

## KLASIFIKASI STATUS PENGAJUAN KPR RUMAH SEDERHANA MENGGUNAKAN ALGORITMA RANDOM FOREST

Raksi Andika<sup>1</sup>, Lastri Widya Astuti<sup>2\*</sup>, Muhammad Haviz Irfani<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, \*Penulis Koresponden

e-mail: [1raksiandika@gmail.com](mailto:1raksiandika@gmail.com), [2lastriwidya@uigm.ac.id](mailto:2lastriwidya@uigm.ac.id), [3m.haviz@uigm.ac.id](mailto:3m.haviz@uigm.ac.id)

### ABSTRACT

The property industry has a strategic role in supporting economic growth, especially through Home Ownership Credit (KPR) financing which enables people to own housing. In big cities in Indonesia, various property developers are innovating to meet the demand for modern, affordable housing, but some of them are facing obstacles in the KPR feasibility evaluation process. Developers experience difficulties in managing the eligibility selection process, because inaccurate analysis can make consumers disappointed after paying the down payment (DP) if their KPR application is ultimately rejected. This research aims to classify the status of consumer mortgage applications, with historical data of 969 samples covering previous projects, which contains features such as consumer income, number of dependents, employment, credit history, and type of property being applied for, as well as 1 status feature which states the results of the application. The Random Forest algorithm is used to produce a KPR status classification model into three main categories: Contract, Reject Bank, and Reject Customer. Based on model evaluation using the average multiclass matrix, this model achieved 99% accuracy, with an average precision, recall and f1-score of 98%, which shows this model is very effective in identifying and classifying the three categories of KPR status.

**Keywords:** home ownership credit, classification, random forest, multiclass, consumer

### INTISARI

Industri properti memiliki peran strategis dalam mendukung pertumbuhan ekonomi, terutama melalui pembiayaan Kredit Pemilikan Rumah (KPR) yang memungkinkan masyarakat memiliki hunian. Kota-kota besar di Indonesia, berbagai pengembang properti berinovasi untuk memenuhi permintaan akan hunian modern yang terjangkau, namun beberapa diantaranya menghadapi kendala dalam proses evaluasi kelayakan KPR. Pengembang mengalami kesulitan dalam mengelola proses seleksi kelayakan, karena analisis yang kurang akurat dapat membuat konsumen kecewa setelah pembayaran *down payment* (DP) seandainya pengajuan KPR mereka akhirnya ditolak. Penelitian ini bertujuan untuk mengklasifikasikan status pengajuan KPR konsumen, dengan data historis sebanyak 969 sampel yang mencakup proyek-proyek sebelumnya, yang berisi fitur seperti pendapatan konsumen, jumlah tanggungan, pekerjaan, riwayat kredit, dan jenis properti yang diajukan, serta 1 fitur status yang menyatakan hasil pengajuannya. Algoritma Random Forest digunakan untuk menghasilkan model klasifikasi status KPR ke dalam tiga kategori utama: Akad, Reject Bank, dan Reject Customer. Berdasarkan evaluasi model menggunakan rata-rata matriks multiclass, model ini mencapai akurasi 99%, dengan rata-rata presisi, recall, dan f1-score sebesar 98%, yang menunjukkan model ini sangat efektif dalam mengidentifikasi serta mengklasifikasikan ketiga kategori status KPR.

**Kata kunci:** kredit pemilikan rumah, klasifikasi, *random forest*, *multiclass*, konsumen

### 1. PENDAHULUAN

Industri properti memiliki peran penting dalam pertumbuhan ekonomi, terutama melalui pembiayaan Kredit Pemilikan Rumah (KPR) yang mempermudah masyarakat memiliki hunian. Di Indonesia, banyak pengembang properti yang terus berinovasi untuk menjawab kebutuhan pasar akan perumahan modern yang terjangkau. Salah satu pengembang yang berfokus di kota besar mengembangkan proyek hunian baru setelah keberhasilan proyek-proyek sebelumnya. Dengan mengedepankan transparansi, kepercayaan, dan kerjasama yang kuat dengan para mitra serta debitur, perusahaan ini menghadapi tantangan dalam pengelolaan proses KPR di tengah dinamika ekonomi dan regulasi yang berkembang.

Keputusan terkait status pengajuan KPR memerlukan penilaian cermat terhadap sejumlah faktor, termasuk data keuangan dan kredit pelanggan (Satria, 2018). Faktor-faktor tersebut menjadi penentu penting dalam menilai kelayakan seorang individu atau peminjam dalam memperoleh KPR. Perusahaan pengembang properti sering kali mengalami analisa yang salah diawal, sehingga banyak konsumen yang telah membayar *down paymen* (DP) merasa kecewa karena pengajuan KPR ditolak (Takawira & Muteba Mwamba, 2022), maka untuk penilaian kelayakan kredit yang harus diperhatikan adalah kesalahan dalam perhitungan pendapatan, hutang peminjam, ketidakseimbangan resiko portofolio, ketidakseimbangan struktur pinjaman, tidak memperhitungkan resiko pasar, dan ketidakpatuhan terhadap prosedur internal, maka pemanfaatan teknologi klasifikasi seperti algoritma Random Forest dapat menjadi langkah inovatif untuk meningkatkan transparansi dan ketepatan dalam mengambil keputusan.

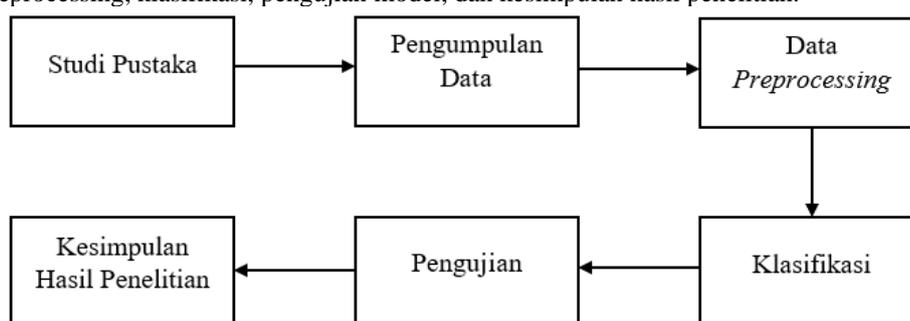
Algoritma Random Forest, yang diciptakan oleh J. Ross Quinlan, diberi nama demikian karena berkembang dari metode ID3 dalam pembuatan pohon keputusan. Menurut Larose (2013) dalam (Zailani & Hanun, 2020:8) Random Forest termasuk algoritma yang sesuai untuk tugas klasifikasi dalam bidang *machine learning* dan data mining. Algoritma Random Forest juga salah satu metode klasifikasi yang andal dalam *machine learning* karena mampu mengombinasikan banyak pohon keputusan guna menghasilkan prediksi yang lebih akurat dan stabil. (Primajaya & Sari, 2018) menyatakan bahwa algoritma ini sangat cocok untuk pengolahan data yang kompleks dan berskala besar, karena dapat menjaga akurasi meskipun menghadapi berbagai jenis data.

Di bidang keuangan, seperti pada evaluasi pengajuan KPR, Random Forest digunakan untuk mengelompokkan data berdasarkan atribut tertentu dalam analisis risiko kredit (Lestari & Sirodj, 2022; Pahlevi & Handrianto, 2023; Takawira & Muteba Mwamba, 2022). (Primajaya & Sari, 2018) juga menggarisbawahi bahwa teknik ini mampu mengurangi *overfitting* melalui pendekatan ensemble, menggabungkan beberapa pohon keputusan acak untuk menghasilkan prediksi yang lebih konsisten (Suci Amaliah et al., 2022). Algoritma ini semakin banyak digunakan dalam penelitian terbaru untuk membantu pengambilan keputusan berbasis data di berbagai sektor seperti (Aini et al., 2024; Irma Purnamasari & Ali, 2024; Madaerdo Sotarjua et al., 2022; Suci Amaliah et al., 2022)

Penelitian ini akan fokus pada pengklasifikasian status pengajuan KPR menjadi tiga kategori yaitu Akad, Reject Bank dan Reject Customer. Penelitian ini akan menerapkan algoritma Random Forest sebagai metode klasifikasi utama untuk mengolah data status pengajuan KPR. Penelitian ini difokuskan pada pengelompokan status pengajuan KPR ke dalam tiga kategori: Akad, Reject Bank, dan Reject Customer. Untuk itu, algoritma Random Forest akan digunakan sebagai metode utama dalam klasifikasi data status pengajuan KPR.

## 2. METODE PENELITIAN

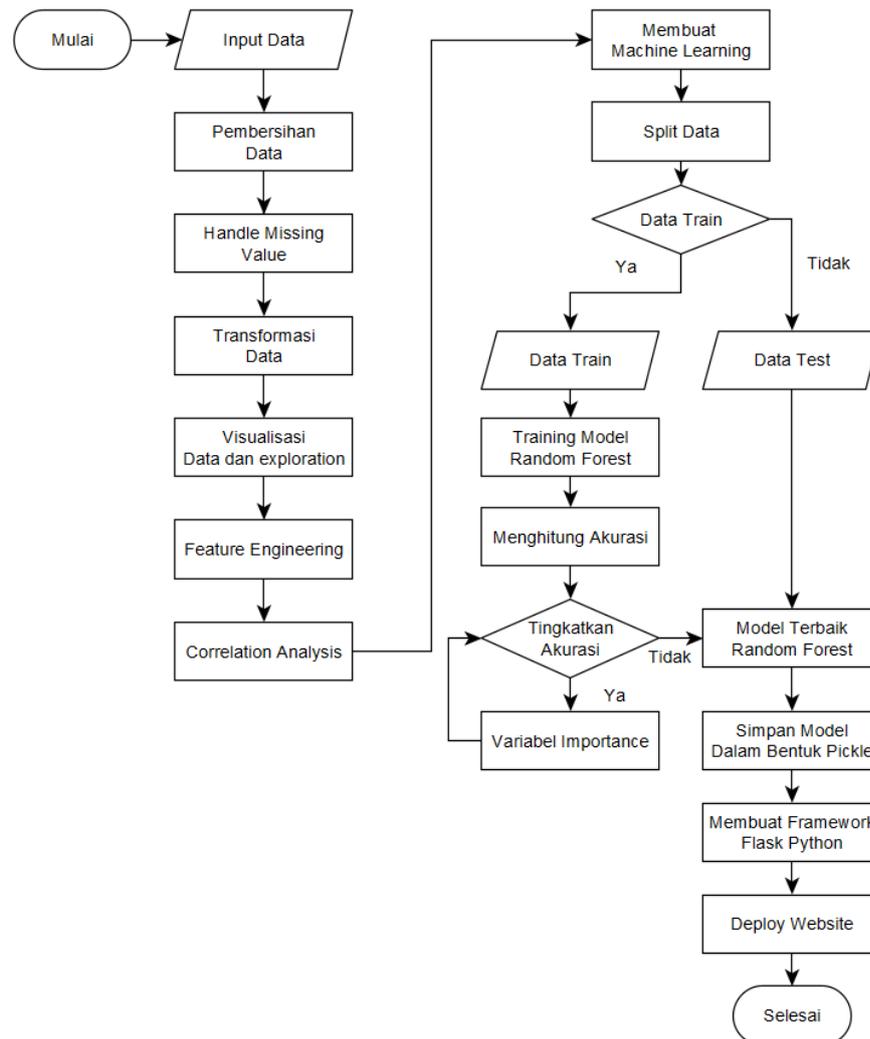
Alur penelitian ini terlihat pada Gambar 1, terdapat beberapa tahap alur penelitian yaitu studi ustaka, pengumpulan data, data preprocessing, klasifikasi, pengujian model, dan kesimpulan hasil penelitian.



**Gambar 1.** Alur Tahapan Penelitian Alur Penelitian

Setelah proses pengumpulan data, selanjutnya dilakukan pengolahan awal data dengan stuktur variabel-variabel dan dilakukan pembersihan data untuk proses selanjutnya. Data yang diperoleh berjumlah 969 data historis, 9 (sembilan) buah atribut yang akan menjadi kelas pada saat diektrak. Hasil preprocessing diperoleh 35 record missing dari atribut feature bank. Selanjutnya dilakukan pembersihan data sekaligus penghapusan kolom yang tidak diperlukan dari data yang sudah ditabulasi, penggantian nama kolom, melengkapi nilai data yang hilang, dan penyesuaian tipe data. Kemudian membuat tipe dictionary untuk masing-masing kategori ke dalam nilai numerik, serta menggunakan fungsi `map()` untuk penggantian nilai kategori. Selain itu, dilakukan label encoder untuk mengubah tipe data string menjadi integer, visualisasi data menggunakan plot bar dan heatmap untuk penyajian data agar mudah dipahami karakteristiknya, memperoleh perhitungan hasil korelasi antar fitur (kelas), agar dapat

memperjelas hubungan antara fitur dengan target yang dituju.



**Gambar 2.** Alur Proses Random Forest dengan *Hyperparameter Tuning*

Aliran proses yang terjadi dengan menggunakan algoritma *random forest* untuk parameter-parameter yang akan dilakukan proses tuning berkali-kali percobaan (Gambar 2) dalam mendapatkan model terbaik. *Machine learning* dengan Bahasa Python tidak mendukung *multiclass* atau tidak bisa digunakan jika tipe data pada fitur (x) dan label (y) berbeda jenisnya, sehingga perlu dilakukan label *encoder*. Label *encoder* adalah proses mengubah tipe data string menjadi tipe data integer. Pada data rekam nasabah yang telah dikumpulkan, terdapat beberapa informasi yang berbentuk string, sehingga perlu dilakukan label *encoder*. Berikut adalah tabel yang menjelaskan tentang jenis data awal dari konsumen yang telah terkumpul.

**Tabel 1.** Label *Encoder* untuk Semua Fitur

Nama Variabel	Tipe Data Awal	Hasil Penyesuaian
Harga_Jual	<i>Integer</i>	<i>Integer</i>
Usia	<i>Integer</i>	<i>Integer</i>
Status_Konsumen	<i>String</i>	<i>Integer</i>
Jumlah_Anak	<i>String</i>	<i>Integer</i>
Status_Pekerjaan	<i>String</i>	<i>Integer</i>
TotalGaji	<i>Integer</i>	<i>Integer</i>
CicilanBln	<i>Integer</i>	<i>Integer</i>
Hasil_Slik	<i>String</i>	<i>Integer</i>
Bank	<i>String</i>	<i>Integer</i>
Status_Berkas	<i>String</i>	<i>Integer</i>

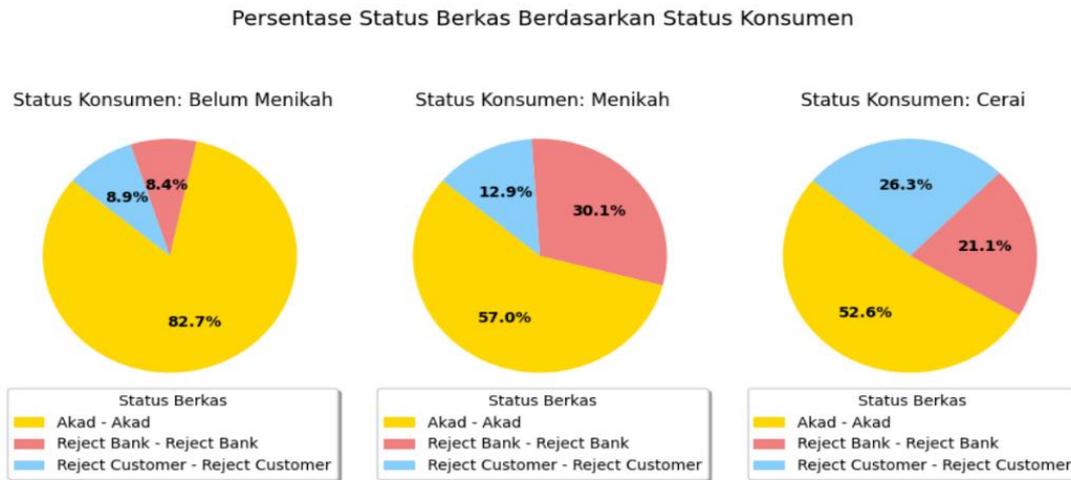
Ada enam data yang awalnya berjenis String atau objek dalam catatan data konsumen. Oleh karena itu, dilakukan label encoder untuk mengubah jenis data tersebut menjadi integer (Tabel 1). Selanjutnya membagi data menjadi

beberapa folder untuk ekstraksi fitur pelatihan dan validasi, pelatihan untuk memperoleh model dengan algoritma Random Forest (rfmodel) data latih yang sudah dibagi (X\_train sebagai fitur dan y\_train sebagai target). Model tersebut akan siap untuk digunakan untuk melakukan klasifikasi atau evaluasi lebih lanjut. Visualisasi pohon keputusan sebagai luaran dari model Random Forest yang telah dilatih sebelumnya.

### 3. HASIL DAN PEMBAHASAN

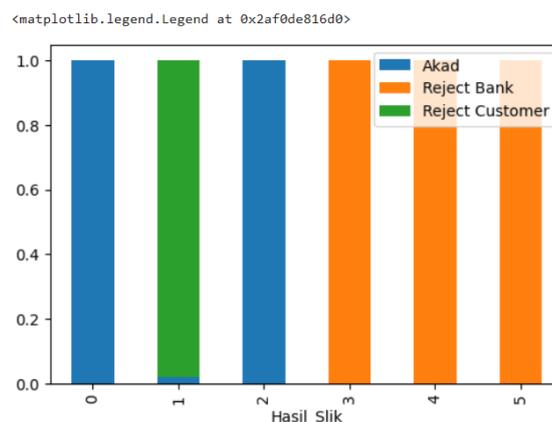
#### 3.1. Deskripsi Konsumen berdasarkan Status Berkas

Gambaran sampel data yang dilakukan preprocessing sebelum pengujian ditunjukkan pada Gambar 3 dalam bentuk diagram pie untuk 3 buah kategori status konsumen. Kategori “Status Akad” memperlihatkan dengan jumlah persentase yang paling tinggi, selanjutnya untuk reject bank menempati urutan kedua, dan yang paling rendah presentasenya adalah *reject customer*.



Gambar 3. Status Konsumen berdasarkan Status Berkas

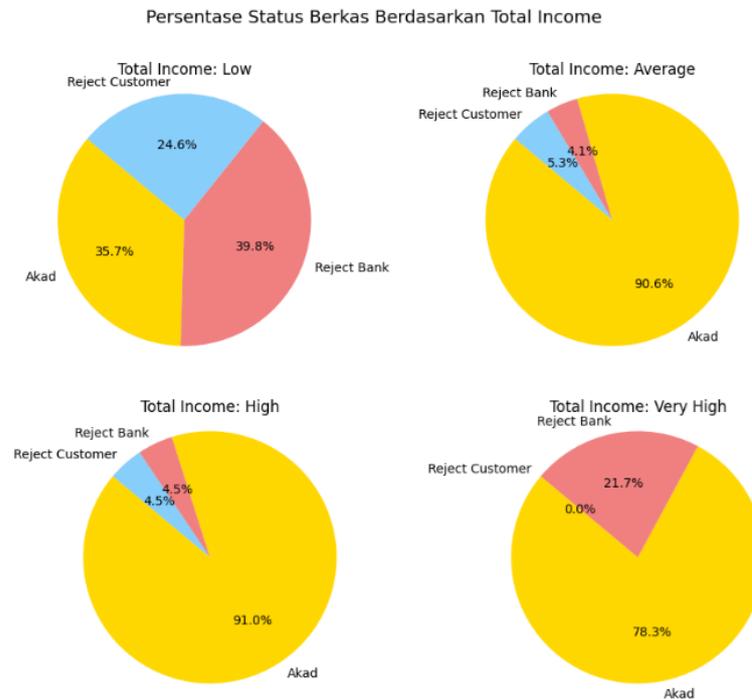
Selanjutnya melihat perbandingan hasil slik dari status berkas pada Gambar 4 dimana untuk presentase akad yang paling tinggi adalah untuk kategori 0 (Tidak ada pinjaman) dan 2 (Kol 2) serta terdapat juga pada kol 1, untuk reject bank pada hasil slik dengan kategori 3, 4, dan 5 (kol 3, kol 4, kol 5) dan untuk reject customer terdapat pada 2 (Kol 1 atau lancar) ini terjadi dikarenakan berkas yang diajukan konsumen mengalami turun plafond sehingga konsumen mengundurkan diri dari kredit KPR.



Gambar 4. Hasil slik berdasarkan Status Berkas

Diagram batang bertumpuk menunjukkan hubungan antara kategori penghasilan ('Low', 'Average', 'High', 'Very High') yang dibuat dari perhitungan total penghasilan dengan status berkas (Akad, Reject Bank, Reject Customer). Setiap bar pada diagram tersebut mewakili persentase dari tiap status berkas dalam setiap kategori penghasilan. Ini membantu dalam melihat proporsi status berkas dalam setiap kategori penghasilan. Melalui diagram ini, dapat

dilihat apakah ada pola atau perbedaan signifikan dalam persentase status berkas diantara kategori penghasilan yang berbeda. Hal ini dapat memberikan wawasan mengenai hubungan antara penghasilan dengan keputusan status berkas terkait aplikasi atau pengajuan tertentu.

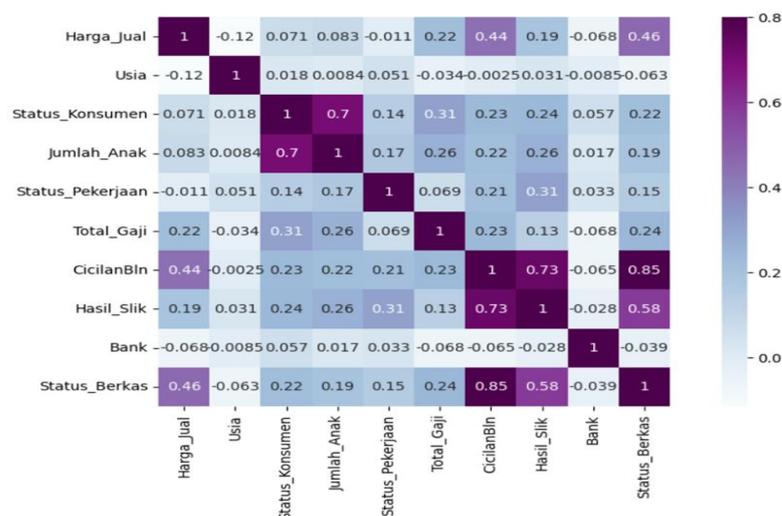


**Gambar 5.** Total Income dengan Status Berkas

Gambar 5 memperlihatkan bahwa kategori Akad untuk semua kategori income memiliki jumlah Akad Kredit yang besar dari kategori Income Low sampai Very High, untuk Reject Bank paling besar pada kategori income “Low dan Very High”, serta pada kategori Reject Bank menunjukan bahwa banyak konsumen yang mengundurkan diri karena income berada pada karegori “Low”.

### 3.2 Korelasi antar variabel

Korelasi dengan heat map pada Gambar 6 menunjukkan bahwa tidak ada nilai 0 yang menunjukkan bahwa tidak ada korelasi. Nilai korelasi berkisar antara -1 hingga 1, dimana nilai 1 menunjukkan korelasi positif sempurna, nilai -1 menunjukkan korelasi negatif sempurna.



**Gambar 6.** Matriks Heat Map untuk representasi nilai korelasi



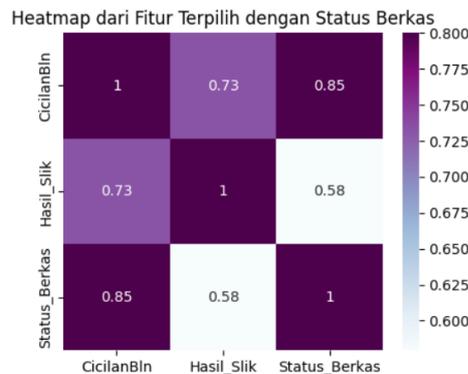


Gambar 8. Visualisasi Pohon Data Train 60% dan Test 40%



Gambar 9. Visualisasi Pohon Data Train 70% dan Test 30%

Nilai value masing-masing kelas dengan urutan X1, masuk kedalam kelas adalah 351, 71 dan 46. Jika nilai skor 'X1' >= 31.5 maka akan masuk kelas X6. Namun, jika nilai skor 'X1' <= 31.5 akan bernilai TRUE masuk ke kelas X0, dan berlanjut ke kelas selanjutnya. Sampai ke kelas terakhir dari 500 populasi diambil 313 sampel. Memilih Fitur dan melatih ulang model dengan fitur terpilih. Proses pemilihan fitur ini bertujuan untuk meningkatkan performa model dengan memfokuskan penggunaan hanya pada fitur-fitur yang dianggap penting, sehingga dapat memperbaiki akurasi atau efisiensi model klasifikasi yang dibangun.



Gambar 10. Matriks *Heatmap* dari Fitur Terpilih

Berdasarkan fitur terpilih *Heat Map* pada Gambar 10 dapat disimpulkan bahwa semakin gelap warna dari *heat map* maka menunjukkan bahwa korelasi antara fitur semakin erat hubungan atau korelasi antara pasangan fitur dalam dataset. Seperti fitur CicilanBln dengan Hasil Slik menunjukkan nilai 0.73 yang menandakan bahwa semakin dekat nilainya dengan 1 maka korelasi atau hubungan antara fitur semakin erat. Kemudian untuk CicilanBln dengan Status Berkas memiliki korelasi sebesar 0.85, dari fitur terpilih dilatih ulang sehingga dapat meningkatkan akurasi dan peforma model pada *random forest*.

Melakukan perhitungan yang lebih rinci berdasarkan confusion matrix yang telah dihasilkan sebelumnya. maka rincian tentang kinerja model klasifikasi untuk setiap kelas, mencakup jumlah TP, TN, FP, dan FN yang dapat memberikan wawasan tentang seberapa baik model dapat memngklasifikasi setiap kelas target ('Akad', 'Reject Bank', 'Reject Customer').

```

Kelas 0:
Total True Positive (TP): 361
Total True Negative (TN): 102
Total False Positive (FP): 2
Total False Negative (FN): 2

Kelas 1:
Total True Positive (TP): 62
Total True Negative (TN): 403
Total False Positive (FP): 0
Total False Negative (FN): 2

Kelas 2:
Total True Positive (TP): 40
Total True Negative (TN): 425
Total False Positive (FP): 2
Total False Negative (FN): 0
    
```

Gambar 11. Nilai TP, TN, FP dan FN

Pada klasifikasi multiclass, *confusion matrix* menampilkan jumlah *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) untuk masing-masing kelas. Untuk menghitung *confusion matrix* seperti akurasi (*accuracy*) yang menunjukkan proporsi prediksi benar terhadap seluruh prediksi, presisi (*precision*) mengukur seberapa banyak prediksi benar dari total prediksi positif untuk suatu kelas, *recall* mengukur seberapa banyak instance positif yang terdeteksi dengan benar dari total instance aktual positif, dan *F1-score* untuk rata-rata harmonik dari presisi dan recall, di bawah ini rumus *accuracy*, *precision*, *recall* dan *f1 Score* dalam *confusion matrix multiclass*.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Presisi + Recall}$$

Selanjutnya menghitung masing masing kelas dari *confusion matrix* terlihat pada Tabel 2.

**Tabel 2.** *Confusion Matrix* variabel Akad, *Reject bank*, dan *Reject Customer*

Aktual (Akad)	Prediksi	
	Positive	Negative
Positive	361	2
Negative	2	102

Aktual ( <i>Reject Bank</i> )	Prediksi	
	Positive	Negative
Positive	62	2
Negative	0	403

Aktual ( <i>Reject Customer</i> )	Prediksi	
	Positive	Negative
Positive	40	0
Negative	2	425

**Tabel 3.** Hasil pengukuran metrics *Accuracy*, *precision*, *Recall*, *F1 Score* untuk kelas Akad, *Reject Bank*, dan *Reject Customer*

No	Kelas	Measurement (%)			
		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0	Akad	99,14	99,45	99,44	99,45
1	<i>Reject Bank</i>	99,57	100	96,87	98,41
2	<i>Reject Customer</i>	99,57	95,23	100	97,55

#### 4. KESIMPULAN

Berdasarkan uraian pembahasan paragraf sebelumnya dan Tabel 3 diperoleh Kesimpulan dari hasil penelitian yang telah dilakukan bahwa klasifikasi status pengajuan kredit KPR menggunakan algoritma Random Forest pada PT Yuki Prima Anugrah berhasil dilakukan. Jumlah data yang digunakan sebanyak 969 data dan jumlah fitur yang digunakan sebanyak 9 fitur, didapatkan hasil akurasi sebesar 99% yang termasuk ke dalam klasifikasi yang sangat baik sekali berada pada rentang 90%-100%. Selain itu, proses seleksi fitur dapat meningkatkan performa model klasifikasi dengan fokus pada fitur-fitur yang memiliki pengaruh signifikan, sehingga dapat meningkatkan akurasi dan efisiensi kinerja model.

#### DAFTAR PUSTAKA

- Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1), 11–16. <https://doi.org/10.31294/inf.v11i1.20637>
- Irma Purnamasari, A., & Ali, I. (2024). PERBANDINGAN TINGKAT AKURASI ALGORITMA DECISION TREE DAN RANDOM FOREST DALAM MENGLASIFIKASI PENERIMA BANTUAN SOSIAL BPNT DI DESA SLANGIT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).
- Lestari, T. S., & Sirodj, D. A. N. (2022). Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest. *Jurnal Riset Statistika*, 1(2), 160–167. <https://doi.org/10.29313/jrs.v1i2.525>
- Madaerdo Sotarjua, L., Budhi Santoso, D., Singaperbangsa Karawang Ji Ronggo Waluyo, U. H., Telukjambe Timur, K., Karawang, K., & Barat, J. (2022). PERBANDINGAN ALGORITMA KNN, DECISION TREE,\*DAN RANDOM\*FOREST PADA DATA IMBALANCED CLASS UNTUK KLASIFIKASI PROMOSI KARYAWAN. *JURNAL INSTEK*, 7(2), 192–200.
- Pahlevi, O., & Handrianto, Y. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. *Jurnal Infortech*, 5(1). <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech>
- Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)*, 1(1), 27–31.
- Satria, M. R. (2018). ANALISIS PERBANDINGAN PEMBERIAN KREDIT PEMILIKAN RUMAH (KPR)

PADA BANK KONVENSIONAL DENGAN PEMBIAYAAN MURABAHAH (KPR) PADA BANK SYARIAH. *Amwaluna: Jurnal Ekonomi Dan Keuangan Syariah*, 2(1), 105–118. <https://doi.org/10.29313/amwaluna.v2i1.2880>

- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variasiunm31>
- Takawira, O., & Muteba Mwamba, J. W. (2022). AN ANALYSIS OF SOVEREIGN CREDIT RATINGS USING RANDOM FOREST. *INTERNATIONAL JOURNAL OF ECONOMICS AND FINANCE STUDIES*, 14(1), 29–87. <https://doi.org/10.34109/ijefs>
- Zailani, A. U., & Hanun, N. L. (2020). PENERAPAN ALGORITMA KLASIFIKASI RANDOM FOREST UNTUK PENENTUAN KELAYAKAN PEMBERIAN KREDIT DI KOPERASI MITRA SEJAHTERA. *Infotech: Journal of Technology Information*, 6(1), 7–14. <https://doi.org/10.37365/jti.v6i1.61>