Jurnal Elektronika dan Teknik Informatika Terapan
Vol. 1 No. 3 September 2023
e-ISSN : 2988-0874, p-ISSN : 2988-0866, Hal 140-147
DOI : https://doi.org/10.59061/jentik.v1i3.383

Prediksi Kemampuan Pembayaran Klien Home Credit Menggunakan Model *Random Forest, Decision Tree, Dan Logistic Regression*

Nurul Chairunnisa
Universitas Islam 45

Alamat: l. Cut Mutia No.83, RT.004/RW.009, Margahayu, Kec. Bekasi Tim.,
Kota Bks, Jawa Barat 17113
Email: nurulchair12@gmail.com

Abstract.*Home Credit is a global financial company that provides consumer loan services. The purpose of this research is to predict the ability of clients to pay in order to make it easier for companies to provide loans or not. Not being careful in analyzing lending will cause credit risk. So to reduce these risks, the company needs an analysis to predict the client's repayment ability to determine whether to pay or not as a reference for the company in providing credit loans. By using the previous member criteria data, predictions of the smoothness of payments can be made using data mining. The data mining techniques used are Random Forest Classifier, Decision Tree Classifier, and Logistic Regression Classifier. The models used are Random Forest, Decision Tree, Logistic Regression, which determine the likelihood or opportunity based on the data of previous members, and the results. The criteria used consist of the ten best features selected based on the results of best feature importance. The evaluation results of the random forest model are able to predict the ability to pay home credit clients with a high level of test accuracy score of 0.9967, ROC value of 0.9967, recall value of 1.00 compared to the other two models.*

Keywords: *Home Credit, Random Forest, Decision Tree, Logistic Regression.*

Abstrak.Home Credit adalah perusahaan keuangan global yang menyediakan layanan pinjaman konsumen. Tujuan dilakukannya penelitian ini adalah untuk memprediksi kemampuan pembayaran klien agar memudahkan perusahaan untuk memberikan pinjaman atau tidak. Tidak hati-hatinya dalam menganalisis pemberian pinjaman akan menimbulkan risiko kredit. Maka untuk mengurangi risiko yang tersebut, perushaan membutuhkan analisa untuk memprediksi kemampuan pembayaran klien untuk menentukan apakah akan membayar atau tidak sebagai acuan bagi Perusahaan dalam memberikan pinjaman kredit. Dengan menggunakan data kriteria anggota sebelumnya, prediksi kelancaran pembayaran dapat dibuat dengan menggunakan data mining. Teknik data mining yang digunakan adalah *Random Forest Classifier, Decision Tree Classifier*, dan *Logistic Regression Classifier*. Model yang digunakan yaitu *Random Forest, Decision Tree, Logistic Regression* yang menentukan kemungkinan atau peluang berdasarkan data anggota sebelumnya, dan hasilnya Kriteria yang digunakan terdiri dari sepuluh fitur terbaik yang dipilih berdasarkan hasil *best feature importance*. Hasil evaluasi model *random forest* mampu memprediksi kemampuan pembayaran klien home credit dengan tingkat nilai akurasi test yang tinggi sebesar 0.9967, nilai ROC sebesar 0.9967, nilai recall sebesar 1.00 dibandingkan dengan dua model lainnya.

Kata kunci: *Home Credit, Random Forest, Decision Tree, Logistic Regression.*

LATAR BELAKANG

Saat ini, teknologi semakin berkembang dan mulai masuk ke berbagai industri. Teknologi sangat berpengaruh bagi seluruh kegiatan yang dilakukan oleh sebuah Perusahaan bahkan tidak terlepas dari itu. Para pengusaha dapat menggunakan berbagai aplikasi komputer yang telah tersedia untuk mengelola bidang usaha mereka. Oleh karena itu, tidak dapat dipungkiri bahwa sistem selalu diperlukan untuk setiap tindakan yang dilakukan oleh perusahaan agar dapat mempengaruhi pertumbuhannya. Seperti sistem untuk mengetahui kemampuan pembayaran klien apakah layak atau tidak untuk dipertahankan (Marvin, 2018). Salah satu masalah yang sangat penting bagi lembaga keuangan adalah penilaian risiko

Received Juni 30, 2023; Revised Juli 2, 2023; Accepted Agustus 14, 2023
* Nurul Chairunnisa, *nurulchair12@gmail.com*

kelayakan kredit (Pahlevi et al., 2023). Sehingga perusahaan dapat mengetahui atau menilai klien tersebut layak diberikan piutang atau tidak.

PT Home Credit Indonesia termasuk perusahaan yang membutuhkan sistem ini. Home Credit pertama kali memasuki pasar Indonesia pada tahun 2013 dengan kantor pusat yang terletak di Jakarta. Bisnis mereka telah tumbuh dan meluas ke seluruh wilayah Indonesia, termasuk kota-kota seperti Bandung, Makassar, Surabaya, Yogyakarta, Semarang, Malang, Denpasar, Pekanbaru, Medan, Batam, Palembang, Banjarmasin, Pontianak, Manado, dan Balikpapan, sejak permulaan tahun tersebut. Perusahaan berencana untuk menyediakan layanannya ke seluruh kota di Indonesia pada tahun 2021. Individu yang berminat untuk membeli barang seperti furnitur, perangkat elektronik, ponsel, serta peralatan rumah tangga dapat memperoleh pembiayaan di berbagai toko yang bekerja sama dengan perusahaan ini (Vanessa, 2021).

Pinjaman dapat menghasilkan keuntungan, tetapi disisi lain juga memiliki resiko kredit yang tinggi sehingga dapat mengalami kerugian (Suwati et al., 2022). Risiko kredit adalah risiko kerugian yang berkaitan dengan kemungkinan bahwa pihak lain tidak akan memenuhi kewajibannya atau bahwa peminjam tidak akan membayar pinjamannya Kembali (Handayani et al., 2021). Dalam upaya untuk mengatasi masalah kredit, Home Credit menetapkan bahwa "Tanggal Jatuh Tempo merujuk pada tanggal yang tercantum dalam Lembar Tagihan di mana setidaknya Pembayaran Minimum harus dilakukan oleh Pemegang Kartu kepada Home Credit agar menghindari Biaya Keterlambatan serta status kolektibilitas yang buruk; yakni 21 (dua puluh satu) hari dihitung sejak lembar tagihan dicetak. Jika tanggal jatuh tempo jatuh pada hari libur atau akhir pekan, maka akan digeser ke hari berikutnya."

Maka dari itu untuk mengurangi risiko tersebut, perusahaan memerlukan analisa untuk memprediksi kemampuan pembayaran klien untuk menentukan apakah akan membayar atau tidak sebagai acuan bagi Perusahaan dalam memberikan pinjaman kredit.

KAJIAN TEORITIS

Home Credit adalah perusahaan keuangan global yang menyediakan layanan pinjaman konsumen. Mereka fokus pada memberikan akses ke layanan keuangan kepada individu yang mungkin sulit mendapatkan kredit dari lembaga keuangan tradisional, terutama bagi mereka yang memiliki riwayat kredit yang terbatas atau tidak ada sama sekali. Home Credit sering kali menargetkan segmen pasar yang tidak terlayani oleh lembaga keuangan lainnya, seperti penduduk dengan pendapatan rendah atau tanpa rekening bank. Visi dari Home Credit sendiri

yaitu menjadikan penyedia solusi keuangan terdepan yang mampu menjangkau Masyarakat untuk mencapai tujuan finansial mereka dan mewujudkan Impian hidup. Sedangkan misi dari Home Credit yaitu menghadirkan akses finansial, meningkatkan kualitas hidup, inovasi berkelanjutan, kemitraan yang kuat, serta pertanggungjawaban dan integritas (Vanessa, 2021).

Decision tree merupakan salah satu cara untuk memprediksi suatu masalah adalah dengan cara membentuk suatu pohon keputusan kemudian memecahnya menjadi kumpulan yang lebih kecil dan secara bertahap meningkatkan proses pengambilan keputusan (Rizky & Andriyansyah, 2023). Dalam data mining, *decision tree* juga disebut sebagai pohon keputusan, membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil dengan memperhatikan variable tujuan (Muningsih, 2022). Menurut penelitian (Handayani et al., 2021), algoritma decision tree memiliki kemampuan untuk mengidentifikasi pengetahuan atau pola-pola kesamaan karakteristik dalam kelompok atau kelas tertentu. Hasil evaluasi confusion matrix menunjukkan akurasi sebesar 79%.

Metode Pendekatan *Random Forest* merupakan salah satu cara yang efisien dan mampu memberikan rekomendasi optimal dibandingkan dengan pendekatan-pendekatan pembelajaran mesin lainnya (Putra, 2019). *Random forest* merupakan salah satu bentuk metode *decision tree*, sehingga karakteristik yang dimiliki yang mirip dalam jalannya. Berbeda dari *decision tree, random forest* merupakan gabungan beberpa pohon atau pohon yang dibuat sebagai model (Rizky & Andriyansyah, 2023). Dibuktikan oleh hasil penelitian (Prasojo & Haryatmi, 2021) yang menunjukkan bahwa algoritma *random forest* memiliki tingkat akurasi 0,83, atau 83%, yang menempatkannya dalam kategori model yang sangat baik.

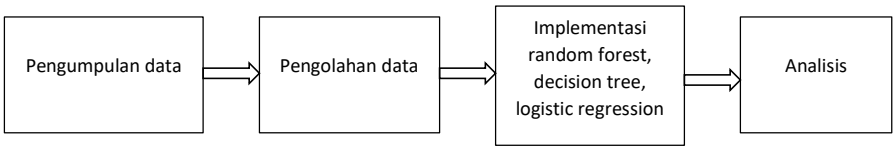
Regresi logistik adalah bentuk analisis regresi yang dipergunakan untuk menguraikan koneksi antara variabel tergantung serta variabel independen. Rentangnya terletak antara 0 dan 1, dapat menggambarkan kebenaran atau ketidakbenaran, serta ukurannya dapat bervariasi. Jenis variabel independennya termasuk dalam kategori. (Pramakrisna et al., 2022). Hasil dari penelitian (Kasidi & Christanto, 2022) menggunakan pemodelan *regresi logistic* biner. Nilai ROC dan AUC digunakan sebagai nilai kelayakan, dan tujuh variabel berdampak pada model, dengan nilai SMOTE *logistic biner* sebesar 0.72 dan nilai AUC biner asli sebesar 0,68.

METODE PENELITIAN

Tahapan Penelitian

Dalam riset ini, pendekatan yang digunakan adalah menggunakan informasi historis mengenai kriteria pelanggan yang akan dimanfaatkan untuk meramalkan kemampuan pembayaran melalui penerapan teknik data mining. Menggunakan perbandingan dari beberapa

model yaitu *random forest, decision tree, logistic regression* untuk mengatasi permasalahan yang dihadapi.



Gambar 1 Flowchart Metode Penelitian

Pengumpulan Data

Informasi yang digunakan dalam penelitian ini diperoleh dari https://www.kaggle.com/c/home-credit-default-risk pada bulan Juni 2023.. Data tersebut berasal dari *Kaggle* yaitu salah satu platform komunitas online untuk praktisi data (Dqlab, 2023). Kaggle didirikan oleh *Goldbloom* pada tahun 2010 (Nirla05, 2022). Dataset terdiri dari 307511 data dengan 122 fitur kolom.

Pengelolaan Data

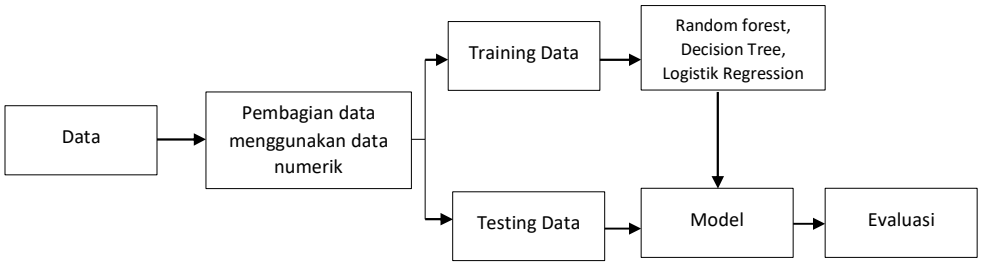
Pada tahap pengolahan data dilakukan pembersihan data, informasi data. Terdapat 67 kolom yang memiliki data kosong. Hasil dari pembersihan data didapatkan sebanyak 72 kolom yang dapat diolah untuk tahap selanjutnya. Terdapat anomali pada kolom DAYS_EMPLOYED sebesar 18%, pada kolom DAYS_BIRTH semua datanya bernilai negatif. Kemudian untuk kolom DAYS_EMPLOYED dilakukan handling untuk menghilangkan anomali dan untuk kolom DAYS_BIRTH dilakukan handling. Kemudian untuk melihat *outlier* disini menggunakan *boxplot* dan terlihat bahwa ada *outlier* yang begitu banyak. Namun untuk data ini outlier tidak dianggap sebagai data outlier, karena untuk data peminjaman kemungkinan ada klien yang meminjam dalam jumlah yang besar. Fase pengolahan data selanjutnya adalah *data encoding* yaitu mengubah variabel kategorikal menjadi bentuk numerik, kemudian dilakukan *best feature importance* untuk memilih 10 fitur terbaik yang diurutkan berdasarkan importance tertinggi sehingga dapat memasukan semua fitur data dalam bentuk numerik ke dalam algoritma pemodelan, yang umumnya hanya menerima input numerik.

	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_TYPE_SUITE	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	NAME_FAMILY_STATU
0	1	0	1	0	1	0	202500.0	406597.5	24700.5	351000.0	6	7	4	
1	0	0	0	0	0	0	270000.0	1293502.5	35698.5	1129500.0	1	4	1	
2	0	1	1	1	1	0	67500.0	135000.0	6750.0	135000.0	6	7	4	
3	0	0	0	0	1	0	135000.0	312682.5	29686.5	297000.0	6	7	4	
4	0	0	1	0	1	0	121500.0	513000.0	21865.5	513000.0	6	7	4	

Gambar 2 Hasil Encoding Kategorik menjadi Numerik

Implementasi *Random Forest, Decision Tree dan Logistic Regression*

Diberikan informasi terkait dengan Langkah impementasi alogaritma random forest, decision tree dan logistic regression untuk prediksi kemampuan pembayaran klien home credit. Pertama yaitu melakukan input data hasil dari *data encoding* dan *best feature importance*. Setelah itu, data dipisahkan menjadi data *training* dan data *testing* dengan memanfaatkan teknik *splitting* data, yaitu pendekatan yang memisahkan data menjadi dua kelompok atau lebih untuk membentuk subset data (Trivusi, 2022). Data *training* dimanfaatkan sebagai masukan bagi algoritma *random forest,* sementara data *training* digunakan untuk menguji serta menilai hasil atau model yang terbentuk melalui penerapan algoritma *random forest.*



Gambar 3 Langkah Implementasi

Pengujian kinerja ketiga model ini dilaksanakan dengan menggunakan beberapa indikator pengukuran, mencakup akurasi, recall, presisi, nilai ROC-AUC, dan f1-score. Akurasi adalah parameter yang umum serta sederhana digunakan untuk menilai performa algoritma klasifikasi, yakni angka yang menunjukkan kesesuaian antara prediksi sistem dengan prediksi manusia. Nilai presisi merujuk pada nilai sensitivitas atau ketepatan sistem dalam mengidentifikasi data kelas negatif dan positif dengan benar. Nilai recall mengindikasikan tingkat keberhasilan atau spesifikasinya dalam mengenali kembali informasi yang benar terkait data kelas negatif atau positif. (Azhari et al., 2021). F1-skor merupakan hasil harmonisasi dari presisi dan recall. Nilai F1-skor terbaik adalah 1.0 sementara yang terburuk adalah 0. Dalam representasinya, ketika F1-skor memiliki nilai yang tinggi, ini menggambarkan bahwa model klasifikasi menunjukkan baiknya presisi dan recall. (Setiawan, 2020). Kurva ROC merupakan sebuah alat evaluasi untuk mengukur performa dalam permasalahan klasifikasi dengan maksud menentukan ambang batas suatu model. ROC dapat menampilkan hasil dalam bentuk akurasi dan membandingkan klasifikasi secara grafis. Sementara itu, Area Di Bawah Kurva (AUC) berfungsi untuk menghitung performa umum dari model.

AUC	Interpretation
1.0 (100%)	Model Sempurna

0.9 – 0.99 (90 – 99%)	Model Luar Biasa
0.8 – 0.89 (80 – 89%)	Model Sangat Baik
0.7 – 0.79 (70 – 79%)	Model Cukup Baik
0.51 – 0.69 (51 – 69%)	Model Buruk
< 0.05 (50%)	Model Tidak Bermakna

Tabel 1 Klasifikasi Nilai AUC

Dari tabel di atas, dapat diamati bahwa AUC mengukur nilai yang terletak di bawah kurva ROC, dan semakin mendekati nilai satu, ROC akan menjadi semakin baik.

Analisis

Pada langkah ini, evaluasi dilakukan terhadap model yang dihasilkan dalam konteks studi kasus prediksi kemampuan pembayaran klien Home Credit. Selain itu, hasil pengujian berdasarkan parameter pengujian juga dieksplorasi untuk menilai kualitas dari model yang telah dibuat.

HASIL DAN ANALISIS

Dibawah ini adalah hasil penilaian dari model yang terbentuk melalui algoritma *random forest, decision tree, logistic regression* dimana data numerik digunakan sebagai data *training* dan *testing*.

	Models	Training Accuracy Score	Testing Accuracy Score	Recall	ROC Score
1	Random Forest	1.0000	0.9967	1.00	0.9967
2	Decision Tree	1.0000	0.9087	0.91	0.9087
0	Logistic Regression	0.6728	0.6745	0.67	0.6745

Gambar 4 Hasil Evaluasi Model

Berdasarkan hasil penilaian model yang terbentuk, terlihat bahwa akurasi uji dari *random forest* adalah 0.9967 dengan recall mencapai 1.00 dan skor ROC mencapai 0.9967. Pada model *decision tree*, nilai akurasi uji mencapai 0.9087 dengan recall sekitar 0.91 dan skor ROC sekitar 0.9087. Sementara pada *logistic regression*, akurasi uji mencapai 0.6745 dengan recall sekitar 0.67 dan skor ROC sekitar 0.6745. Dengan mempertimbangkan tingkat akurasi dari ketiga model, maka dapat disimpulkan bahwa model *random forest* memiliki tingkat akurasi uji yang lebih tinggi dibandingkan dengan dua model lainnya.

KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi dari ketiga model yang telah dilakukan, maka dapat disimpulkan model *random forest* mampu memprediksi kemampuan pembayaran klien home

credit dengan tingkat nilai akurasi test yang tinggi sebesar 0.9967, nilai ROC sebesar 0.9967, nilai recall sebesar 1.00 dibandingkan dengan dua model lainnya.

DAFTAR PUSTAKA

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, *5*(2), 640–651. https://doi.org/10.30865/mib.v5i2.2937

Dqlab. (2023). *4 Platform Kekinian untuk Portofolio Data Analyst*. Dqlab. https://dqlab.id/4-platform-kekinian-untuk-portofolio-data-analyst#:~:text=Kaggle adalah salah satu platform,menyelenggarakan kompetisi di ilmu data

Handayani, N., Wahyono, H., Trianto, J., & Permana, D. S. (2021). Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45. *JURIKOM (Jurnal Riset Komputer)*, *8*(6), 198–204. https://doi.org/10.30865/jurikom.v8i6.3643

Kasidi, & Christanto, J. (2022). Perancangan Model untuk Prediksi Potensi Churn pada Debitur KPR dengan Regresi Logistik. *Institut Teknilogi Sepuluh November*. https://repository.its.ac.id/92487/

Marvin, K. (2018). *Klasifikasi Potensi Pembayaran Kredit Customer Dengan Metode C4. 5 Pada Pt. Autochem Industry*. *34–56*, 1–191. http://repositori.buddhidharma.ac.id/830/%0Ahttp://repositori.buddhidharma.ac.id/830/1/Marvin Kristianto - 20141000034.pdf

Muningsih, E. (2022). Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan Split Data. *Jurnal Teknoinfo*, *16*(1), 113–118. https://doi.org/10.33365/jti.v16i1.1561

Nirla05. (2022). *Mengenal Lebih Jauh Apa itu Kaggle, Fungsi Kaggle dan Manfaatnya*. IDMETAFORA. https://idmetafora.com/news/read/1827/Mengenal-Lebih-Jauh-Apa-Itu-Kaggle-fungsi-Kaggle-dan-Manfaatnya.html

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

Pramakrisna, F. D., Adhinata, F. D., & Tanjung, N. A. F. (2022). Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression. *Teknika*, *11*(2), 90–97. https://doi.org/10.34148/teknika.v11i2.466

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 79–89. https://doi.org/10.25077/teknosi.v7i2.2021.79-89

Putra, M. I. (2019). Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest pada BRI Kantor Cabang Pelaihari. *Jurnal Teknik Informatika Dan Sistem Informasi, UNIVERSITAS ISLAM NEGERI SUNAN AMPEL*, *13*(1), 61. https://core.ac.uk/download/pdf/232849774.pdf

Rizky, M., & Andriyansyah, R. (2023). *Komparasi Performa Model Terhadap Klasifikasi Sinyal Mit-Bih Arrhythmia Database* (M. Y. H. Setyawan (ed.); satu). Penerbit Buku Pedia.

Setiawan, S. (2020). *Membicarakan Precision, Recall, dan F1-Score*. Medium.

https://stevkarta.medium.com/membicarakan-precision-recall-dan-f1-score-e96d81910354

Suwati, Yesputra, R., & Sapta, A. (2022). Prediksi Kelancaran Pembayaran Angsuran Pada Koperasi Dengan Metode Naive Bayes Classifier. *Indonesian Journal of Computer Science*, *11*(2), 635–644. https://doi.org/10.33022/ijcs.v11i2.3080

Trivusi. (2022). *Data Splitting: Pengertian, Metode, dan Kegunaannya*. Trivusi. https://www.trivusi.web.id/2022/08/data-splitting.html#:~:text=Data splitting atau pemisahan data,lainnya digunakan untuk melatih model

Vanessa, Y. (2021). Pelaksanan Perjanjian Finansial Thecnologi Antara Nasabah Dengan PT Home Credit Indonesia Di Kecamatan Senapelan. *Uin Suska Riau*. https://repository.uin-suska.ac.id/49299/2/SKRIPSI YOKO VANESSA.pdf