

Implementasi *K-Nearest Neighbor* dengan Pemilihan Fitur pada Aplikasi Prediksi Kelayakan Pengajuan Pinjaman

Fani Fadillah Hermawan¹, Yuni Yamasari²

^{1,2} Jurusan Teknik Informatika, Prodi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

¹fani.18003@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak— Dalam kehidupan manusia tidak lepas dari yang dinamakan pemenuhan kebutuhan sehari-hari. Baik itu kebutuhan primer ataupun sekunder. Dalam pemenuhan kebutuhan tersebut pasti dibutuhkan dengan alat transaksi yang disebut dengan uang. Uang dapat didapatkan oleh manusia dari berbagai macam cara, mulai dari bekerja sebagai pencaharian utama atau primer hingga melakukan peminjaman uang sebagai pendapatan penunjang seorang manusia selain bekerja. Pada umumnya proses seleksi pemberian pinjaman, pihak kreditur akan melakukan proses seleksi dengan metode analisis kredit atau bisa disebut dengan prinsip 5C. Namun seiring dengan majunya komputerisasi proses tersebut bisa diprediksi dengan menggunakan salah satu metode dalam *data mining*. Metode tersebut adalah *K-Nearest Neighbor* dengan seleksi fitur menggunakan korelasi Pearson. Fitur dengan peringkat tiga teratas dengan hasil korelasi paling tinggi yang akan digunakan untuk proses prediksi. Penelitian ini bertujuan untuk membantu kreditur dalam memprediksi kelayakan pengajuan pinjaman seorang debitur. Selama uji coba dengan memakai fitur yang telah terseleksi dan nilai K sebesar 3. Pencatatan dilakukan saat hasil kinerja terendah hingga tertinggi algoritma *K-NN*. Kinerja terendah terjadi pada saat rasio perbandingan 5% data uji dan 95% data latih. Hal tersebut diindikasikan dengan hasil pengukuran *MAE* sebesar 0,26315789, *Precision* sebesar 76,92, *Recall* sebesar 83,3 dan *Accuracy* sebesar 73,68%. Kemudian kinerja tertinggi terjadi pada saat rasio perbandingan 20% data uji dan 80% data latih. Hal tersebut diindikasikan dengan hasil pengukuran *MAE* sebesar 0, *Precision* sebesar 100, *Recall* sebesar 100 dan *Accuracy* sebesar 100%.

Kata Kunci— Aplikasi, Prediksi, *K-Nearest Neighbor*, Kredit, Seleksi Fitur.

I. PENDAHULUAN

Dalam kehidupan, manusia tidak lepas dari yang dinamakan pemenuhan kebutuhan sehari-hari. Baik itu kebutuhan primer ataupun sekunder. Dalam pemenuhan kebutuhan tersebut pasti dibutuhkan dengan alat transaksi yang disebut dengan uang. Uang dapat didapatkan oleh manusia dari berbagai

macam cara, mulai dari bekerja sebagai pencaharian utama atau primer hingga meminjam uang sebagai pendapatan sekunder seorang manusia selain bekerja. Pihak yang mengajukan pinjaman disebut dengan debitur dan pihak yang memberikan pinjaman disebut dengan kreditur. Pinjaman adalah sebuah uang yang diberikan oleh pihak kreditur (bisa sebuah lembaga keuangan atau perusahaan) ke pihak penerima yang mana uang tersebut harus dibayarkan dalam waktu dan uang yang telah disepakati bersama sebelumnya [1]. Sebelum menerima pengajuan kredit dari debitur, pihak kreditur akan memproses secara selektif dan cermat. Hal tersebut untuk menghindari kredit macet yang mungkin bisa dialami oleh calon penerima kredit. Akibat yang akan terjadi jika dialami pihak kreditur apabila seorang debitur mengalami kredit macet adalah tingkat profitabilitas dari bunga kredit akan menurun [2]. Untuk terhindar dari risiko kredit yang dialami debitur. Pihak kreditur biasanya akan mengukur kelayakan atau kredibilitas seorang calon penerima kredit dengan menggunakan prinsip analisis kredit atau prinsip 5C [3]. Kelima unsur C pada analisis kredit tersebut antara lain *Character*, *Collateral*, *Capacity*, *Capital*, dan *Condition of economy* [4].

Pada penelitian ini penulis akan membuat sistem prediksi dengan memanfaatkan metode pada *data mining* yaitu klasifikasi. Tujuan dari klasifikasi adalah untuk memperkirakan sifat dari sebuah item dalam suatu kelas yang tersedia [5]. Sebuah label kelas dapat diprediksi juga dengan menggunakan metode klasifikasi [6]. Jenis-jenis klasifikasi antara lain *Bayesian Belief Networks*, *Backpropagation*, *Support Vector Machine*, klasifikasi dengan menggunakan pola (asosiatif dan diskriminatif), *Lazy Learners* (*k-Nearest neighbor* dan *Case-Based Reasoning*) dan algoritma klasifikasi lainnya seperti Algoritma Genetika [7]. Berdasarkan pemaparan sebelumnya, sistem yang akan digunakan untuk memprediksi kelayakan calon nasabah penerima

kredit menggunakan metode *K-Nearest Neighbor*. Algoritma tersebut tergolong dalam algoritma *Supervised Learning*, dimana hasil dari perhitungan yang baru nanti akan digolongkan berdasarkan kategori terbanyak [8]. Algoritma *K-Nearest Neighbor* adalah algoritma yang mengklasifikasikan sebuah objek baru dengan tetangga terdekatnya berdasarkan nilai K yang ditetapkan sebelumnya. Algoritma ini sangat bergantung pada jumlah K -tetangga terdekatnya. Apabila K bernilai 1 maka jumlah tetangga terdekatnya berjarak 1 dari data uji, begitu juga seterusnya hingga K bernilai n maka jumlah tetangga terdekatnya berjarak n dari data uji [9].

Alasan pemilihan algoritma *K-Nearest Neighbor* adalah hasil rata-rata akurasi dan pengukuran AUC yang didapatkan pada penelitian prediksi kelayakan pengajuan adalah 85.71% dan 0.836 dengan K sebesar 5 [10]. Hal tersebut mengindikasikan algoritma tersebut layak digunakan. Beralih ke penelitian lain, algoritma *K-NN* juga diterapkan pada penelitian dengan topik yang sama. Pada penelitian tersebut algoritma *K-NN* digunakan untuk klasifikasi penentuan pengajuan kartu kredit. Hasil pengujian mendapatkan rata-rata *precision* sebesar 92, *recall* sebesar 93, dan *accuracy* sebesar 93% [11]. Tidak hanya itu, algoritma KNN juga dapat digunakan dalam kasus lain selain dalam dunia perbankan. Contohnya adalah untuk mengklasifikasi kelayakan penerimaan Kartu Indonesia Sehat (KIS) bagi warga kurang mampu [12]. Pada penelitian tersebut menggunakan gabungan algoritma *K-NN* dengan algoritma *Naive Bayes*. Hasil pengujian menunjukkan akurasi sebesar 64% dan rata-rata waktu komputasi sebesar 0.01428 detik apabila hanya menggunakan algoritma *K-NN* saja. Namun jika dikombinasikan dengan algoritma *Naive Bayes* maka hasil akurasinya menjadi 96% dan waktu komputasinya sebesar 0.00118 detik lebih cepat daripada sebelumnya.

Kelebihan dari algoritma *K-Nearest Neighbor* adalah apabila data latih berjumlah banyak maka sangat efektif untuk mengklasifikasikannya, menghasilkan data yang akurat, metode yang baik dalam hal ruang pencarian (kelas tidak harus linear dipisahkan), dan apabila data latih mempunyai *noise*, maka metode ini sangat cocok untuk digunakan [13]. Namun kekurangan dari algoritma *K-Nearest*

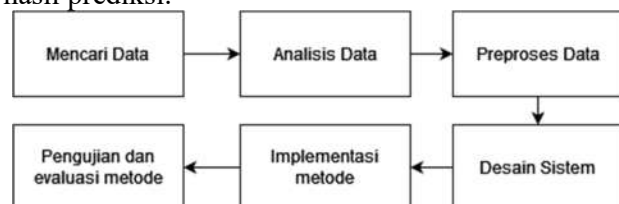
Neighbor antara lain algoritma tersebut tidak dapat mengetahui fitur-fitur yang berpengaruh dalam proses klasifikasi dan membutuhkan komputasi yang tinggi [14]. Untuk menutupi dua kekurangan tersebut diperlukan seleksi fitur dengan menggunakan korelasi Pearson terlebih dahulu sebelum masuk dalam proses prediksi dengan algoritma *K-NN*.

Tujuan dari pemilihan fitur adalah untuk mengurangi kompleksitas, menaikkan akurasi dan menentukan fitur yang optimal berdasarkan kumpulan fitur data [15]. Metode yang digunakan untuk pemilihan fitur adalah dengan menggunakan seleksi korelasi Pearson. Berdasarkan pada penelitian [16]. Hasil pengujian menunjukkan tingkat akurasi algoritma *Naive Bayes* yang didapatkan tanpa seleksi fitur dengan korelasi Pearson adalah sebesar 68.26%. Sedangkan hasil pengujian ketika menggunakan seleksi fitur dengan korelasi Pearson meningkat sebanyak 10.87% dengan akurasi akhir sebesar 79.13%.

Sistem yang akan dibangun adalah berupa aplikasi berbasis desktop. Dalam aplikasi akan tersemat algoritma *K-Nearest Neighbor* dengan fitur yang telah diseleksi oleh korelasi Pearson. Kemudian sebagai tahap uji kinerja algoritma akan dilakukan pengujian dengan menggunakan *Mean Absolute Error* (MAE) dan evaluasi hasil dengan melakukan uji *accuracy*, *precision*, dan *recall*.

II. METODE PENELITIAN

Agar penelitian dapat berjalan dengan terstruktur, maka diperlukan sebuah diagram alur dalam sebuah penelitian. Penelitian ini mengacu dalam diagram alur berikut yang terdapat pada Gbr 1. Alur penelitian diawali dari pencarian data, kemudian data dianalisis, beranjak ke proses berikutnya adalah mengkonversikan nilai data pada tahap *preprocessing data*, desain sistem, implementasi metode dan uji kinerja algoritma beserta evaluasi hasil prediksi.



Gbr 1. Diagram alur metode penelitian

A. Pencarian Data

Data yang dipakai adalah data nasabah Koperasi Bakti Sukaraja periode pencatatan 2016/2017. Data yang digunakan berjumlah 220 data nasabah. Data tersebut terbagi menjadi 200 data latih dan 20 data nasabah untuk data uji. Data yang dipakai dalam penelitian ini diambil dari penelitian [10] dan termuat dalam laman internet [17].

B. Analisis Data

Setelah data didapatkan pada tahap pencarian data. Tahap berikutnya adalah menganalisis data. Tujuan menganalisis data adalah untuk memilih atribut atau fitur yang terdapat pada data nasabah. Fitur atau atribut tersebut nantinya akan digunakan untuk proses penyeleksian fitur dan juga proses prediksi dengan menggunakan algoritma *K-Nearest Neighbor*. Fitur atau atribut yang akan digunakan dapat dilihat pada Tabel I.

TABEL I
TABEL ATRIBUT ATAU FITUR NASABAH

Atribut	Tipe Data	Isi
Jenis Kelamin	Karakter	Laki-laki, Perempuan
Pekerjaan	Karakter	Karyawan Swasta, Mengurus Rumah Tangga
Keperluan	Karakter	Pembelian sesuatu, Modal usaha
Jenis jaminan	Karakter	SPPT, Tidak ada
Jumlah pendapatan	Numerik	Masukkan dari pengguna
Jumlah pengeluaran	Numerik	Masukkan dari pengguna
Pinjaman sebelumnya	Numerik	Masukkan dari pengguna
Jumlah pengajuan	Numerik	Masukkan dari pengguna
Lama pinjam	Numerik	12 Bulan, 18 Bulan, dan 24 Bulan
Status kelayakan	Karakter	Diterima, Dikurangi

C. Preprocessing data

Data yang telah dipilih pada tahap analisis tidak dapat digunakan secara langsung untuk proses

seleksi. Data yang pada awalnya bertipe data non-numerik atau karakter harus diubah menjadi tipe data numerik. Sedangkan untuk data yang sudah bertipe data numerik seperti pendapatan, pengeluaran, pinjaman sebelumnya, jumlah pengajuan dan lama meminjam juga tidak dapat secara langsung digunakan dalam proses seleksi maupun prediksi. Beberapa data tersebut apabila digunakan langsung maka akan memberatkan algoritma Korelasi Pearson dalam menyeleksi fitur dan *K-Nearest Neighbor* dalam mengklasifikasi data. Hal tersebut dikarenakan ukuran data yang terlalu besar dan akan membuat nilai komputasinya sangat tinggi. Maka dari itu beberapa data tersebut harus disederhanakan kembali agar ukuran data yang dipakai menjadi lebih kecil dari sebelumnya. Cara yang akan digunakan adalah dengan menggolongkan beberapa data tersebut ke beberapa angka numerik sederhana (seperti 1, 2, dan seterusnya). Pemberian label nilai pada atribut seperti itu tidak dapat diberikan secara langsung, melainkan harus ada pertimbangan. Hal ini terkait dengan bobot suatu data yang diajukan. Adapun pada proses ini, data yang telah dikonversikan menjadi sebuah nilai untuk lebih detailnya dapat dilihat pada Tabel II.

TABEL II
TABEL KONVERSI NILAI ATRIBUT DATA

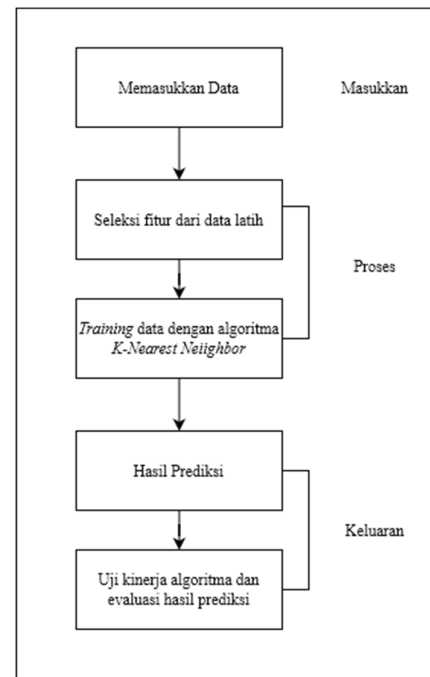
No.	Atribut	Nilai Atribut	Hasil Konversi
1	Jenis Kelamin	Laki-laki	2
		Perempuan	1
2	Keperluan	Modal usaha	2
		Pembelian barang	1
3	Pekerjaan	Karyawan swasta	4
		Mengurus rumah tangga	2
4	Jenis jaminan	SPPT	2
		Tidak ada	1
5	Lama pinjam	24 Bulan	3
		18 Bulan	2
		12 Bulan	1
6	Pendapatan	> Rp. 7.200.000	7

		Rp. 6.250.000 – Rp. 7.000.000	6
		Rp. 5.400.000 – Rp. 6.250.000	5
		Rp. 4.040.000 – Rp. 5.400.000	4
		Rp. 3.300.000 – Rp. 4.040.000	3
		Rp. 2.250.000 – Rp. 3.300.000	2
		Rp. 1.000.000 – Rp. 2.250.000	1
7	Pengeluaran	> Rp. 4.500.000	1
		Rp. 3.500.000 – Rp. 4.500.000	2
		Rp. 3.000.000 – Rp. 3.500.000	3
		Rp. 2.500.000 – Rp. 3.000.000	4
		Rp. 2.000.000 – Rp. 2.500.000	5
		Rp. 1.750.000 – Rp. 2.000.000	6
		Rp. 1.500.000 – Rp. 1.750.000	7
		Rp. 1.000.000 – Rp. 1.500.000	8
8	Pinjaman Sebelumnya	Tidak ada (Rp. 0)	4
		Rp. 1.000.000 – Rp. 2.000.000	3
		Rp. 3.000.000 – Rp. 4.000.000	2
		> Rp. 5.000.000	1
9	Jumlah pengajuan	Rp. 1.000.000 – Rp. 3.000.000	5
		Rp. 3.000.000 – Rp. 6.000.000	4
		Rp. 6.000.000 – Rp. 12.000.000	3
		Rp. 12.000.000 – Rp. 18.000.000	2
		Rp. 18.000.000 – Rp. 25.000.000	1

10	Status kelayakan	Diterima	1
		Dikurangi	0

D. Desain sistem

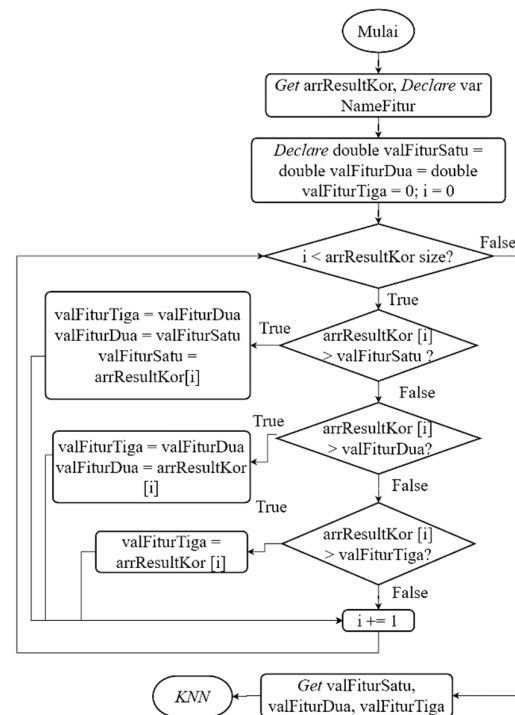
Sistem dirancang secara runtut mulai dari pengguna memasukkan data latih dan data uji hingga sistem menampilkan hasil prediksi. Adapun secara umum sistem akan melakukan tiga proses dasar. Ketiga proses dasar tersebut antara lain masukan, proses, dan keluaran. Untuk lebih jelas dapat dilihat pada Gbr 2.



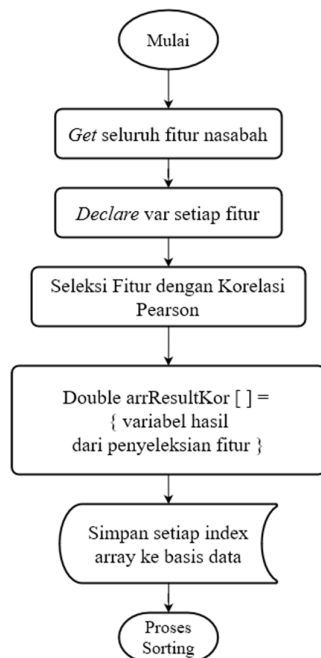
Gbr 2. Diagram alur proses secara umum

Proses akan dimulai ketika pengguna memasukkan data latih ke dalam sistem. Setiap fitur dari data nasabah akan diseleksi dengan korelasi Pearson terlebih dahulu. Dalam penyeleksian fitur, setiap fitur akan dihitung korelasinya dengan status kelayakan. Seperti korelasi antara jenis kelamin dengan status kelayakan, keperluan dengan status kelayakan, jenis jaminan dengan status kelayakan dan seterusnya. Pada tahap ini sistem menghasilkan perhitungan korelasi Pearson. Fitur dengan korelasi tertinggi dengan status penerimaan yang akan digunakan untuk proses prediksi dengan algoritma *K-Nearest Neighbor*. Untuk mengurutkan fitur

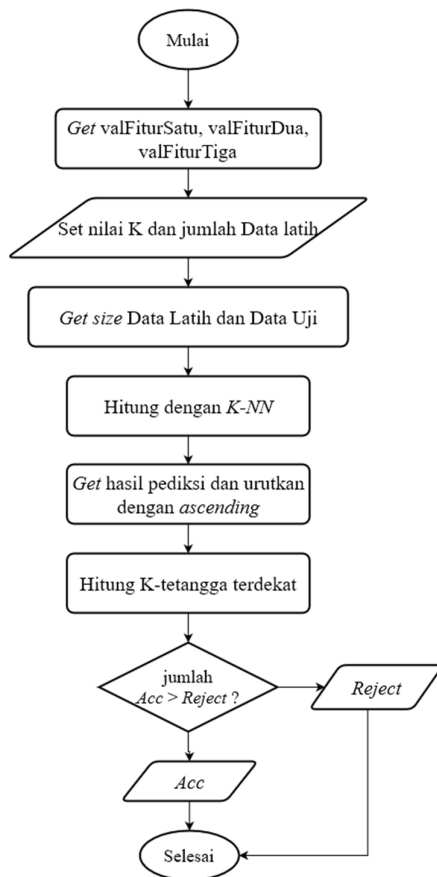
dengan korelasi tertinggi ke terendah digunakan algoritma *Sorting*. Tujuan dari algoritma *Sorting* adalah untuk mendapatkan tiga fitur dengan korelasi tertinggi. Fitur tersebut kemudian akan digunakan untuk memprediksi dengan menggunakan *K-Nearest Neighbor*. Pengguna akan menentukan ukuran data latih dan juga nilai K yang akan dipakai dalam proses prediksi. Setelah proses prediksi, hasil akan ditampilkan pada layar aplikasi. Hasil prediksi bisa menjadi pertimbangan bagi pihak kreditur untuk menerima pengajuan pinjaman dari seorang nasabah atau sebaliknya. Kemudian langkah terakhir untuk mengetahui apakah sistem layak digunakan dalam prediksi maka perlunya untuk menguji kinerja algoritma *K-Nearest Neighbor*. Metode yang digunakan adalah dengan menggunakan *Mean Absolute Error*. Selain itu evaluasi juga dibutuhkan. Evaluasi dilakukan dengan perhitungan *accuracy*, *precision*, dan *recall*. Hasil uji kinerja dan evaluasi akan ditampilkan pada halaman uji algoritma sebagai penutup dari semua menu. Uraian penjelasan mulai dari seleksi fitur, proses *sorting*, dan proses prediksi secara berurutan digambarkan pada diagram alur 3, 4 dan 5



Gbr 4. Diagram alur seleksi dan *sorting* fitur



Gbr 3. Diagram alur korelasi Pearson



Gbr 5. Diagram alur prediksi dengan *K-Nearest Neighbor*

E. Implementasi metode

Pada tahap ini akan dirancang sebuah sistem aplikasi berbasis desktop. Kemudian untuk membantu penyimpanan data nasabah, perhitungan seleksi, dan data pengguna dibutuhkan suatu basis data. Basis data yang digunakan adalah PostgreSQL. Perhitungan yang akan direkayasa dalam aplikasi antara lain seleksi korelasi Pearson, *K-Nearest Neighbor*, uji kinerja algoritma yang terdiri dari *MAE*, *Accuracy*, *Precision*, dan *Recall*.

1) Korelasi Pearson

Proses seleksi fitur menggunakan perhitungan korelasi Pearson. Tujuan menggunakan korelasi Pearson adalah untuk menentukan arah hubungan, kekuatan hubungan, dan signifikansi kekuatan hubungan antara dua variabel, dengan syarat persebaran data kedua variabel terdistribusi secara normal [18]. Berikut adalah rumus

untuk perhitungan korelasi Pearson seperti Pers. (1)

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \quad (1)$$

Simbol r adalah simbol korelasi Pearson. Dalam perhitungan korelasi Pearson melibatkan variabel bebas (fitur atau atribut selain status kelayakan) dan variabel terikat (fitur atau atribut status). Variabel bebas bisa diganti-ganti selama perhitungan. Berbeda dengan variabel terikat. Variabel tersebut nilainya dibuat tetap untuk mengetahui seberapa pengaruhnya variabel tersebut dalam mempengaruhi variabel bebas. Variabel bebas disimbolkan dengan huruf "x" dan variabel terikat disimbolkan dengan huruf "y".

2) *K-Nearest Neighbor*

Berdasarkan pemaparan sebelumnya tentang algoritma *K-Nearest Neighbor*. Algoritma tersebut merupakan suatu algoritma untuk mengklasifikasikan suatu data baru berdasarkan jumlah K tetangga terdekatnya. Persamaan yang digunakan dalam *K-Nearest Neighbor* dinamakan *Euclidean Distance*. Berikut adalah persamaan *Euclidean Distance* seperti Pers. (2)

$$D = \sqrt{\sum_{i=1}^n (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (2)$$

Keterangan:

D : Jarak terdekat

x_1, y_1, z_1 : Fitur atau atribut data latih

x_2, y_2, z_2 : Fitur atau atribut data uji

n : Jumlah keseluruhan fitur yang akan dihitung

i : Jumlah iterasi perhitungan dari 1 hingga n iterasi

Proses prediksi dengan klasifikasi KNN dapat diimplementasikan ke dalam sistem seperti berikut:

1. Fitur yang telah didapatkan dari perhitungan korelasi Pearson dihitung

dengan menghitung jarak terdekat pada data latih.

2. Setelah itu lakukan proses pengurutan data dengan cara *sorting* dari terkecil hingga terbesar.
3. Menentukan nilai K untuk penentuan jumlah tetangga terdekat setelah proses *sorting*. Untuk K yang ditetapkan bernilai ganjil. Berdasarkan penelitian sebelumnya, penetapan nilai K untuk mendapatkan hasil yang optimal adalah sebesar 3.
4. Mendapatkan hasil status prediksi sesuai dengan nilai K yang telah ditentukan sebelumnya.

3) Uji kinerja algoritma

Hasil kinerja algoritma KNN akan diuji dengan *Mean Absolute Error (MAE)*. Tujuan dari perhitungan *MAE* adalah untuk mengukur jumlah kesalahan atau *error* suatu algoritma dalam mengklasifikasi atau memprediksi suatu model. Selain itu *MAE* juga dapat digunakan untuk mengetahui kedekatan suatu nilai prediksi (*predictor*) dengan nilai sebenarnya (*actual*) [19]. Persamaan *MAE* dapat dilihat dalam Pers. (3).

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i| \quad (3)$$

Keterangan:

n : Total data uji

A_i : Hasil data aktual ke- i

F_i : Hasil data prediksi ke- i

d : Evaluasi prediksi

4) Evaluasi hasil

Hasil prediksi yang dihasilkan oleh algoritma *K-Nearest Neighbor* dengan fitur yang telah diseleksi oleh korelasi Pearson juga harus dievaluasi. Tujuan evaluasi juga hampir sama dengan uji kinerja algoritma yaitu untuk mengevaluasi kelayakan suatu algoritma jika diterapkan kedalam suatu sistem. Evaluasi prediksi dapat dihitung dengan *precision*, *recall* dan *accuracy*. Penjelasan *precision*, *recall* dan *accuracy* telah dijelaskan pada

penelitian [20]. *Precision* adalah suatu kondisi aktual positif yang diprediksi secara benar. Hasil *Precision* dapat diketahui melalui persamaan Pers. (4). *Recall* adalah suatu kondisi aktual positif yang dapat dikenali secara benar. Hasil *Recall* dapat diketahui melalui persamaan Pers. (5). *Accuracy* adalah ketepatan dari semua data yang diprediksi secara benar. Hasil *Recall* dapat diketahui melalui persamaan Pers. (6).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+TN} \quad (5)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Keterangan:

TP (True Positive) : jumlah hasil prediksi benar dan aplikasi dapat memprediksi dengan benar sesuai dengan data aktualnya yaitu benar.

TN (True Negative) : jumlah hasil prediksi salah dan aplikasi dapat memprediksi dengan salah sesuai dengan data aktualnya yaitu salah.

FP (False Positive) : jumlah hasil prediksi benar dan aplikasi tidak dapat memprediksi dengan benar sesuai data aktualnya yaitu salah.

FN (False Negative) : jumlah hasil prediksi salah dan aplikasi tidak dapat memprediksi dengan benar sesuai data aktualnya yaitu benar.

F. Uji metode

Pengujian metode atau algoritma menggunakan teknik pembagian data. Teknik tersebut sering disebut dengan teknik *percentage split*. Sistem akan membaca jumlah data latih dari pengguna. Data tersebut akan digunakan sebagai model untuk membangun model klasifikasi. Setelah total data latih terbaca oleh sistem, kemudian sistem akan otomatis menghitung sisa dari data yang telah digunakan oleh data latih untuk total data uji. Sebagai contoh, apabila pengguna memasukkan jumlah data latih sebesar 10% dari data latih maka sistem akan secara langsung menetapkan sisa dari keseluruhan data untuk total data uji. Jadi total data ujinya adalah 90%. Setelah jumlah data latih dan

data uji didapatkan sistem. Proses berikutnya sistem akan menguji kinerja algoritma *K-Nearest Neighbor* dengan *Mean Absolute Error*, *Precision*, *Recall* dan *accuracy*.

III. HASIL DAN PEMBAHASAN

A. Hasil penelitian

Aplikasi adalah sebagai wadah untuk mewujudkan apa yang telah direncanakan sebelumnya. Selain itu dengan dibuatnya aplikasi, merupakan bentuk nyata penerapan sebuah metode ke dalam suatu sistem. Aplikasi ini juga dapat digunakan oleh pengguna nantinya. Dalam aplikasi terdapat beberapa menu yang dapat diakses. Beberapa halaman tersebut antara lain halaman masuk, halaman daftar, halaman utama, halaman uji, halaman latihan, halaman tambah dan ubah data, halaman prediksi, dan halaman uji algoritma.

1) Halaman masuk

Pada halaman ini pengguna akan memasukkan *username* dan *password* yang telah terdaftar sebelumnya. Sistem akan mengecek *username* dan *password* yang dimasukkan oleh pengguna. Apabila benar pengguna dapat berpindah ke halaman utama.

2) Halaman daftar

Pada halaman ini pengguna yang belum memiliki akun dapat mendaftar. Pengguna dapat memasukkan *username*, *password*, nama pengguna, nomor telepon, dan sistem akan secara langsung melabeli bahwa akun tersebut aktif.

3) Halaman utama

Pada halaman utama, pengguna dapat melihat sekilas info seperti jumlah data latihan, jumlah data uji, total nasabah yang mengajukan pinjaman, jumlah fitur yang digunakan untuk memprediksi kelayakan.

4) Halaman uji

Pada halaman uji, pengguna dapat membaca data, memasukkan data, mengubah data, dan menghapus data. Fitur tambahan yang tersedia pada menu ini antara lain sistem dapat menampilkan data sesuai status penerimaan

pengajuan pinjaman. Status yang dimaksud adalah diterima atau tidaknya sebuah pengajuan.

5) Halaman latihan

Pada halaman latihan, pengguna juga dapat melakukan proses hal yang sama seperti halaman uji. Seperti membaca data, memasukkan data, mengubah data dan menghapus data. Fitur tambahan juga sama dengan menu halaman uji. Fitur tersebut adalah sistem dapat menampilkan data sesuai status penerimaan pengajuan pinjaman

6) Halaman tambah dan ubah data

Jumlah halaman ini adalah hanya satu, namun dapat dipergunakan untuk multi fungsi sekaligus. Pada halaman ini digunakan pada halaman uji dan halaman latihan. Halaman tambah data akan terbuka ketika pengguna menekan tombol tambah data pada halaman data latihan atau halaman data uji. Kemudian apabila pengguna ingin mengubah salah satu data latihan atau data uji pengguna dapat memilih data pada tabel yang tersedia kemudian mengeklik secara dua kali dan halaman ubah data akan terbuka.

7) Halaman prediksi

Halaman ini dapat digunakan oleh pengguna untuk melakukan prediksi pengajuan pinjaman seorang nasabah. Hasil prediksi dikategorikan menjadi layak dan perlu adanya pengurangan dari jumlah pengajuan sebelumnya. Fitur yang dapat digunakan oleh pengguna pada halaman ini antara lain tabel untuk menampilkan hasil perhitungan *KNN*, tabel untuk menampilkan data nasabah yang siap untuk diprediksi, tampilan saran bagi pengguna apakah pengajuan nasabah bersangkutan layak untuk diterima atau tidak. Untuk lebih jelasnya halaman ini dapat dilihat pada Gbr. 6

8) Halaman uji algoritma

Halaman terakhir pada aplikasi ini adalah halaman untuk menguji algoritma. Tujuan dari pengujian algoritma adalah untuk mengukur kinerja dari algoritma *K-NN* dalam

memprediksi kelayakan pinjaman. Semua proses uji kinerja dan evaluasi seperti uji *Mean Absolute Error*, *Precision*, *Recall*, dan *Accuracy* akan ditampilkan dalam halaman ini. Halaman uji algoritma dapat dilihat pada Gbr. 7

B. Pembahasan

Pada poin ini akan membahas hasil dari penelitian dengan topik yang telah dijelaskan sebelumnya. Pembahasan akan dimulai dari penyeleksian fitur dengan korelasi Pearson, kemudian uji coba prediksi dengan menggunakan algoritma *K-Nearest Neighbor* dan berakhir dengan pengujian kinerja beserta evaluasi hasil prediksi menggunakan uji *Mean Absolute Error*, *Precision*, *Recall* dan *Accuracy*.

1) Penyeleksian fitur

Telah diketahui sebelumnya dalam menyeleksi fitur menggunakan suatu metode penyeleksian fitur, metode digunakan adalah korelasi Pearson dan disimbolkan dengan r . Fitur akan dihitung dan diambil dengan peringkat tiga teratas. Data yang digunakan untuk pelatihan (*training*) adalah data latih. Jumlah data yang dipakai adalah sebanyak 200 data nasabah. Fitur yang korelasinya paling kuat dengan status kelayakan antara lain pengeluaran, pendapatan dan fitur yang terakhir adalah pekerjaan nasabah. Fitur dengan hasil dan peringkat di bawah fitur-fitur tersebut tidak akan digunakan untuk proses prediksi dengan menggunakan algoritma *K-Nearest Neighbor*. Adapun lebih jelasnya terdapat pada Tabel. III

Prediksi kelayakan (Prediction)

Alias: DU20220315201134

Pekerjaan: Mengurus Rumah Tangga

Kebutuhan: Modal Usaha

Jenis Agunan: SPPT

Lama: 12, Jenis Kelamin: Perempuan

Pendapatan: Rp. 5.400.000

Pengajuan: Rp. 8.000.000

Pinjaman Sebelumnya: Rp. 5.000.000

Persentase Data Latih: 5%, Nilai K: 3

[+] Log Out

Data Nasabah Siap Prediksi

Alias	JK	Pekerjaan	Kebutuhan	Jaminan	Pendapatan	Pengaj
DU2022031...	Perempuan	Mengurus R...	Modal Usaha	SPPT	Rp5.400.000	Rp8.000

Hasil Prediksi

Alias Nasabah	Hasil Perhitungan	Status Kelayakan
DL20220312113040	7	0
DL20220312113853	5,8309518948453	1
DL20220312114017	4,12310562561766	0
DL20220312114105	3,16227766016838	1

Penjelasan :
Nasabah dengan label DU20220315201134
sebabnya pengajuan kreditnya Terima

Nasabah	Jenis Kelamin	Pekerjaan	Jumlah Pengajuan	Keperluan	Agunan	Saran Sistem
DU202203...	Perempuan	Mengurus...	Rp8.000.000	Modal Usaha	SPPT	Diterima

Gbr 6. Halaman prediksi nasabah

Jumlah Data Latih	40
Jumlah Data Uji	16
Mean Absolute Error	0,0000
Accuracy	$(TP + TN) / (TP + TN + FN + FP)$ $(10 + 6) / (10 + 6 + 0 + 0) \times 100 = 100 \%$
Precision	$TP / TP + FP$ $(10 / 10 + 0) \times 100 = 1$
Recall	$TP / TP + FN$ $(10 / 10 + 0) \times 100 = 1$

Prediksi	Layak	Dikurangi
Layak	10	0
Dikurangi	0	6

Gbr 7. Halaman uji kinerja algoritma

TABEL III
TABEL PENYELEKSIAN FITUR

Fitur	Hasil Korelasi Pearson	Peringkat	Label
Jenis kelamin	0.32913658587373	5	F1
Pekerjaan	0.449457803098817	3	F2
Keperluan	0.0244876722179252	9	F3
Jenis Jaminan	0.206684975147631	6	F4
Jumlah Pendapatan	0.496333029308101	2	F5
Jumlah Pengeluaran	0.607543142257494	1	F6
Pinjaman Sebelumnya	0.38038728784656	4	F7
Pengajuan	0.0569626525165957	8	F8
Lama Pinjam	0.0854484231826708	7	F9

2) Prediksi dengan K-Nearest Neighbor

Setelah tiga fitur dengan hasil korelasi paling tinggi di antara fitur yang lainnya telah didapatkan. Sebelum data nasabah diprediksi, sistem akan melakukan *preprocessing* data dengan mengonversikan isi dari fitur tersebut. Setelah itu sistem akan melakukan prediksi

dengan algoritma *K-Nearest Neighbor* dengan fitur yang telah didapatkan sebelumnya. Total data yang digunakan untuk uji coba *training* algoritma *K-Nearest Neighbor* adalah berjumlah 10 data latih dapat dilihat pada Tabel IV dan 1 data uji dapat dilihat pada Tabel V.

TABEL IV
TABEL DATA LATIH

No	Alias	Label			Status
		F2	F5	F6	
1	DL20220312113040	8	2	4	0
2	DL20220312113853	7	2	2	1
3	DL20220312114017	6	4	2	0
4	DL20220312114105	5	4	2	1
5	DL20220312114311	2	6	2	1
6	DL20220312114334	6	3	2	0
7	DL20220312114407	1	7	2	1
8	DL20220312114445	3	5	2	1
9	DL20220312114559	2	5	2	1

10	DL20220312114 640	7	2	2	1
----	----------------------	---	---	---	---

TABEL V
TABEL DATA UJI

No	Alias	Label			Status
		F 2	F 5	F 6	
1	DU20220315201 134	2	5	2	?

Nilai K digunakan untuk mengurutkan jumlah tetangga terdekat adalah masukkan dari pengguna. Proses di dalam sistem akan diurutkan secara *ascending* berdasarkan nilai K yang telah ditetapkan yaitu sebesar 3. Setelah diurutkan sistem akan menghitung berapa banyak data latih yang status nya ditolak dan diterima. Apabila jumlah yang ditolak lebih banyak dari yang diterima maka sistem akan menampilkan pesan bahwa sebaiknya permohonan pengajuan pinjaman ditolak begitu juga sebaliknya. Untuk lebih detailnya dapat dilihat pada Tabel V untuk hasil perhitungan, Tabel VI untuk sorting berdasarkan K-tetangga terdekat dan Tabel VIII untuk hasil prediksi.

TABEL VI
HASIL PERHITUNGAN K-NEAREST NEIGHBOR

Alias	Urutan	Hasil	Status
DL2022031211 3040	10	7.0	0
DL2022031211 3853	9	5.830951 89	1
DL2022031211 4017	6	4.123105 62	0
DL2022031211 4105	5	3.162277 66	1
DL2022031211 4311	3	1.0	1
DL2022031211 4334	7	4.472135 95	0

DL2022031211 4407	4	2.236067 97	1
DL2022031211 4445	2	1.0	1
DL2022031211 4559	1	0.0	1
DL2022031211 4640	8	5.830951 89	1

TABEL VII
HASIL MENGURUTKAN BERDASARKAN K-
TETANGGA TERDEKAT

Alias	Urutan	Status
DL20220312114559	1	1
DL20220312114445	2	1
DL20220312114311	3	1

TABEL VIII
HASIL PREDIKSI PENGAJUAN PINJAMAN
NASABAH

No.	Alias	Status	Konversi
1	DU20220315201134	1	Diterima

3) Uji kinerja algoritma

Tahap terakhir pada penelitian ini adalah menguji sekaligus mengevaluasi kinerja dari algoritma *K-Nearest Neighbor* dengan fitur yang telah diseleksi dalam memprediksi kelayakan pengajuan pinjaman nasabah. Teknik yang dipakai adalah dengan membagi keseluruhan data (data latih dan data uji) menjadi beberapa persen. Teknik tersebut dinamakan rasio perbandingan data atau sering disebut dengan *percentage split*. Rincian total keseluruhan data adalah 220 data dengan pembagian 200 data latih dan 20 data uji. Skenario yang diambil adalah dengan membagi keseluruhan data menjadi 5% data latih dan 95% data uji, 10% data latih dan 90% data uji, 15% data latih dan 85% data uji, 20% data latih dan 80% data uji. Kemudian nilai K yang dipakai sama seperti saat tahap uji coba yaitu sebesar 3. Hasil dari pengujian dapat dilihat pada Tabel. IX.

TABEL IX
HASIL UJI KINERJA DAN EVALUASI ALGORITMA

Rasio Perbandingan (<i>percentage split</i>)				<i>Mean Absolute Error</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
Persen data latih	Total data latih	Persen data uji	Total data uji				
5%	10	95%	19	0.26315789	76,92	83.33	73.68 %
10%	20	90%	18	0.05555556	100	90.90	94.44 %
15%	30	85%	17	0.05882353	100	90	94.11 %
20%	40	80%	16	0	100	100	100 %
Rata-rata				0.094384245	94.23	91.06	90.56 %

Melalui skenario pengujian dengan *percentage split* atau membagi data latih dan data uji menjadi beberapa bagian, menunjukkan bahwa tingkat akurasi mengalami peningkatan pada saat penambahan data latih sebanyak 5% dari data latih atau sebanyak 10 menjadi 20 % dari data latih atau sebanyak 40. Akurasi yang awalnya sebesar 73.68% meningkat menjadi 94% pada saat data latih bertambah sebanyak 10% dan naik secara pesat hingga saat akurasi mencapai kondisi maksimum di 20% data latih dengan akurasi yang didapatkan sebesar 100%. Untuk pengujian lain juga ikut meningkat seperti pada *precision* dan *recall*. Hal tersebut ditunjukkan dengan hasil pengukuran terendah saat pengujian pada data latih sebanyak 5% dari data latih atau sebanyak 10 dengan *precision* sebesar 76.92 dan *recall* sebesar 83.33. Kemudian hasil tertinggi dari pengujian *precision* dan *recall* terjadi saat penambahan data latih sebanyak 20% dari data latih atau berjumlah 40. Hasil yang didapatkan adalah 100 untuk *precision* dan *recall*.

Hal ini berbanding terbalik dengan hasil pengujian *MAE*. Seiring bertambahnya jumlah data latih yang digunakan maka hasil dari *MAE* akan semakin kecil. Karena hasil dari pengujian *MAE* adalah rentang dari 0 sampai 1. Semakin baik suatu model klasifikasi apabila *error* nya yang mendekati 0. Dapat ditarik sebuah kesimpulan melalui hasil uji kinerja algoritma dan evaluasi hasil prediksi. Apabila semakin besar selisih antara data latih dan data uji yang digunakan, maka akan membentuk sebuah model pengklasifikasian berdasarkan status aktual dari data latih tersebut dan jumlah tersebut akan semakin lengkap seiring bertambahnya data latih. Hal ini yang menyebabkan akurasi yang didapatkan

akan jauh lebih baik dari sebelumnya saat melakukan proses prediksi dengan data uji baru.

IV. KESIMPULAN

Kesimpulan yang dapat ditarik selama penelitian ini mulai dari awal sampai akhir dan hingga terwujudnya sebuah sistem yang dapat memprediksi kelayakan pengajuan dengan memanfaatkan algoritma *K-Nearest Neighbor* beserta seleksi Korelasi Pearson antara lain sebagai berikut.

Sebelum sistem melakukan penyeleksian fitur dan prediksi. Sistem akan melakukan *preprocessing* data. Pada tahap tersebut sistem akan melakukan pengkonversian setiap fitur. Tahap ini sangat penting dikarenakan algoritma *K-NN* sangat bergantung sekali dengan pelabelan suatu data. Selain itu mengetahui kekurangan dari *K-NN* adalah tingkat komputasinya yang sangat tinggi maka langkah ini dapat diambil untuk menutupi kekurangan tersebut dan juga mempercepat kinerja dari algoritma *K-NN*.

Lalu berpindah ke proses seleksi fitur nasabah. Sistem dapat menyeleksi fitur secara dinamis dengan bantuan algoritma seleksi korelasi Pearson beserta teknik *sorting* data. Adapun total fitur yang diambil berjumlah tiga tertinggi dari fitur lainnya. Selama penelitian fitur yang terseleksi antara lain pekerjaan, pendapatan dan pengeluaran.

Fitur yang telah didapatkan akan dilatih untuk masuk kedalam proses prediksi. Data uji yang telah dimasukkan oleh pengguna akan dilatih dengan data latih. Sistem akan melakukan proses klasifikasi secara otomatis ketika pengguna menekan tombol prediksi pada aplikasi. Hasil dari prediksi secara langsung ditampilkan pada layar. Keputusan kelayakan penerimaan pengajuan berdasarkan hasil perhitungan jumlah *K* tetangga terdekat yang

dimasukkan oleh pengguna. Kemudian untuk mengetahui kinerja algoritma dan mengevaluasi hasil prediksi sistem dapat secara langsung menampilkan hasilnya. Selama penelitian rata-rata yang didapatkan adalah *accuracy* sebesar 90.56%, *precision* sebesar 94.23, *recall* sebesar 91.06 dan uji *mean absolute error* sebesar 0.094384245

V. SARAN

Saran yang bisa penulis sampaikan kepada peneliti yang ingin meneruskan atau mengembangkan penelitian yang telah dilakukan antara lain. Telah dijelaskan bahwa sistem ini pada saat seleksi fitur dimasukkan sebuah teknik *sorting* sehingga tampak dinamis. Untuk penelitian berikutnya bisa dilakukan seleksi fitur secara manual dengan memilih fitur yang berdasarkan pilihan pengguna. Kemudian sebaiknya perlu pembandingan dengan algoritma lain untuk mengetahui algoritma terbaik pada kasus seperti ini. Lalu terakhir saran dari penulis adalah media aplikasi sangat boleh sekali diterapkan dengan teknologi yang modern di penelitian mendatang.

UCAPAN TERIMA KASIH

Rasa terima kasih pertama akan saya panjatkan kepada Allah SWT. Berkat rahmat dan hidayah-Nya penelitian ini dapat diselesaikan sesuai dengan target yang telah direncanakan dengan baik. Tidak lupa sholawat dan salam semoga tetap tercurahkan ke Nabi Muhammad SAW. Kemudian akan saya sampaikan kepada kedua orang tua saya, baik Ibu dan Ayah yang telah mendukung dan mendoakan saya sepenuh hati. Tidak lupa terimakasih akan saya berikan kepada Dr. Yuni Yamasari, S.Kom., M.Kom. sebagai dosen pembimbing, berkat arahan dan bimbingan dari beliau penelitian ini dapat terlaksana dengan baik hingga selesai. Rasa terima kasih berikutnya akan saya berikan kepada teman-teman saya dan juga pihak yang telah menyukseskan penelitian ini baik itu berupa ilmu ataupun dukungan kepada saya.

REFERENSI

- [1] Y. Mbulu, P. B. Kellen, L. M. Ndoen and B. Marselinus, "Pengaruh Jumlah Simpanan

dan Jumlah Pinjaman Anggota Terhadap Sisa Hasil Usaha pada Koperasi Kredit Handayani Bajawa," *Jurnal Akuntansi: Transparansi dan Akuntabilitas*, vol. Vol.7 No.2, pp. 165-172, 2019.

- [2] Afriyeni, "Analisis Pengaruh Pemberian Kredit Terhadap Profitabilitas PT. Bank Pembangunan Daerah (BPD) Sumatera Barat," *Jurnal KBP*, vol. I, pp. 95 - 107, 2013.
- [3] I. Rapida, "Faktor 5C dalam Pembiayaan Kepemilikan Rumah di Bank BJB Syariah Kantor Cabang Pembantu Rancaekkek," *Jurnal Manajemen Perbankan Syariah*, vol. 4, pp. 101-111, 2021.
- [4] B. Margono, "Usaha Perbankan Dalam Mengatasi Credit Crunch Perbankan di Indonesia Dengan Mengikuti Kebijakan Pemerintah dan Kredit Selektif Menggunakan Prinsip 5C," *Jurnal Ilmiah Akuntansi dan Keuangan*, vol. IV, pp. 272-285, 2021.
- [5] A. O. Adebayo and S. M. Chaubey, "Data Mining Classification Techniques On The Analysis Of Student's Performance," *Global Scientific Journals*, pp. 79-95, 2019.
- [6] I. Cholissodin and E. Riyandi, Analisis Big Data (Teori dan Aplikasi), Malang: Fakultas Ilmu Komputer, Universitas Brawijaya, 2018.
- [7] J. K. Han, M and J. Pei, Data Mining: Concept and Techniques, Third Edition, Waltham: Morgan Kaufmann Publishers, 2012.
- [8] F. Gorunescu, Data Mining: Concepts, Models and Techniques, Berlin Heidelberg: Springer, 2011.
- [9] H. Pratiwi, M. A. Mukid and T. Widiarihi, Credit Scoring Analysis Using Pseudo Nearest Neighbor, Semarang: Departemen Statistika Universitas Diponegoro, 2019.
- [10] Roviani, D. Supriadi and I. D. Iskandar, "Prediction Of Cooperative Loan Feasibility Using The K-Nearest Neighbor Algorithm,"

- Jurnal PILAR Nusa Mandiri*, vol. 17, pp. 39 - 46, 2021.
- [11] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *Jurnal Ilmiah MATRIK*, pp. 73-82, 2020.
- [12] Y. F. Safri, R. Arifudin and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healty Card Indonesia Giving to The Poor," *Scientific Journal of Informatics*, vol. 5, pp. 9-17, 2018.
- [13] T. C. Pratama, "Penerapan Metode K-Nearest Neighbour Dalam Menentukan Kelayakan Calon Nasabah yang Layak Untuk Kredit Mobil (Studi Kasus: PT. Astra Internasional, Tbk-Toyota)," *Jurnal Riset Komputer (JURIKOM)*, vol. Vol. 5 No. 4, pp. 402-408, 2018.
- [14] E. Nasri and A. S. A W, "Aplikasi Seleksi Penentuan Nasabah Untuk Penjualan Barang Secara Kredit Dengan Algoritma K-Nearest Neighbor," *Jurnal Sains dan Teknologi*, vol. 4, pp. 1-11, 2020.
- [15] H. Wen and J. Zhao, "Aspect term extraction of E-commerce comments based on model ensemble," in *13th International Computer Conference on Wavelet Active Media Technology and Information Processing ICCWAMTIP*, Chengdu, China, 2017.
- [16] Khoirudin and M. B. Hanif, "Sistem Aplikasi Prediksi Penyakit Diabetes Menggunakan Fiture Selection Korelasi Pearson dan Klasifikasi Naive Bayes," *Pengembangan Rekayasa dan Teknologi*, vol. 16, pp. 199-205, 2020.
- [17] I. D. Iskandar, "Github," 12 Maret 2021. [Online]. Available: <https://github.com/iqbaldzi13/Data-Set>. [Accessed 21 Februari 2022].
- [18] E. Roflin and F. E. Zulvia, *Kupas Tuntas Analisis Korelasi, Pekalongan, Jawa Tengah: PT Nasya Expanding Management*, 2021.
- [19] C. . N. Villavicencio, J. J. Escudero M, X. . A. Inbaraj, J.-H. Jeng and J.-G. Hsieh, "COVID-19 Prediction Applying Supervised Machine Learning Algorithms with Comparative Analysis Using WEKA," *Journal Algorithms*, vol. 14, pp. 1-22, 2021.
- [20] A. P, "A comparasion of machine learning techniques for customer churn prediction," *International J.Pure Application Math*, vol. 3 No.1, pp. 1-9, 2018.