dirancang sesuai dengan perhitungan metode secara manual.

#### 4.4 Tampilan Halaman Utama Admin

Halaman utama admin merupakan tampilan halaman yang tampil setelah melakukan proses *login* dengan hak akses admin atau level 1. Pada menu halaman admin akan menginformasikan data resiko peminjaman anggota dan menu-menu lain yang berfungsi untuk mengelola data peminjaman anggota dan user berupa menambahkan data, mengubah data dan menghapus data. Tampilan halaman utama admin seperti pada Gambar 8 berikut :



Gambar 8 Tampilan Halaman Utama Admin

#### 4.5 Tampilan Halaman Data Peminjaman Anggota

Halaman data peminjaman anggota merupakan halaman yang menampilkan seluruh data peminjaman anggota. Tampilan halaman data peminjaman anggota seperti pada Gambar 9 berikut :

No	Nama	Alamat	Jumlah Pinjaman	Tanggal Pinjaman	Status Pinjaman	Aksi
1	Peg Sareh	Bandung	10000	01/02/2019	6	Tinggi
2	Lasari	Bandung	10000	01/02/2019	6	Rendah
3	P Sareh	Bandung	10000	21/02/2019	2	Tinggi
4	Agus Setiawan	Bandung	10000	01/02/2019	12	Rendah
5	Lasari	Bandung	10000	21/02/2019	3	Tinggi
6	Lasari	Bandung	10000	01/02/2019	6	Tinggi
7	Lasari	Bandung	10000	01/02/2019	3	Tinggi
8	Lasari	Bandung	10000	21/02/2019	4	Tinggi
9	Peg Sareh	Bandung	20000	01/02/2019	10	Rendah
10	Lasari	Bandung	10000	01/02/2019	10	Rendah
11	Lasari	Bandung	10000	01/02/2019	6	Rendah
12	Lasari	Bandung	10000	01/02/2019	10	Rendah
13	Lasari	Bandung	10000	01/02/2019	6	Rendah
14	Lasari	Bandung	10000	01/02/2019	6	Rendah
15	Lasari	Bandung	10000	01/02/2019	12	Rendah
16	Lasari	Bandung	10000	01/02/2019	4	Rendah
17	Lasari	Bandung	10000	21/02/2019	10	Rendah

Gambar 9 Tampilan Halaman Data Peminjaman Anggota

#### 4.6 Tampilan Halaman User

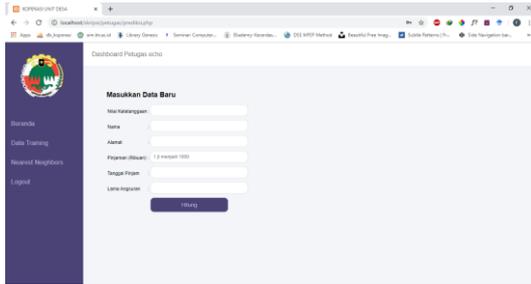
Halaman *user* merupakan halaman yang menampilkan seluruh data pengguna aplikasi baik sebagai admin atau petugas berupa nama lengkap, username dan level hak akses. Tampilan halaman user seperti pada Gambar 10 berikut :

No	Nama	Username	Level
1	Agus Setiawan	agus	Admin
2	Lasari	lasari	Admin

Gambar 10 Tampilan Halaman User

#### 4.7 Tampilan Halaman Utama Petugas

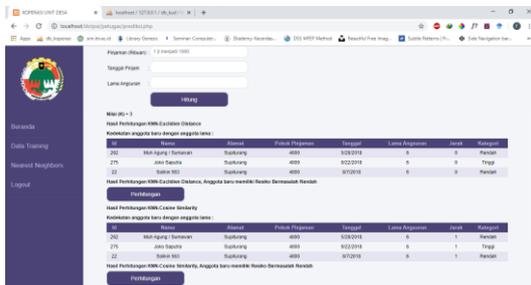
Halaman utama petugas merupakan tampilan halaman yang tampil setelah melakukan proses *login* dengan hak akses petugas atau level 2. Pada menu halaman petugas berfungsi untuk melakukan proses klasifikasi data peminjaman anggota baru. Menu-menu lain yang terdapat pada halaman petugas berupa beranda dan data *training*. Tampilan halaman utama petugas seperti pada Gambar 11 berikut :



Gambar 11 Tampilan Halaman Utama Petugas

#### 4.8 Tampilan Hasil Klasifikasi

Hasil klasifikasi merupakan hasil dari perhitungan metode yang tampil dalam satu halaman dengan halaman utama petugas setelah menekan tombol hitung pada halaman utama petugas. Tampilan hasil klasifikasi seperti pada Gambar 12 berikut :



Gambar 12 Tampilan Hasil Klasifikasi

#### 4.9 Pengujian Browser

Pada tahap pengujian *browser* menu *website* dilakukan dengan menggunakan 2 *browser* yaitu *Google Chrome 78.0.3904.108* dan *Mozilla Firefox 70.0.1*. Pengujian ini dilakukan untuk mengetahui fungsional dari fitur-fitur pada aplikasi berbasis *web*. Hasil pengujian ditunjukkan pada Tabel 15 :

Tabel 15 Pengujian Sistem

Hak Akses	Fungsi	GC	MF
	Halaman <i>Login</i>	√	√
	Tombol <i>Login</i>	√	√
	Halaman Registrasi	√	√
	Tombol Registrasi	√	√
Admin	Halaman Beranda	√	√
	Halaman Tampil Data	√	√
	Halaman Upload Data	√	√
	Tombol Upload Data	√	√
	Halaman Tambah Data	√	√
	Tombol Tambah Data	√	√
	Halaman Edit Data	√	√
	Tombol Edit Data	√	√
	Halaman Hapus Data	√	√
	Tombol Hapus Data	√	√
	Halaman Tampil User	√	√
	Halaman Edit User	√	√
	Tombol Edit User	√	√
	Halaman Hapus User	√	√
	Tombol Hapus User	√	√
	Logout	√	√

Tabel 16 Pengujian Sistem (lanjutan)

Hak Akses	Fungsi	GC	MF
Petugas	Halaman Beranda	√	√
	Halaman Tampil Data <i>training</i>	√	√
	Halaman Klasifikasi	√	√
	Tombol Hitung	√	√
	Tombol Perhitungan	√	√
	Halaman Perhitungan	√	√
	Logout	√	√

#### 4.1 Pengujian user

Pengujian *user* pada sistem analisis data kredit nasabah menggunakan metode *k-nearest neighbor (KNN)* ini dilakukan dengan memberikan pertanyaan kepada *user*. Pengujian *user* diajukan kepada 10 orang responden untuk memberikan penilaian terhadap aplikasi. Hasil pengujian *user* ditunjukkan pada Tabel 17

Tabel 4.17 Pengujian user

No.	Pertanyaan	Pilihan Jawaban		
		Setuju	Cukup	Tidak Setuju
1.	Apakah warna <i>background</i> dan warna tulisan pada sistem terlihat jelas pada pandangan mata anda ?	7	3	-
2.	Apakah warna <i>background</i> dengan warna semua elemen halaman (tombol, menu, ikon) sudah sesuai ?	7	3	-
3.	Apakah aplikasi membantu mempermudah proses analisa resiko kredit nasabah ?	5	5	-
4.	Apakah fungsi-fungsi yang tersedia berjalan sesuai fungsinya ?	10	-	-
5.	Apakah aplikasi dapat bermanfaat untuk kedepannya ?	9	1	-
	<b>Presentase</b>	76 %	24 %	

Dari hasil kuisisioner pengujian user dapat disimpulkan bahwa :

- 7 dari jumlah responden berpendapat bahwa warna *background* dan warna tulisan terlihat dengan jelas. 3 dari jumlah responden menganggap cukup jelas tulisan pada sistem.
- 7 dari jumlah responden berpendapat tampilan warna *background* selaras dengan elemen-elemen pada sistem. 3 pendapat cukup pada tampilan *website*.

3. 5 dari jumlah responden berpendapat bahwa sistem mempermudah proses analisa. 5 pendapat cukup membantu dalam analisa.
4. 10 dari jumlah responden berpendapat bahwa semua fungsi-fungsi yang ada pada sistem berjalan sesuai fungsinya.
5. dari jumlah responden berpendapat bahwa kedepannya aplikasi tersebut bermanfaat dan membantu pekerjaan. 1 perbendapat cukup bermanfaat untuk kedepannya.

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Setelah dilakukan pengujian aplikasi Implementasi Analisis Data Kredit Nasabah Menggunakan Metode *K-Nearest Neighbor*, maka penulis mendapatkan kesimpulan sebagai berikut :

1. Sistem analisis data kredit nasabah ini dapat diimplementasikan di Koperasi Unit Desa. Berdasarkan hasil pengujian sistem yang telah dilakukan, perhitungan sistem yang dilakukan dengan menggunakan *Microsoft excel* sesuai dengan perhitungan yang dilakukan pada sistem sebesar 100%.
2. Berdasarkan pengklasifikasian algoritma KNN dengan nilai ketetanggan 3 menggunakan metode *euclidien distance* diperoleh tingkat akurasi terbesar 79%, sedangkan pengklasifikasian algoritma KNN dengan nilai ketetanggan 3 menggunakan metode *cosine similarity* akurasi sebesar 76%. Terlihat bahwa metode *Euclidien Distance* lebih baik dari pada metode *Cosine Similarity* pada pengklasifikasian dengan algoritma KNN.
3. Sistem analisis data kredit nasabah dapat dijalankan pada *browser Google Chrome* dan *Mozilla Firefox* dengan baik.
4. Berdasarkan total skor dari responden sebesar 76 % responden setuju atas hasil dari sistem dan 24 % cukup setuju atas hasil dari pembuatan sistem. Bahwa sistem analisis data kredit nasabah menggunakan metode KNN baik dalam pengklasifikasian.

### 5.2 Saran

Berdasarkan penelitian yang telah dilakukan maka penulis dapat memberikan saran-

saran untuk pengembangan selanjutnya karena penelitian ini banyak kekurangan, sehingga untuk penyempurnaan dapat ditambahkan:

1. Sistem klasifikasi analisis data kredit nasabah menggunakan metode *k-nearest neighbor* dapat dilakukan perbandingan dengan metode klasifikasi yang lain.
2. Untuk pengembangan lebih lanjut data yang digunakan harus dilakukan pengujian kualitas data.

## DAFTAR PUSTAKA

- [1] Harsoyo, Y., 2006. *Ideologi koperasi menatap masa depan*. Pustaka Widyatama.
- [2] Leidiyana, H., 2013. Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan kendaraan bermotor. *Penelitian Ilmu Komputer Sistem Embedded dan Logic*, 1(1).
- [3] Marcos, H. and Hidayah, I., 2014. Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank 'X' Menggunakan Classification Rule. *Semin. Nas. Teknol. Inf. dan Multimed*, pp.1-7.
- [4] Amin, R.K., Indwiarti, I. and Sibaroni, Y., 2015. Implementasi Klasifikasi Decision Tree Dengan Algoritma C4. 5 Dalam Pengambilan Keputusan Permohonan Kredit Oleh Debitur (Studi Kasus: Bank Pasar Daerah Istimewa Yogyakarta). *eProceedings of Engineering*, 2(1).
- [5] Menarianti, I., 2015. Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*, 1(1/November).
- [6] Fahrudin, F. and Prayogi, P., 2018. Sistem Pendukung Keputusan Pemberian Kredit Modal Kerja Menggunakan Metode Simple Additive Weighting Pada Bank BPR Kabupaten Cirebon. *Jurnal ICT: Information Communication & Technology*, 17(1), pp.1-5.
- [7] Wu, X. and Kumar, V. eds., 2009. *The top ten algorithms in data mining*. CRC press.
- [8] Imbar, R.V., Adelia, A., Ayub, M. and Rehatta, A., 2014. Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, 10(1), pp.31-42.