P-ISSN: 1411-3635
E-ISSN: 2962-5645
https://doi.org/10.37817/Tekinfo.v26i1

Semi Supervised Learning Model Untuk Prediksi Kelancaran Pembayaran Kredit Kendaraan Bermotor

[1]Marsella Dwi Utami, [2]Noviandi, [3]Habibullah Akbar, [4]Diah Aryani

[1234]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul, DKI Jakarta

E-mail: [1]marselladwiutami@student.esaunggul.ac.id, [2]noviandi@esaunggul.ac.id, [3]habibullah.akbar@esaunggul.ac.id, [4]diah.aryani@esaunggul.ac.id

ABSTRAK

Tingginya permintaan kredit kendaraan sepeda motor menyebabkan meningkatnya jumlah perusahaan pembiayaan untuk membantu masyarakat atau konsumen dalam membeli sepeda motor dengan cara kredit. Namun terdapat tantangan dalam pembayaran kredit kendaraan dimana sebagian besar nasabah sering kali tidak mematuhi tenggat waktu pembayaran angsuran, bahkan ada yang mencoba menghindari kewajiban pembayaran mereka yang berdampak negatif pada keuangan perusahaan. Untuk mengatasi masalah ini, maka dibuat analisa dengan menggunakan teknik semi supervised learning yang mampu memprediksi sejauh mana kelancaran pembayaran kredit kendaraan oleh nasabah. Dalam penelitian ini menggunakan 2 algoritma yaitu, algoritma K-Means Clustering dan algoritma Random Forest. Algoritma K-Means Clustering digunakan untuk membuat label target untuk pengujian analisa klasifikasi. Dari hasil analisa, diperoleh k = 2, dengan nilai SSE sebesar 62,23 dan nilai Silhouette Score sebesar 0,53. Sehingga dapat dilakukan penamaan dari setiap cluster, yaitu cluster lancar dan macet. Setelah diketahui labelnya, maka dilakukanlah Analisa klasifikasi Random Forest. Terdapat beberapa fitur yang sangat berpengaruh dalam kelancaran pembayaran kredit kendaraan bermotor ini seperti 'PEKERJAAN','PENGHASILAN_PER_BULAN','JUMLAH_TANGGUNGAN' dan 'CICILAN_NASABAH. Dari hasil analisa menghasilkan skor accuracy sebesar 98%, precission sebesar 100%, recall sebesar 97%, dan F1 Score sebesar 98%.

Kata kunci : ***Kredit Kendaraan Bermotor, K-Means Clustering, Semi Supervised Learning, Random Forest.***

ABSTRACT

The high demand for motorcycle credit has led to an increase in the number of financing companies to help people or consumers buy motorcycles on credit. However, there are challenges in vehicle credit payments where most customers often do not comply with installment payment deadlines, some even try to avoid their payment obligations which have a negative impact on the company's finances. To overcome this problem, an analysis was made using a semi-supervised learning technique that is able to predict the extent to which vehicle credit payments are smooth by customers. This study uses 2 algorithms, namely the K-Means Clustering algorithm and the Random Forest algorithm. The K-Means Clustering algorithm is used to create target labels for classification analysis testing. From the results of the analysis, k = 2 was obtained, with an SSE value of 62.23 and a Silhouette Score value of 0.53. So that each cluster can be named, namely smooth and stalled clusters. After the label is known, the Random Forest classification analysis is carried out. There are several features that greatly influence the smooth payment of motor vehicle credit such as 'PEKERJAAN','PENGHASILAN_PER_BULAN','JUMLAH_TANGGUNGAN' and

https://doi.org/10.37817/Tekinfo.v26i1
P-ISSN: 1411-3635
E-ISSN: 2962-5645

‘CICILAN_NASABAH. The analysis results showed an accuracy score of 98%, a precision of 100%, a recall of 97%, and an F1 Score of 98%.

Keyword : ***Motor Vehicle Credit; K-Means Clustering; Semi Supervised Learning; Random Forest.***

1. PENDAHULUAN

Ketergantungan masyarakat Indonesia pada kendaraan pribadi sebagai sarana transportasi utama dalam kehidupan sehari-hari telah menghasilkan berbagai variasi dalam jenis kendaraan yang digunakan (Rusli et al., 2021). Menurut data dari Korlantas Polri, jumlah kendaraan bermotor yang terdaftar di Indonesia pada tanggal 26 Maret 2023 mencapai 154.236.431 unit (Putri Arny Christika, 2023). Meningkatnya jumlah masalah terkait pemberian kredit dalam dunia finansial menyebabkan banyak pihak merasa bahwa pemberian kredit dapat menjadi solusi untuk mengatasi tantangan finansial, baik dalam skala individu maupun perusahaan (Jasmir et al., 2022). Kredit Kendaraan Bermotor merupakan solusi yang terbaik dalam memiliki kendaraan dengan cara mengangsur atau mencicil baik itu kendaraan baru ataupun bekas (Siti Qomah, 2021). Proses penilaian kelayakan kredit adalah tantangan yang cukup kompleks karena melibatkan sejumlah faktor penilaian yang banyak (Fernando et al., 2023).

Pada penelitian sebelumnya menerapkan algoritma K-Nearest Neighbor dengan tingkat akurasi sebesar 81,82% (Supriana et al., 2019). Model Naïve Bayes Classifier memberikan hasil lebih rendah dari KNN dengan akurasi 53,76% dikarenakan tingkat akurasi kevalidannya tidak terlalu tinggi (Dewi, 2019). Namun, dalam penelitian sebelumnya, belum ada pembahasan dengan melakukan prediksi dengan dataset yang tidak memiliki label. Sehingga pembuatan label harus dilakukan secara manual (Givari et al., 2022). Selain itu, keterbatasan data juga dapat menghambat kemampuan model untuk mengenali dan memahami pola serta karakteristik dalam data tersebut (Aulia Akbar, 2020).

Berdasarkan dari permasalahan diatas, maka dilakukanlah penelitian berupa analisis guna memprediksi apakah nasabah dalam masa kredit sini memiliki pembayaran yang lancar atau terhambat (macet). Penelitian ini akan mengelompokkan kategori pembayaran nasabah berdasarakan fitur-fiturnya dengan melakukan pengelompokkan dengan algoritma *K-Means Clustering* dan dilanjutkan klasifikasi dengan algoritma Random Forest. Penelitian ini bertujuan untuk mengembangkan model yang dapat memprediksi kelancaran pembayaran kredit kendaraan bermotor.

2. LANDASAN TEORI

2.1 Kredit Kendaraan Bermotor

Secara definisi, KKB diartikan sebagai fasilitas kredit yang diberikan oleh bank atau perusahaan pembiayaan lainnya, untuk pembelian kendaraan baru maupun bekas (Aulia Akbar, 2020). Perjanjian kredit adalah kesepakatan pinjaman antara kreditur (bank/leasing) dan nasabah.

2.2 Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Ariadi, 2020).

P-ISSN: 1411-3635
E-ISSN: 2962-5645
https://doi.org/10.37817/Tekinfo.v26i1

2.3 Semi Supervised Learning

Semi-supervised learning adalah metode yang dapat dapat memanfaatkan data yang tidak berlabel (unlabeled data) untuk membantu pembentukan model dalam mempelajari pola data. Teknik ini merupakan kombinasi antara supervised learning dan *unsupervised learning*. *Semi-supervised learning* dibagi menjadi dua jenis, yaitu semi-supervised classification dan *semi-supervised clustering* (Fiarni et al., 2017).

2.4 *K-Means Clustering*

K-Means Clustering merupakan salah satu metode pengelompokan data dalam bentuk kelompok ataupun kluster (Fitriani et al., 2023). Pemanfaatan metode pengelompokan data, seperti yang terlihat dalam penerapan algoritma *Clustering*, telah diimplementasikan dalam berbagai konteks.

2.5 *Random Forest*

Random forest merupakan suatu teknik klasifikasi yang menggabungkan beberapa pohon keputusan dengan cara memilih data dan variabel secara acak. Metode *Random Forest* menghasilkan satu set pohon acak. Kelas yang dihasilkan berasal dari proses klasifikasi yang dipilih dari kelas yang paling banyak (modus) yang dihasilkan oleh pohon keputusan yang ada (Rachmi, 2020).

3. METODOLOGI

Metodologi penelitian ini terdiri dari tahapan pengumpulan data, pre-processing, pembuatan label data, split data, klasifikasi data dan evaluasi model, ditunjukkan pada Gambar 1.

Gambar 1 Tahap Penelitian

1. Pengumpulan Data

Pengumpulan data awal berupa data sekunder yang beruapa data kuantitatif dan kualitatif dari dataset kredit kendaraan bermotor nasabah Adira Finance bulan Oktober tahun 2023 yang berjumlah 573 record dengan 11 atribut.

2. Data *Preprocessing*

Data *pre-processing* adalah proses mengubah data mentah menjadi data yang siap untuk dianalisis. Proses ini penting untuk memastikan bahwa data dapat dianalisis dengan benar dan menghasilkan hasil yang akurat (Givari et al., 2022).

3. Eksplorasi Data

Analisis data eksploratif (EDA) adalah proses eksplorasi data secara visual dan statistik untuk memahami karakteristik data dan mengidentifikasi pola atau tren yang mungkin ada. Tujuan dari EDA adalah untuk mendapatkan pemahaman yang lebih baik tentang data sebelum menerapkan metode analisis data yang lebih formal (Fadilah, 2018).

4. Pembuatan Label Data

Pembuatan label data menggunakan *K-Means* adalah proses pengelompokan data ke dalam beberapa kelompok (kluster) berdasarkan kedekatannya. Proses ini meliputi :

(a) Menentukan jumlah kluster yang diinginkan. Jumlah kluster dapat ditentukan dengan menggunakan berbagai metode, seperti metode *Elbow*.

(b) Menentukan pusat kluster awal.

(c) Menghitung jarak *Euclidean* setiap data ke pusat kluster.

(d) Mengelompokkan data ke dalam kluster berdasarkan jarak terdekat.

(e) Menghitung pusat kluster baru.

https://doi.org/10.37817/Tekinfo.v26i1
P-ISSN: 1411-3635
E-ISSN: 2962-5645

(f) Mengulang langkah 2 hingga 5 hingga tidak ada perubahan pada pengelompokan data.

5. Split Data

Pembagian data merupakan kegiatan membagi data menjadi beberapa bagian. Keandalan dan kemampuan generalisasi model dipastikan dengan memilih rasio 80:20 untuk pembagian data latih dan uji, yang menjaga keseimbangan antara data pelatihan dan pengujian (Ricky, 2020).

6. Klasifikasi Data

Proses pengkalsifikasian data menggunakan *Random Forest*, berikut adalah langkah-langkah dalam melakukan klasifikasi dengan metode *Random Forest* (Zhalsabilla, 2022).

(a) Tahap pertama adalah mengambil sampel secara acak dari data latih (*train*) dengan memungkinkan data yang sama muncul kembali.

(b) Tahap selanjutnya adalah pembentukan pohon keputusan secara acak, yang dilakukan dengan memilih fitur secara acak di setiap simpul.

(c) Selanjutnya, menghitung nilai *entrophy* dari seluruh fitur yang digunakan untuk membangun pohon keputusan.

(d) Setelah menghitung nilai entropi dari masing-masing fitur, dilakukan perhitungan nilai perolehan *informasi gain* dari nilai entropi yang didapatkan. Fitur dengan nilai perolehan *informasi gain* paling besar akan dijadikan simpul akar pada pohon yang dibangun.

(e) Selanjutnya, dilakukan pemisahan (*split*) hingga semua simpul berbentuk simpul daun atau tidak ada pemisahan yang dapat dilakukan lagi.

(f) Jika belum selesai membangun pohon sebanyak *k* pohon, ulangi langkah 1-5. Jika pohon yang dibangun sudah mencapai *k* pohon, dilakukan pemungutan suara terbanyak untuk prediksi akhir.

7. Evaluasi Model

Evaluasi model dilakukan dengan melihat seberapa baik model dapat memprediksi kelas positif dan negatif, dengan menggunakan *Confusion Matrix* (Pahlevi et al., 2023) ditunjukkan pada Tabel 1.

Tabel 1 *Confusion Matrix*, k = 2

	Predicted Positive	**Predicted Negative**
Actual Positive	*True Positive (TP)*	*False Negative (FN)*
Actual Negative	*False Positive (FP)*	*True Negative (TN)*

$$Accuracy = \frac{TP+TN}{Total} \quad (1)$$

$$Precision = \frac{TP}{FP+TP} \quad (2)$$

$$Recall = \frac{TP}{FN+TP} \quad (3)$$

P-ISSN: 1411-3635
E-ISSN: 2962-5645
https://doi.org/10.37817/Tekinfo.v26i1

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} (4)$$

4. HASIL DAN PEMBAHASAN

4.1 Data *Preprocessing*

Langkah-langkah dalam tahap *preprocessing* data adalah sebagai berikut :

1. Pembersihan data adalah proses untuk menghilangkan data yang tidak valid, tidak lengkap, atau tidak relevan. Pada tahap *preprocessing*, salah satu hal yang dilakukan adalah mencari *missing value* terdapat pada Gambar 2.

```
In [4]: df.isnull().sum()

Out[4]: NO                      0
        PENDIDIKAN              0
        USIA                    0
        JENIS_KELAMIN           0
        PEKERJAAN               0
        PENGHASILAN_PER_BULAN   0
        JUMLAH_TANGGUNGAN       0
        TOTAL_HARGA             0
        DP                      0
        POKOK_HUTANG            0
        JUMLAH_BUNGA            0
```

Gambar 2 Jumlah *Missing Value*

2. Transformasi data adalah proses untuk mengubah format atau nilai data agar lebih mudah dipahami dan dianalisis.
   (a) Menambah variabel baru dan mengubah nama variabel pada dataset yang terdapat pada Gambar 3.

Gambar 3 Menambah Variabel Baru

Penambahan variabel CICILAN_NASABAH, berguna untuk membaca dataset lebih mudah. Dikarenakan dengan adanya variabel tersebut menunjukkan cicilan yang harus dibayar oleh nasabah secara keseluruhan.

Gambar 4 Perubahan Nama Variabel

Pada Gambar 4 dilakukan perubahan nama variabel dari ‘TOTAL_HARGA’ menjadi ‘HARGA_UNIT’ memudahkan pembacaan dataset. Dikarenakan ‘TOTAL_HARGA’ memiliki pengertian yang cukup luas, maka dari itu nama ‘HARGA_UNIT’ memiliki pengertian yang sesuai karena menunjukkan harga unit kendaraan.

(b) Mengubah tipe data menjadi integer yang ditunjukkan pada Gambar 5.

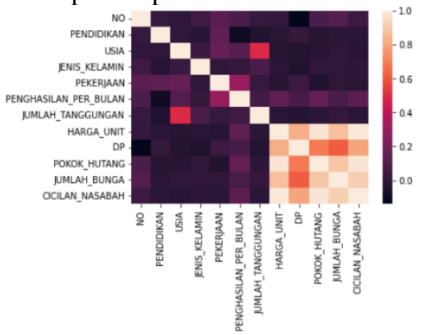
```
In [10]: data.dtypes

Out[10]: NO                      int64
         PENDIDIKAN              int32
         USIA                    int64
         JENIS_KELAMIN           int32
         PEKERJAAN               int32
         PENGHASILAN_PER_BULAN   int64
         JUMLAH_TANGGUNGAN       int64
         HARGA_UNIT              int64
         DP                      int64
         POKOK_HUTANG            int64
         JUMLAH_BUNGA            int64
         CICILAN_NASABAH         int64
```

Gambar 5 Tipe Data

(c) Merubah kategori pada variabel JENIS_KELAMIN dan PEKERJAAN. Agar memudahkan proses analisis data maka dilakukan *labelling encod*.

3. Proses pemilihan variabel, atau seleksi fitur, dilakukan untuk menentukan variabel-variabel yang akan digunakan dalam analisis penelitian. Dari 12 variabel awal, beberapa dipilih berdasarkan visualisasi heatmap yang ditampilkan pada Gambar 6.



Gambar 6 Hasil Analisa Korelasi antar Variabel

https://doi.org/10.37817/Tekinfo.v26i1
P-ISSN: 1411-3635
E-ISSN: 2962-5645

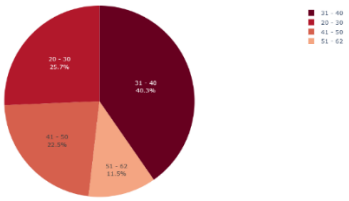
4. Normalisasi data adalah proses untuk mengubah data ke dalam skala yang sama. Hal ini penting agar data dapat dibandingkan dengan benar. Penskalaan data menggunakan normalisasi min-max, yang mengubah nilai data menjadi rentang 0 sampai 1, dilakukan dalam penelitian ini, dan visualisasi dari metode ini dapat ditemukan pada Gambar 7.

Gambar 7 Normalisasi Data dengan *Min-Max*

4.2 Eksplorasi Data

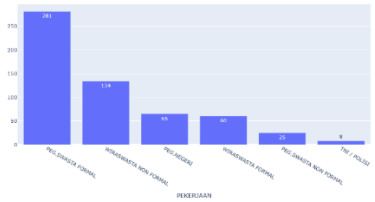
Berikut analisa dari masing-masing variabel.

1. Usia, pada Gambar 8 merupakan grafik rentang usia terbanyak yang melakukan transaksi kredit kendaraan bermotor.



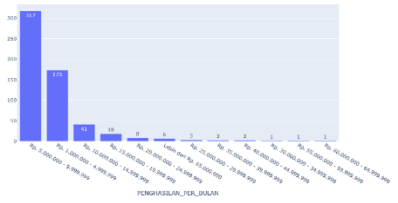
Gambar 8 Analisa Variabel Usia

2. Pekerjaan, dalam grafik yang ditunjukkan pada Gambar 9. Pekerjaan yang paling mendominasi yaitu Pegawai Swasta Formal.



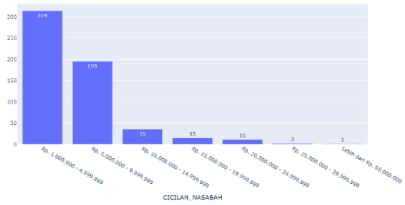
Gambar 9 Analisa Variabel Pekerjaan

3. Penghasilan per Bulan, dalam grafik yang ditunjukkan pada Gambar 10 menampilkan rentang penghasilan per bulan para nasabah.



Gambar 10 Analisa Variabel Penghasilan Per Bulan

4. Cicilan nasabah, grafik yang ditunjukkan pada Gambar 11 menampilkan rentang cicilan nasabah yang harus dibayarkan.



Gambar 11 Analisa Variabel Cicilan Nasabah

4.3 Pembuatan Label Data

1. Menentukan jumlah kluster yang diinginkan. Dalam penelitian ini, metode elbow digunakan untuk menentukan *k*. Serta penggunaan persamaan SSE.
2. Pusat kluster awal dapat ditentukan secara acak atau dengan menggunakan metode tertentu. Dari hasil grafik *elbow* dan perhitungan dengan SSE diatas,maka pusat kluster awal yaitu 2.
3. Menghitung jarak *Euclidean* setiap data ke pusat kluster. Berikut datanya yang terdapat pada Tabel 2.

Tabel 2 Sampel Data yang telah di Normalisasi

USIA	PEKERJAAN	PENGHASILAN_PER_BULAN	JUMLAH_TANGGUNGAN	HARGA_UNIT	CICILAN_NASABAH
0.545	0	0.026	0.333	0.122	0.12
0.25	0.2	0.036	0.167	0.017	0.014
0.159	0.2	0.007	0	0.119	0.105
0.363	0.2	0.058	0.167	0.263	0.252
0.704	0.2	0.019	0.333	0.042	0.043

4. Setelah didapatkan data yang telah dinormalisasi, maka dilakukan proses klusterisasi dengan $k = 2$ yaitu *cluster* 1 (C1) dan *cluster* 2 (C2) dengan algorima *K-Means*.

P-ISSN: 1411-3635
E-ISSN: 2962-5645
https://doi.org/10.37817/Tekinfo.v26i1

5. Hasil perhitungan jarak *Euclidean* dengan *centroid* pada iterasi pertama dan kedua, kemudian dikelompokkan data sesuai dengan hasil *cluster*nya yang ditunjukkan pada Tabel 3.

Tabel 3 Pengelompokkan Data pada Iterasi Pertama

USIA	PEKERJAAN	PENGHASILAN_PER_BULAN	JUMLAH_TANGGUNGAN	HARGA_UNIT	CICILAN_NASABAH	C1	C2	CLUSTER
0.545	0	0.026	0.333	0.122	0.12	0.14	0.41	1
0.25	0.2	0.036	0.167	0.017	0.014	0.44	0.37	2
0.159	0.2	0.007	0	0.119	0.105	0.58	0.152	2
0.363	0.2	0.056	0.167	0.253	0.292	0.41	0.22	2
0.704	0.2	0.019	0.333	0.042	0.043	0.14	0.52	1

Selanjutnya, dihitung nilai rata-rata di setiap *centroidnya* dengan menjumlahkan anggota masing-masing *cluster* dan kemudian dibagi dengan jumlah anggota yang memiliki *cluster* yang sama. Ditunjukkan dengan Tabel 4.

Tabel 4 Nilai Rata-Rata Centroid Iterasi Pertama dan Kedua

	USIA	PEKERJAAN	PENGHASILAN_PER_BUL	JUMLAH_TANGGUN	HARGA_UNIT	CICILAN_NASABAH
C1	0.625	0.1	0.023	0.333	0.082	0.0815
C2	0.26	0.2	0.034	0.11	0.133	0.124

	USIA	PEKERJAAN	PENGHASILAN_PER_BUL	JUMLAH_TANGGUN	HARGA_UNIT	CICILAN_NASABAH
C1	0.625	0.1	0.023	0.333	0.082	0.0815
C2	0.26	0.2	0.034	0.11	0.133	0.124

6. Jika hasil rata-rata iterasi pertama dan kedua belum sama, maka lakukan perhitungan jarak *Euclidean* kembali hingga hasil rata-rata tidak mengalami perubahan.

4.4 Split Data

Pembagian data ini dilakukan dengan proporsi 80% untuk data training dan 20% untuk data testing.

4.5 Klasifikasi Data

Setelah membagi data menjadi data latih dan data uji, langkah berikutnya adalah melakukan analisis klasifikasi menggunakan metode *Random Forest*. Dalam pembangunan setiap pohon, data digunakan dengan teknik *bootstrap*, dimana sampel data dipilih secara acak dari set data asli dan dapat terpilih lebih dari satu kali. Didapatkan 4 variabel independen yang memiliki nilai korelasi yang tinggi dengan variabel target yaitu variabel 'PEKERJAAN','PENGHASILAN_PER_BULAN','JUMLAH_TANGGUNGAN dan 'CICILAN_NASABAH'. Sebelum dilakukan perhitungan dengan *bootstrap* set maka dilakukan perhitungan nilai *entrophy* dari 7 sampel data berikut perhitungannya.

$$E\,|S| = -\frac{3}{7}log2\left(\frac{3}{7}\right) - \frac{4}{7}log2\left(\frac{4}{7}\right) = 0.97 \quad (5)$$

Dari hasil perhitungan *entrophy* semesta diatas berguna untuk menghitung *information gain* untuk membuat suatu pohon Keputusan (Permataning Tyas et al., 2022).

Dataset *Bootstrap* 1

Tabel 5 Dataset *Bootstrap* 1

PKJ	PPB	KET
>1	<= 6 JUTA	0
<=1	> 6 JUTA	0
<=1	<= 6 JUTA	0
>1	<= 6 JUTA	1
>1	> 6 JUTA	1
>1	<= 6 JUTA	1
>1	<= 6 JUTA	0

Variabel yang dipilih pada dataset ini yaitu PEKERJAAN dan PENGHASILAN_PER_BULAN. Berikut perhitungan *entrophy* dari masing-masing variabel.

a. PEKERJAAN

$$\text{PEKERJAAN} > 1\ [3,5] = -\frac{3}{5}log2\left(\frac{3}{5}\right) - \frac{2}{5}log2\left(\frac{2}{5}\right) = 0.95 \quad (6)$$

$$\text{PEKERJAAN} \leq 1\ [0,2] = -\frac{0}{2}log2\left(\frac{0}{2}\right) - \frac{2}{2}log2\left(\frac{2}{2}\right) = 0 \quad (7)$$

P-ISSN: 1411-3635
E-ISSN: 2962-5645

$$E(T, PEKERJAAN) = \left(\frac{5}{7} \times 0.95\right) + \left(\frac{2}{7} \times 0\right) = 0.67 \quad (8)$$

$$Gain(PEKERJAAN) = 0.97 - 0.67 = 0.3 \quad (9)$$

b. PENGHASILAN_PER_BULAN

$$PPB > 6\ JUTA\ [1,2] = -\frac{1}{2}log2\left(\frac{1}{2}\right) - \frac{1}{2}log2\left(\frac{1}{2}\right) = 1 \quad (10)$$

$$PPB \leq 6\ JUTA\ [2,5] = -\frac{2}{5}log2\left(\frac{2}{5}\right) - \frac{3}{5}log2\left(\frac{3}{5}\right) = 0.95 \quad (11)$$

$$E(T, PPB) = \left(\frac{2}{7} \times 1\right) + \left(\frac{5}{7} \times 0.95\right) = 0.95 \quad (12)$$

$$Gain(PPB) = 0.97 - 0.95 = 0.02 \quad (13)$$

Berdasarkan nilai *information gain* variabel diatas, maka variabel PEKERJAAN merupakan node pertama pada pohon. Dan lakukan hingga semua variabel dihitung sehingga terbentuk *Random Forest* yang ditunjukkan pada Gambar 12.

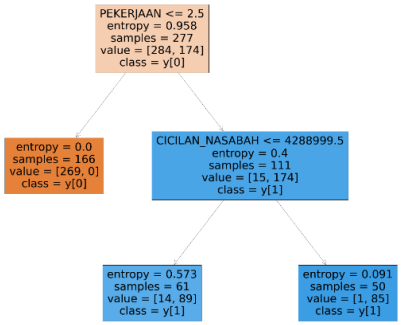
Gambar 12 Hasil Pohon *Random Forest*

Tuning parameter juga dilakukan untuk mendapatkan model dengan kinerja terbaik. Dengan memanfaatkan scikit-learn, berikut hasil dari tuning parameter ditampilkan pada Tabel 6.

Tabel 6 Parameter *Random Forest*

Parameter	Values
n_estimators	100
criterion	*entrophy* dan *information gain*
max_depth	2
max_features	sqrt
bootstrap	True

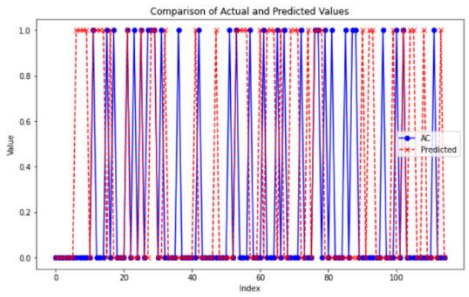
Dari hasil tuning parameter diatas didapatkan skor akurasi sebesar 0.98. Berikut Gambar 13 dibawah ini merupakan contoh dari pohon *Random Forest*. Pada salah satu pohon keputusan *Random Forest* terdapat dua kali percabangan, sesuaia dengan parameter yang sudah di tuning dengan nilai *max_depth* sebanyak 2 kali.



Gambar 13 Contoh Pohon *Random Forest*

Pada Gambar 14, merupakan perbandingan dari kedua variabel. Perbedaan ini bisa disebabkan oleh berbagai faktor, termasuk perbedaan dalam metode pengukuran atau prediksi. Variabel 'PREDICTED' kemungkinan besar berasal dari model prediksi yang menggunakan algoritma atau asumsi tertentu. Akurasi model, kualitas data pelatihan, dan parameter yang digunakan dalam model prediksi dapat mempengaruhi hasil prediksi dan menyebabkan perbedaan dengan data sebenarnya.

P-ISSN: 1411-3635
E-ISSN: 2962-5645
https://doi.org/10.37817/Tekinfo.v26i1



Gambar 14 Perbandingan *Data Actual* dengan *Data Predicted*

4.6 Evaluasi Model

Dengan mengevaluasi model digunakan nilai akurasi precission, recall, dan f1 score.

```
Confusion Matrix:
[[74  2]
 [ 0 39]]
```

Gambar 15 *Confusion Matrix* Data Setelah Diklasifikasi

Berikut perhitungannya.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (14)$$

$$= \frac{(74+39)}{(74+39+0+2)} \quad (15)$$

$$= \frac{113}{115} = 0.98 \quad (16)$$

$$Precission = \frac{TP}{(TP+FP)} \quad (17)$$

$$= \frac{74}{(74+0)} \quad (18)$$

$$= \frac{74}{74} = 1 \quad (19)$$

$$Recall = \frac{TP}{(FN+TP)} \quad (20)$$

$$= \frac{74}{(2+74)} \quad (21)$$

$$= \frac{74}{76} = 0.97 \quad (22)$$

$$F1\ Score = 2 \times \frac{Precission \times Recall}{Precission+Recall} \quad (23)$$

$$= 2 \times \frac{1 \times 0.97}{1+0.97} \quad (24)$$

$$= 2 \times \frac{0.97}{1.97} = 0.98 \quad (25)$$

Setelah analisis dilakukan, model yang telah dibuat disimpan seperti yang ditunjukkan pada Gambar 16. Model ini disimpan dalam format pickle.

```
import pickle

# Menyimpan model
with open('kredit-lancar-macet-4var.pkl', 'wb') as file:
    pickle.dump(rf_model, file)
```

Gambar 16 Menyimpan Model yang telah Dianalisa

5. KESIMPULAN

Dapat disimpulkan bahwa model *K-Means Clustering* membuat label untuk data yang belum memiliki label dengan tujuan menghasilkan kategori yang dapat digunakan untuk klasifikasi lebih lanjut. Hasil dari proses ini adalah label LANCAR dan MACET. Model *Random Forest* melakukan pemrediksian kelancaran pembayaran kredit kendaraan bermotor nasabah dengan membuat dataset *bootstrap*, kemudian menjalankan prosedur *decision tree* untuk mencari keputusan pada setiap data *bootstrap*. Menggunakan 4 fitur yang berpengaruh dengan variabel target yaitu ‘PEKERJAAN’,’PENGHASILAN_PER_BULAN’,’JUMLAH_TANGGUNGAN’ dan ‘CICILAN_NASABAH dengan menghitung *entrophy* dari semua variabel dan menentukan *node* pertama dengan bantuan *information gain*. Dari analisis Random Forest dengan menggunakan beberapa parameter yaitu *n_estimator* = 100, *max_depth* = 2 dan *max_features* = ‘sqrt’ diperoleh untuk memprediksi kelancaran pembayaran kredit kendaraan bermotor nasabah memiliki nilai akurasi sebesar 98%. Dari kesimpulan diatas, terdapat saran agar penelitian ini lebih baik yaitu dengan menggunakan dataset yang lebih banyak karena mempengaruhi pola dan hasil keakuratan yang didapatkan. Dan untuk penelitian berikutnya, dapat dipertimbangkan penggunaan pendekatan lain untuk mengevaluasi kinerja dari proses klasifikasi yang dihasilkan oleh algoritma yang digunakan.

6. UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih kepada Universitas Esa Unggul atas dukungan yang diberikan untuk penelitian ini. Ucapan terima kasih juga ditujukan kepada semua pihak yang terlibat termasuk dosen pembimbing dan PT Adira Dinamika Multi Finance Tbk yang telah mengizinkan untuk memberikan dataset untuk dilakukan analisa.

DAFTAR PUSTAKA

Aulia Akbar. (2020). *Pengertian Kredit Kendaraan Bermotor dan Syarat Pengajuan*. Lifepal. https://lifepal.co.id/media/kredit-kendaraan-bermotor/

Dewi, S. (2019). Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan. *Jurnal Khatulistiwa Informatika*, *7*(1), 59–65. https://doi.org/10.31294/jki.v7i1.5744

Fernando, S. D., Purnamasari, S. D., Syaputra, H., Adha, N., & Saputri, O. (2023). *Prediksi Kredit Sepeda Motor pada Showroom Astra Motor Sako Palembang dengan Metode Naive Bayes*. *7*(1), 17–23. http://e-journal.unipma.ac.id/index.php/doubleclick

Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *Nuansa Informatika*, *16*(1), 141–149. https://doi.org/10.25134/nuansa.v16i1.5406

Pahlevi, O.-, Amrin, A.-, & Handrianto, Y.-. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. In *Jurnal Infortech* (Vol. 5, Issue 1, pp. 71–76). https://doi.org/10.31294/infortech.v5i1.15829

Rachmi, A. N. (2020). *Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn*. 1–101.

Ricky, L. et al. (2020). *Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing*.

Rusli, A., Tambing, B., Universitas, D., Palopo, M., Universitas, M., & Palopo, M. (2021). *Faktor-faktor yang memengaruhi permintaan kredit motor di fifgroup pos toraja kabupaten toraja utara*.

Siti Qomah, N. (2021). Klasifikasi Pengelolaan Kredit Menggunakan Metode Naïve Bayes. In *Jurnal Data Science & Informatika (Jdsi)* (Vol. 2, Issue 1, pp. 35–40).

Supriana, I. W., Raharja, M. A., & Gunawan, P. W. (2019). *Sistem Informasi Prediksi Penilaian Kredit Perbankan Menggunakan Algoritma K-Nearest Neighbor Classification*. *8*(1), 44–54.

Zhalsabilla, S. (2022). *INTEGRASI ALGORITMA KLUSTERISASI K-MEANS DAN KLASIFIKASI NAIVE BAYES PADA PENGELOMPOKKAN JUMLAH SEKOLAH, MURID, DAN GURU DI PROVINSI JAWA BARAT*. 1–23.