



Artikel Penelitian

Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode *Random Forest*

Budi Prasajo^{a,*}, Emy Haryatmi^a

Program Studi Sistem Informasi, Jurusan Sistem Informasi Bisnis, Universitas Gunadarma, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 19 Maret 2020

Revisi Akhir: 29 Agustus 2021

Diterbitkan Online: 02 September 2021

KATA KUNCI

Algoritma Random Forest,

Data Kredit

KORESPONDENSI

E-mail: ubud28@gmail.com

A B S T R A C T

Perkembangan di Negara Indonesia saat ini tidak terlepas dari peminan lembaga keuangan dengan salah satunya adalah perbankan. Bank berkontribusi dalam meningkatkan pertumbuhan dan perkembangan Negara Indonesia. Perbankan berfungsi sebagai lembaga yang memiliki peran sentral dalam meningkatkan pertumbuhan ekonomi Negara. Analisa pemberian kredit perbankan harus memerhatikan Prinsip-prinsip pemberian kredit. Jumlah kredit yang disalurkan oleh bank kepada pihak lain ataupun masyarakat dipengaruhi oleh beberapa faktor. Data dalam jumlah besar pada perbankan khususnya perkreditan tersebut dapat diolah menggunakan beberapa metode tertentu akan memberikan informasi baru yang dapat mendukung dan membantu perbankan mengambil keputusan atau suatu kebijakan, salah satu kebijakannya yaitu dapat memprediksi kelayakan kredit pinjaman secara dini untuk mengetahui nasabah yang layak atau tidak layak, atau menggunakan salah satu teknik melakukan prediksi yang dapat digunakan adalah dengan teknik penggalian data menggunakan algoritma *random forest*. Penelitian ini bertujuan untuk mengetahui penerapan metode klasifikasi dengan algoritma *random forest* serta analisa hasil terbaik dari algoritma *random forest* pada setiap kreditur. Variabel yang dianalisa yaitu V1 sampai dengan V20 dilakukan menggunakan perangkat lunak R. Tahapan metode penelitian menggunakan CRIPS-DM. Tahap pelatihan menggunakan 80% data dan Pengujian menggunakan 20% data secara acak dari 1000 data. Hasil performa dari algoritma *random forest* tersebut yaitu memiliki tingkat akurasi sebesar 0,83 atau 83% sehingga termasuk pada kategori klasifikasi modelnya sangat bagus.

1. PENDAHULUAN

Perkembangan di Negara Indonesia saat ini tidak terlepas dari peminan lembaga keuangan dengan salah satunya adalah perbankan. Bank berkontribusi dalam meningkatkan pertumbuhan dan perkembangan Negara Indonesia [1]. Dalam melakukan analisa pemberian kredit perbankan harus memerhatikan Prinsip-prinsip pemberian kredit [3]. Jumlah kredit yang disalurkan dari pihak bank kepada nasabah terdapat beberapa faktor yang mempengaruhinya [4]. Data dalam jumlah besar pada perbankan khususnya perkreditan tersebut dapat diolah menggunakan beberapa metode tertentu akan memberikan informasi baru yang dapat mendukung dan membantu perbankan mengambil keputusan atau kebijakan. Adapun kebijakan yang dapat diimplementasikan yaitu melakukan mitigasi terhadap putusan kredit untuk setiap pengajuan kredit dari nasabah

sehingga pemutus kredit dapat dengan mudah melakukan analisa kredit, system ini dapat disebut juga dengan teknik data mining atau penggalian data. Seiring dengan perkembangan teknologi sistem informasi yang sangat signifikan, perbankan memanfaatkan perkembangan teknologi sistem informasi untuk melakukan efisiensi dan efektifitas analisa keputusan pemberian kredit. Salah satu perkembangan teknologi informasi tersebut adalah Artificial Intelligence (kecerdasan buatan) yang membuat mesin mampu menjawab pertanyaan secara cerdas sekarang ini dalam bidang rekayasa yang sangat luas [5]. Data mining dapat disebut sebagai bagian dari *Knowledge Discovery in Database* (KDD) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data. Data mining dikenal sebagai knowledge extraction, pattern analysis, information harvesting, dan Business intelligence [6] menggunakan klasifikasi algoritma *Random Forest*. Penelitian ini bertujuan untuk mengetahui penerapan metode klasifikasi dengan algoritma *Random Forest* pada german

data credit dan menganalisis hasil terbaik dari algoritma random forest pada data german credit [7]. Data german credit memiliki data input yaitu V1 sampai dengan V20 dan Data Output pada variable V21. Analisis ini menggunakan perangkat lunak R.

2. METODE

2.1. Data Set

Data set adalah kumpulan dari objek dan sifat atau karakteristik dari suatu objek itu sendiri (atribut). Dataset terdiri dari sekumpulan file atau dokumen [7]. Pada statistik, dataset biasanya berasal dari hasil pengamatan aktual yang diperoleh dari dengan mengambil sampel populasi. Dataset dapat menghasilkan algoritma yang bisa digunakan untuk pengujian pada perangkat lunak.

2.2. Machine Learning Repository

Machine learning yaitu aplikasi dari disiplin ilmu kecerdasan buatan (Artificial Intelligence) yang menggunakan teknik statistika sehingga dapat menghasilkan suatu model yang otomatis dari sekumpulan data, agar dilakukan analisa oleh computer untuk “belajar” [8].

2.3. Decision tree

Digunakan untuk melakukan klasifikasi dan prediksi karena kemudahan dalam interpretasi hasil [9]. Menurut Berry & Linoff Decision Tree yaitu suatu struktur yang dapat digunakan sebagai pemecah data yang besar menjadi record yang lebih kecil dengan mengimplementasikan alur proses keputusan. alur pada decision tree berawal dari simpul akar ke simpul daun yang dapat memprediksi kelas, sehingga decision tree dengan mudah melakukan proses konversi klasifikasi (classification rule). Konsep yang digunakan dalam decision tree tercermin dalam bentuk tabel dengan atribut dan record.[10]. Proses Decision Tree yaitu merubah data dalam bentuk tabel menjadi bentuk Tree, Menrubah Tree menjadi Rule dan melakukan penyederhanaan Rule (basuki&syarif,2003).

2.4. Random Forest

Algoritma Random Forest dapat menggunakan untuk mengklasifikasi *big data*. Prunning atau pemangkasan variable seperti descission tree tidak terdapat dalam algoritma *random forest* namun keunggulan dari *random forest* dapat menggabungkan banyak pohon dan untuk single tree yang terdiri atas satu pohon dalam melakukan klasifikasi dan prediksi kelas [11]. Random Forest memiliki 3 metode yaitu General Random Forest, Conditional Random Forest dan Improve Results of Logistic using Random Forest. Random forest termasuk kedalam teknik klasifikasi baru, apabila dilakukan perbandingan dengan metode bagging dan metode boosting. Metode random forest bertujuan dapat melakukan perbaikan atas prediksi dari metode bagging. Implementasi *random forest* telah dilakukan beberapa penelitian. Penerapan metode random forest dan metode decision tree pada 20 variabel yang ada pada data set mendapatkan hasil, bahwa secara keseluruhan metode random forest lebih baik jika dibandingkan dengan metode decision tree pada data dengan jumlah observasi yang banyak, sedangkan pada jumlah observasi yang sedikit. metode random forest memiliki kemampuan yang tidak berbeda dengan metode decision tree (Ali dkk., 2012),

dalam membangun pohon keputusan terdiri dari root, internal dan leaf node dengan cara mengambil atribut dan data secara *random* berdasarkan ketentuan. Root node yaitu simpul yang berada paling atas atau akar dari pohon keputusan. Internal node yaitu simpul yang bercabang, node tersebut memiliki minimal dua output dan satu input. Sedangkan leaf node atau terminal node merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai output. Pohon keputusan dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan nilai information gain [5].

2.5. Entropy

Entropy telah dikembangkan oleh Claude E. Shannon, yaitu salah satu yang paling metrik penting yang digunakan dalam teori informasi [12]. Metode Random Forest (RF) yaitu suatu metode yang digunakan untuk meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap node dilakukan secara acak. Metode yang digunakan untuk membangun suatu pohon keputusan terdiri dari root, internal, dan leaf node dengan cara mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. Root node merupakan simpul yang berada diatas, atau dapat disebut akar pohon keputusan. Internal node yaitu suatu simpul percabangan, adapun node tersebut memiliki minimal dua output dan hanya ada satu input. Sedangkan leaf atau terminal node adalah simpul akhir yang terdiri dari satu input yang tidak memiliki output. Pohon keputusan memulai dengan cara menghitung dari suatu nilai entropy yang dapat menjadi penentu tingkat atribut yang tidak murni dan nilaidari information gain. [14] Untuk menghitung nilai entropy digunakan rumus sebagai berikut:

$$\text{Entropy}(X) = \sum_{j=1}^k p_j * \log_2 \frac{1}{p_j} = - \sum_{j=1}^k p_j * \log_2 p_j$$

Keterangan,:

X : Himpunan Kasus

k : jumlah partisi X

Pi : Proporsi Xj terhadap X

2.6. Gain

Gain didasarkan pada penurunan entropi setelah dataset yang kemudian dibagi pada setiap atribut, Metode gain adalah pengembangan dari metode Iterative Dichotomiser 3 (ID3) [13]. Proses Pemilihan atribut yang akan menjadi simpul, baik akar (root) atau simpul internal menggunakan data atau nilai information tertinggi dari setiap atribut yang ada [15]. Untuk Nilai dari gain ratio didapatkan dari hasil perhitungan information gain yang telah dibagi dengan split information [16].

2.7. Preprocessing data

Preprocessing data adalah proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas, Preprocessing data merupakan langkah penting dalam data mining [16]. Adapun tahapan dalam preprocessing data sebagai berikut:

2.7.1. Data Cleaning

Data Cleaning yang digunakan salah satunya untuk mengisi *missing value*, identifikasi outlier, memperbaiki data noise, memperbaiki ketidakkonsistenan data dan memperbaiki dampak dari integrasi data yang dapat menimbulkan reduksi.

2.7.2. Data Integration

Data Integration yaitu suatu langkah untuk menggabungkan data dari beberapa sumber. Integrasi data dilakukan apabila terdapat data yang berasal dari berbeda-beda tempat. Langkah yang dilakukan antara lain mengintegrasikan skema, mengidentifikasi masalah entitas, dan mendeteksi sekaligus menyelesaikan konflik pada nilai data.

2.7.3. Data Transformation

Data Transformation adalah mengubah suatu data supaya diperoleh data yang lebih berkualitas. Yang akan dilakukan antara lain menghilangkan noise dari data (*smoothing*), meng-agregasi data, generalisasi data, normalisasi data, dan pembentukan atribut/fitur.

2.7.4. Data Reduction

Data Reduction yaitu langkah untuk mereduksi dimensi, atribut ataupun jumlah data. Yang akan dilakukan antara lain agregasi data cube, reduksi dimensi, diskretisasi, dan kompresi data

2.8. Kredit

Kredit adalah Penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga, imbalan atau pembagian hasil keuntungan [17].

2.9. Evaluasi dan Validasi

Evaluasi adalah tahapan lanjutan dari tujuan dari data mining. Evaluasi dan validasi dilakukan secara mendalam dengan tujuan agar hasil pemodelan (Random Forest) sesuai dengan sasaran yang ingin dicapai dalam penelitian ini. Hasil Evaluasi dan Validasi menggunakan Confusion Matrix secara manual dan otomatisasi menggunakan system R serta Kurva ROC.

2.10. Confusion Matrix

Hasil evaluasi algoritma dapat ditampilkan dengan menggunakan Confusion Matrix (Tan, 2005). Confusion Matrix adalah tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi pada suatu set data testing yang nilai-nilai yang sebenarnya sudah diketahui seperti pada tabel 1.

2.11. Precision dan Recall Curve

Pada penelitian ini precision dan recall curve hasil dari sistem R menjadi satu output, Precision didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut, sedangkan Recall Curve didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia.

Tabel 1. Confusion Matrix

		True Class	
		Good	Bad
Predicated Class	Good	True Positive (TP)	False Negative (FN)
	Bad	False Positive (FP)	True Negative (TN)

Perhitungan Akurasi Dengan Tabel 2.1 Confusion Matrix adalah sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \quad (1)$$

2.12. Precision dan Recall Curve

Kurva ROC menurut (Attenberg & Ertekin, 2013) adalah ukuran numerik untuk membedakan kinerja model, dan menunjukkan seberapa sukses dan benar peringkat model dengan memisahkan pengamatan positif dan negative. ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan false positives sebagai garis horisontal dan true positive sebagai garis vertikal[5].

2.13. Perangkat Lunak R

Software “R” adalah salah satu sistem perangkat lunak yang paling populer dan banyak digunakan untuk statistik, data mining, dan machine learning [18].R mampu menganalisa data dengan sangat efektif dan dilengkapi dengan operator pengolahan array dan matriks.

2.14. R Packages

R packages atau R library adalah kumpulan fungsi, data dan kode yang berada dalam bahasa R. R memiliki lebih dari 6 ribu packages yang berada pada repository CRAN. Ketika instalasi perangkat lunak R maka secara otomatis beberapa packages terinstal. Untuk packages yang tidak terinstal secara otomatis maka dapat diinstal secara manual.

2.15. Penelitian Terdahulu

Penelitian terdahulu yaitu acuan penulis dalam melakukan penelitian ini sehingga penulis dapat memperkaya teori yang digunakan untuk mengkaji penelitian yang dilakukan. Selain itu penelitian terdahulu digunakan untuk melihat perbedaan antara penelitian yang telah dilakukan dengan penelitian yang dilakukan.

Pertama penelitian oleh Ibnu Alfarobi, Taransa Agasya Tutupoly, Ade Suryanto (2018) dengan judul “Komparasi algoritma c4.5, naive bayes, dan random Forest untuk klasifikasi data kelulusan mahasiswa Jakarta” penelitian tersebut menggunakan *Algoritma C4.5, Naive Bayes & Random Forest* dengan kesimpulan Algoritma C4.5 mendapatkan rata-rata Akurasi Sebesar 86,52%, algoritma C4.5 mendapatkan rata-rata akurasi sebesar 85,32%, algoritma Random Forest mendapatkan rata-rata akurasi sebesar 78,55%.

Kedua penelitian oleh Iveline Anne Marie, Lukmanul Hakim, Dedy Sugiarto, Winnie Septiani (2019) dengan judul “Perbandingan Performansi Teknik Klasifikasi Breakdown Mesin pada Proses Produksi Pembuatan Battery Mobil”

penelitian tersebut menggunakan *Decision Tree*, *Bagging*, *Boosting* dan *Random Forest* dengan kesimpulan metode *Decision Tree* mendapatkan rata-rata akurasi sebesar 93,67%, metode *Decision Tree* mendapatkan rata-rata akurasi sebesar 96,2%, metode *Boosting* mendapatkan rata-rata akurasi sebesar 93,67%, metode *Random Forest* mendapatkan rata-rata akurasi sebesar 96,2%

Ketiga penelitian oleh Vanissa Wanika Siburian, Ika Elvina Mulyana (2018) dengan judul “Prediksi Harga Ponsel Menggunakan Metode *Random Forest*” penelitian tersebut menggunakan *Random Forest* dengan kesimpulan Metode *Random Forest* mendapatkan rata-rata akurasi sebesar 81%.

Keempat penulis oleh Nariswari Karina Dewi, Utami Dyah Syafitri, Soni Yadi mulyadi (2018) dengan judul “Penerapan metode *random forest* dalam driver analysis” penelitian tersebut menggunakan *Random Forest* dengan kesimpulan *Metode random forest* untuk data yang lebih dari 500 memberikan akurasi prediksi yang tinggi dan stabil, yaitu dengan tingkat misklasifikasi berkisar antara 33% dan 35.5% dengan nilai rataannya sebesar 34.5%. tingkat misklasifikasi berkisar antara 33% dan 35.5% dengan nilai rataannya sebesar 34.5%.

Kelima penulis oleh Abdachul Charim, Setio Basuki, Denar Regata Akbi (2018) dengan judul “Detect Malware in Portable Document Format Files (PDF) Using Support Vector Machine and *Random Decision Forest*” penelitian tersebut menggunakan Support Vector Machine and *Random Decision Forest* dengan kesimpulan metode SVM mendapatkan rata-rata akurasi sebesar 0.58 atau sebesar 58%, metode *Random Forest* mendapatkan rata-rata akurasi sebesar 0.84 atau sebesar 84%

3. HASIL

3.1. Alur Penelitian

Data utama pada penelitian ini yaitu German Data credit [7]. Metode penelitian digambarkan dalam skema penelitian sehingga

dapat memberikan gambaran yang jelas, teratur dan sistematis seperti ditunjukkan pada gambar 1.



Gambar 1. Metode Penelitian

3.2. Data Set

Data set yang digunakan dalam penelitian didapat dari *machine learning repository* yaitu *german credit* data yang disediakan dari Professor Dr. Hans Hofmann Institut f"ur Statistik und "Okonometrie Universit"at Hamburg. *Machine learning repository* yaitu sekumpulan database, teori domain, dan data generator yang digunakan komunitas Machine learning untuk menganalisa empiris algoritma pembelajaran mesin.

Data set berjumlah 1000 data kreditur dengan kelas yang terdiri dari 800 good (1) dan 200 bad (2), sehingga data yang akan digunakan untuk proses latih sebesar 80% dan proses uji sebesar 20% secara acak dari total 1000 data dengan rincian data sesuai. Tabel 2 adalah tabel data set yang terdiri dari 1000 data kreditur terdiri dari 21 variable mulai dari data input V1 sampai dengan V20 sedangkan V21 adalah target atau output dari hasil Inputan yang telah diproses seperti terlihat pada tabel 3.

Tabel 2. Variabel Data Input dan Data Output

Variabel	Nama Variabel	Kriteria
V1	Checking Checking account status	1) A11: ... < 0 DM 2) A12: 0 <= ... < 200 DM 3) A13: ... >= 200 DM / Salary Assignment for at least 1 year 4) A14: No checking account
V2	Loan Duration	1 – 72 Month
V3	Credit History	1) A30: No Credits Taken/All Credits Paid Back Duly 2) A31: All Credits At This Bank Paid Back Duly 3) A32: Existing Credits Paid Back Duly Till Now 4) A33: Delay In Paying Off In The Past 5) A34: Critical Account/ Other Credits Existing (Not At This Bank)
V4	Purpose of the loan	1) A40 : Car (New) 2) A41 : Car (Used) 3) A42: Furniture/Equipment 4) A43: Radio/Television 5) A44 : Domestic Appliances 6) A45 : Repairs 7) A46 : Education 8) A47 : (Vacation - Does Not Exist?) 9) A48 : Retraining 10) A49 : Business 11) A410 :Others
V5	Credit Amount	Credit (Loan) Amount
V6	Savings account/	1) A61: ... < 100 DM

	<i>bonds</i>	2) A62: 100 <= ... < 500 DM 3) A63: 500 <=...<1000 DM 4) A64: ... >= 1000 DM 5) A65: Unknown/ No Saving Account
V7	<i>Present employment since</i>	1) A71: unemployed 2) A72: ... < 1 year 3) A73: 1 <= ... < 4 years 4) A74: 4 <= ... < 7 years 5) A75:... >= 7 years
V8	<i>Instalment Rate In Percentage Of Disposable Income</i>	1) 0>10% 2) 11% > 20 % 3) 31% > 40% 4) 41 > 50%
V9	<i>Personal Status And Sex</i>	1) A91 : Male:Divorced/Separated 2) A92:Female:Divorced/Separated/Married 3) A93 : Male:Single 4) A94 : Male:Married/Widowed 5) A95 : Female:Single
V10	<i>Other debtors / guarantors</i>	1) A101 : none 2) A102 : co-applicant 3) A103 : guarantor
V11	<i>Present residence since</i>	1) 1: 0 > 2 year 2) 2: 3 > 4 year 3) 3: 5 > 6 years 4) 4: 7 > ... years
V12	<i>Property Type</i>	1) A121 : Real Estate 2) A122 : If Not A121 : Building Society Savings Agreement/ Life Insurance 3) A123 : If Not A121/A122 : Car Or Other, Not In Attribute 6 4) A124 : Unknown / No Property
V13	<i>Age in Years</i>	17 > 75 Years Old
V14	<i>Other Instalment Plans</i>	1) A141:Bank 2) A142:Stores 3) A143:None
V15	<i>Housing Type</i>	1) A151 : Rent 2) A152 : Own 3) A153 : For Free
V16	<i>Number of existing credits at this bank</i>	1) 1: 0 existing credits 2) 2: 1 existing credits 3) 3: 2 existing credits 4) 4: 3 > ... existing credits
V17	<i>Job Status</i>	1) A171 : Unemployed/ Unskilled - Non-Resident 2) A172 : Unskilled - Resident 3) A173 : Skilled Employee / Official 4) A174:Management/ Self-Employed/ Highly Qualified Employee/ Officer
V18	<i>Number of people being liable to provide maintenance for</i>	1) 1 : yes 2) 2 : no
V19	<i>Telephone Number</i>	1) A191 : None 2) A192 : Yes, Registered Under The Customers Name
V20	<i>Foreign Worker</i>	1) A201 : yes 2) A202 : no
V21	<i>Table target</i>	1) 1 : Good 2) 2 : Bad

3.3. Preprocessing Data

Preprocessing data dilakukan pada tahap awal untuk mempersiapkan data mentah sebelum dilakukannya proses pemodelan. Preprocessing dilakukan dengan cara mengubah data menjadi bentuk yang lebih mudah atau mengeliminasi data yang tidak sesuai. Preprocessing data dilakukan dengan tujuan mendapatkan hasil yang lebih akurat, pengurangan waktu perhitungan dan membuat nilai data menjadi lebih kecil. Penelitian ini menggunakan transformasi dengan model normalisasi yaitu metode min-max dari preprocessing data, karena data set pada penelitian ini sudah final dan tidak perlu dilakukan data cleansing. Tabel 3 adalah hasil transformasi data yang dilakukan terhadap 1000 data secara otomatisasi melalui sistem R.

Tabel 3. Data Set Hasil Transformasi

No	V1	V2	V3	V4	V5	...	V21
1	1	6	5	4	1169	...	1
2	2	48	3	4	5951	...	2
3	4	12	5	7	2096	...	1
...
1000	2	45	5	2	4576	...	2

Setelah dilakukan Transformasi terhadap data set didapatkan hasil sesuai tabel 3. Data Set hasil transformasi. Proses selanjutnya yaitu normalisasi data dengan menggunakan metode min max secara otomatisasi dari sistem R ataupun dengan perhitungan manual. Normalisasi Data yaitu Proses memecah tabel tunggal menjadi beberapa tabel yang saling berhubungan seperti dibawah ini:

$$\text{NewData} = (\text{Data} - \text{MinData}) / (\text{MaxData} - \text{MinData})$$

Keterangan:

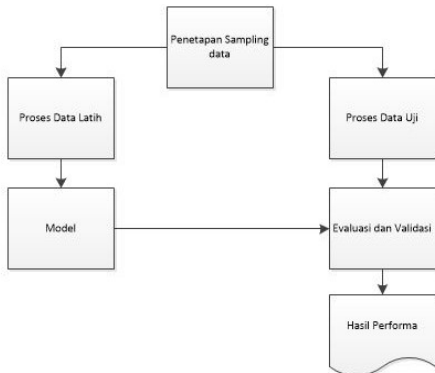
- NewData*: Hasil normalisasi data
- Data*: Data yang akan dinormalisasi (Profil Kreditur)
- MinData*: Nilai minum data
- MaxData*: Nilai maximum data

Tabel 4. Data Set Hasil Normalisasi

No	V1	V2	V3	V4	V5	...	V21
1	0.000	0.0294	1.000	0.333	0.051	...	0.000
2	0.333	0.6471	0.500	0.333	0.314	...	1.000
3	1.000	0.1176	1.000	0.667	0.102	...	0.000
...
1000	1.000	1.0000	1.000	1.000	1.000	...	1.000

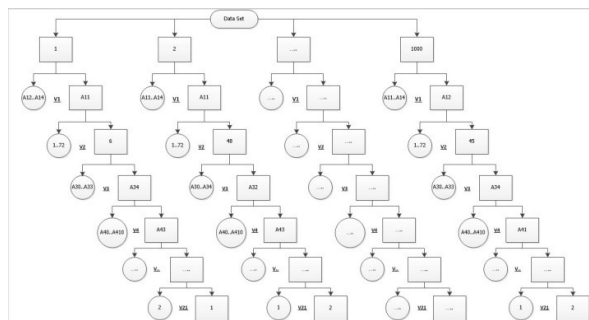
3.4. Perancangan Model

Pada penelitian ini perancangan model dilakukan menggunakan random forest dengan algoritma sebagai berikut:



Gambar 2. Algoritma *Random Forest* Menggunakan Data Set

Berdasarkan Gambar 2 algoritma *random forest* menggunakan data set dengan proses awal yaitu tersedianya data set, data set tersebut dilakukan preprocessing data dengan transformasi dan normalisasi, setelah dilakukan preprocessing data set langkah selanjutnya yaitu pemisahan data menjadi data latih dan data uji, proses tersebut akan dijelaskan dibawah sesuai Gambar 3.4 Algoritma data latih dan data Uji. Proses selanjutnya dari data latih didapatkan model dari hasil output data latih dan untuk data uji akan dilakukan proses evaluasi dan validasi sehingga didapatkan hasil performa.



Gambar 3. Algoritma data latih dan data Uji

Proses algoritma untuk data latih dan data uji sesuai dengan sample data yang telah dijelaskan sebelumnya, berdasarkan

Gambar 3 algoritma data latih dan data Uji dilakukan dengan cara melakukan input data pada variable V1 sampai dengan V20 dan untuk data output terdapat pada variable V21 dari proses tersebut didapatkan nilai *accuracy* dari data latih dan data uji.

Pada saat penentuan root dilakukan penetapan atribut yang dipilih dengan perhitungan nilai gain dari setiap atribut yang ada yaitu variabel V1 sampai dengan V20, sehingga nilai gain tertinggi akan menjadi akar pertama, sebelum mendapatkan nilai gain dari atribut agar dilakukan perhitungan terlebih dahulu untuk nilai entropy total dan setiap atribut pada variable dengan perhitungan yang terlihat pada tabel 5.

Tabel 5. Contoh Jumlah Kasus terhadap data set

Variabel	Nilai	Total Data		
		Good	Bad	
		1000	800	200
V1	0.000	1	1	0
	0.250	272	138	134
	0.333	1	0	1
	0.500	268	164	104
	0.750	63	49	14
	1.000	395	348	47
V3	0.111	1	0	1
	0.182	39	14	25
	0.386	49	21	28
	0.500	2	1	1
	0.556	11	7	4
	0.591	517	355	162
	0.778	1	0	1
	0.795	87	60	27
	1.000	293	242	51

Entropy total kasus (1000):

$$= -\left(\frac{800}{1000}\right) \times \log^2\left(\frac{800}{1000}\right) + \left(-\frac{200}{1000}\right) \times \log^2\left(\frac{200}{1000}\right) = 0.72192809488$$

Dengan perhitungan yang sama dilakukan terhadap tiap variable berdasarkan pada pengelompokan jumlah kasus pada tiap atribut pada variabelnya. Berikut contoh perhitungan entropy tiap subset Variabel V1.

1. Entropy 0.000 (1,0):
 $= -\left(\frac{1}{1000}\right) \times \log^2\left(\frac{1}{1000}\right) + \left(-\frac{0}{1000}\right) \times \log^2\left(\frac{0}{1000}\right) = 0,009965784$
2. Entropy 0.250 (138,134):
 $= -\left(\frac{138}{1000}\right) \times \log^2\left(\frac{138}{1000}\right) + \left(-\frac{134}{1000}\right) \times \log^2\left(\frac{134}{1000}\right) = 0,7828609989$
3. Entropy 0.333 (1,0):
 $= -\left(\frac{1}{1000}\right) \times \log^2\left(\frac{1}{1000}\right) + \left(-\frac{0}{1000}\right) \times \log^2\left(\frac{0}{1000}\right) = 0,009965784$
4. Entropy 0.500 (163,105):
 $= -\left(\frac{163}{1000}\right) \times \log^2\left(\frac{164}{1000}\right) + \left(-\frac{104}{1000}\right) \times \log^2\left(\frac{105}{1000}\right) = 0,7679917198$
5. Entropy 0.750 (49,14):
 $= -\left(\frac{49}{1000}\right) \times \log^2\left(\frac{49}{1000}\right) + \left(-\frac{14}{1000}\right) \times \log^2\left(\frac{14}{1000}\right) = 0,2994206587$
6. Entropy 0.100 (348,47)

$$= (-(\frac{348}{1000}) \times \log^2(\frac{348}{1000})) + (-(\frac{47}{1000}) \times \log^2(\frac{47}{1000})) = 0,7372747799$$

Selanjutnya hitung nilai gain. Contoh perhitungan gain pada Variabel V3 berdasarkan jumlah kasus per subvariabel sebagai berikut:

$$\text{Gain} = 0.72192809488 - ((\frac{1}{1000}) \times 0.00996578428 + (\frac{39}{1000}) \times 0.21926621344 + (\frac{49}{1000}) \times 0.26147882625 + (\frac{2}{1000}) \times 0.01993156856 + (\frac{11}{1000}) \times 0.08197214266 + (\frac{517}{1000}) \times 0.95581007358 + (\frac{1}{1000}) \times 0.00996578428 + (\frac{87}{1000}) \times 0.38422783446 + (\frac{293}{1000}) \times 0.71431619953 = 1.48110977246$$

Selanjutnya hitung semua nilai gain pada setiap variable untuk menentukan nilai gain yang mana yang tinggi. Nilai gain yang tinggi akan dijadikan root. Pada tabel 6 contoh dari hasil entropy dan gain.

Tabel 6. Contoh Dari Hasil Entropy Dan Gain

Variabel	Nilai	Total Data			Entropy	Gain
		1000	800	200		
V1	0.000	1	1	0	0,009965784	1.45384 707391
	0.250	272	138	134	0,7828609989	
	0.333	1	0	1	0,009965784	
	0.500	268	163	105	0,7679917198	
	0.750	63	49	14	0,2994206587	
	1.000	395	348	47	0,7372747799	
V3	0.111	1	0	1	0.00996578428	1.48110 977246
	0.182	39	14	25	0.21926621344	
	0.386	49	21	28	0.26147882625	
	0.500	2	1	1	0.01993156856	
	0.556	11	7	4	0.08197214266	
	0.591	517	355	162	0.95581007358	
	0.778	1	0	1	0.00996578428	
	0.795	87	60	27	0.38422783446	
	1.000	293	242	51	0.71431619953	

Hasil dari Tabel 6 menunjukkan bahwa nilai gain yang tertinggi adalah pada Variabel V3 yaitu 1.48110977246 maka Variabel V3 menjadi node akar (*root node*)

Gambar 4. Contoh Pohon Keputusan

3.5. Pengujian

Pada penelitian ini penulis akan membagi dua proses pengujian dengan data latih dan data uji dengan penjelasan pada Tabel 7.

Tabel 7. Data Pengujian

German Data Credit/Data Set		
	Jumlah Record	%
Data Set	1000	100
Data Latih	800	80
Data Uji	200	20

Berdasarkan Tabel 7 didapatkan informasi bahwa data set berjumlah 1000 data kreditur dengan data yang telah dijelaskan sebelumnya. Data set tersebut dibagi menjadi dua kelompok yaitu data latih dan data uji dan data tersebut dilakukan secara acak menggunakan otomatisasi sistem R.

Tabel 8. Data Latih

No	V1	V2	V3	V4	V5	...	V21
1	0.500	0.4998	0.556	0.194	0.377	...	0.500
2	1.000	0.1663	0.556	0.396	0.166	...	0.500
3	0.500	0.4164	1.000	0.094	0.284	...	1.000
...
...
800	0.500	0.6248	1.000	0.396	0.258	...	1.000

Tabel 9. Data Uji

No	V1	V2	V3	V4	V5	...	V21
1	0.500	0.1996	0.182	0.799	0.077	...	0.500
2	0.500	0.1996	0.591	0.094	0.045	...	1.000
3	0.500	0.3330	0.591	0.194	0.351	...	0.500
...
200	0.250	0.6665	1.000	0.698	0.326	...	1.000

Berdasarkan tabel 8 untuk data latih sebesar 800 data kreditur atau sebesar 80% dari total data set. Berdasarkan tabel 9 untuk data uji sebesar 200 data kreditur atau sebesar 20% dari total data sebanyak 1000 data kreditur.

3.6. Evaluasi dan Validasi

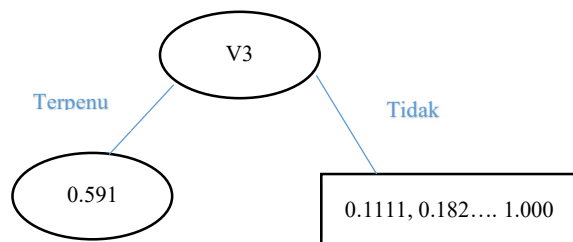
Untuk mengukur evaluasi dan validasi dari model yang dibuat menggunakan perhitungan manual dengan confusion matrix dan otomatisasi dengan sistem R sebagai berikut:

3.6.1. Confusion Matrix

Confusion Matrix berfungsi untuk melihat tingkat akurasi dari hasil yang akan diklasifikasikan secara manual dengan menggunakan tabel matrix. Apabila data set terdapat dua kelas, maka yang satu dianggap positif dan lainnya dianggap negative.

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

Berdasarkan Rumus Perhitungan Confusion matrix akan didapatkan hasil perhitungan tingkat accuracy dari model General Random Forest.



3.6.2. Confusion Matrix Sistem R

Perhitungan Confusion Matrix pada sistem R didapatkan secara otomatisasi dilakukan dengan cara melakukan input data pada sistem R dan melakukan proses sehingga didapatkan hasil confusion matrix kemudian hasil perhitungan tingkat accuracy secara otomatisasi dari model General Random Forest.

3.6.3. Receiver Operating Characteristic (ROC)

Kurva ROC (Receiver Operating Characteristics) yaitu suatu alat ukur performance dari classification problem untuk menentukan threshold suatu model. ROC dapat menampilkan output berupa akurasi serta membandingkan klasifikasi secara visual. ROC dapat disebut juga sebagai tampilan dari confusion matrix. ROC akan menampilkan grafik dua dimensi dengan false positive sebagai garis horizontal dan true sebagai garis vertical.

3.6.4. Precision-Recall Curve

Precision and recall dalam system R outputya menjadi satu, definisi precision yaitu rasio item relevan yang dipilih terhadap seluruh item yang terpilih. Presisi memiliki arti sebagai kesamaan dari permintaan informasi dengan jawaban atas permintaan tersebut, sedangkan definisi *Revall Curve* yaitu rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang ada dan output kurva dari precision-recall curve didapatkan secara otomatisasi sistem R.

3.6.4.1. AUC

AUC berfungsi untuk menghitung performa *General Random Forest* terhadap *german data credit* dengan mengguakan sample data sebanyak 20% atau 200 dari total 1000 data.

Tabel 10. Klasifikasi AUC

AUC	Interpretation
1.0 (100%)	Perfect Model
0.9 – 0.99 (90 - 99%)	Excellent Model
0.8 – 0.89 (80 – 89%)	Very Good Model
0.7 – 0.79 (70 – 79%)	Fair Model
0.51 – 0.69 (51 – 69%)	Poor Model
< 0.5 (50%)	Worthless Model

Berdasarkan tabel 10 didapatkan informasi bahwa AUC akan menghitung nilai yang berada dibawah grafik ROC dan Nilai ROC yang mendekati nilai satu maka akan semakin baik.

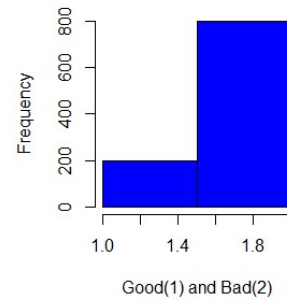
4. PEMBAHASAN

4.1. Data Penelitian

Data penelitian bersumber dari *machine learning repository* dari *german credit data*. Data yang digunakan sebanyak 1000 yang terdiri dari 21 *variable* (data *input* V1 sampai dengan V20 dan data *output* pada V21). Variable target terdiri dari 2 kategori yaitu 1 *Good* dan 2 *Bad*.

Tabel 11. Tabel Persentase Variable Target

	Count	Percentage
Bad	200	20
Good	800	80



Gambar 6. Grafik Variable Target

Berdasarkan Gambar 6 didapatkan informasi bahwa untuk data good credit sebesar 800 kreditur atau sebesar 80% dan data bad sebesar 200 kreditur atau sebesar 20% sehingga total secara keseluruhan sebesar 1000 data. Struktur data set dan *summary* data yang digunakan pada penelitian ini terdapt pada tabel 12.

Tabel 12. Struktur Data Set

No	V1	V2	V3	V4	V5	..	V21
1	A11	6	A34	A43	1169	..	1
2	A12	48	A32	A43	5951	..	2
3	A14	12	A34	A46	2096	..	1
..
1000	A12	45	A34	A41	4576	..	1

Table 12 berfungsi untuk dapat melihat informasi seperti jumlah int, tipe, factor dan target variable, untuk fungsi str pada dataset dapat dilihat struktur dari dataset yang digunakan.

```
chk_ac_status_1 duration_month_2 credit_history_3 purpose_4 credit_amount
5
0 A11: 274 Min. : 4.0 A30: 40 A43 : 280 Min. : 25
6 A12: 269 1st Qu.: 12.0 A31: 49 A40 : 234 1st Qu.: 136
7 A13: 63 Median : 18.0 A32: 530 A42 : 181 Median : 232
0 A14: 394 Mean : 20.9 A33: 88 A41 : 103 Mean : 327
1 3rd Qu.: 24.0 A34: 293 A49 : 97 3rd Qu.: 397
2 Max. : 72.0 A46 : 50 Max. : 1842
4 (Other): 55
savings_ac_bond_6 p_employment_since_7 instalment_pct_8 personal_status_9
A61: 603 A71: 62 Min. : 1.000 A91: 50
A62: 193 A72: 172 1st Qu.: 2.000 A92: 310
A63: 63 A73: 359 Median : 3.000 A93: 548
A64: 48 A74: 174 Mean : 2.973 A94: 92
A65: 183 A75: 253 3rd Qu.: 4.000
Max. : 4.000
other_debtors_or_grantors_10 present_residence_since_11 property_type_12
A101: 907 Min. : 1.000 A121: 282
A102: 41 1st Qu.: 2.000 A122: 332
A103: 52 Median : 3.000 A123: 332
Mean : 2.845 A124: 154
3rd Qu.: 4.000
Max. : 4.000
age_in_yrs_13 other_instalment_type_14 housing_type_15 number_cards_this_16
Min. : 19.00 A141: 139 A151: 179 Min. : 1.000
1st Qu.: 27.00 A142: 47 A152: 713 1st Qu.: 1.000
Median : 33.00 A143: 814 A153: 108 Median : 1.000
Mean : 35.55 Mean : 1.407
3rd Qu.: 42.00 Max. : 4.000
Max. : 75.00
job_17 no_people_liable_for_maintenance_18 telephone_19 foreign_worker_20
A171: 22 Min. : 1.000 A191: 596 A201: 963
A172: 200 1st Qu.: 1.000 A192: 404 A202: 37
A173: 630 Median : 1.000
A174: 148 Mean : 1.155
3rd Qu.: 1.000
Max. : 2.000
good_bad_21
Min. : 1.0
1st Qu.: 1.0
Median : 1.0
Mean : 1.3
3rd Qu.: 2.0
Max. : 2.0
```

Gambar 7. Summary Data Set

Gambar 7 rangkuman secara statistika dengan menggunakan fungsi Summary (dataset) yang terdiri dari variable Chk_ac_Status_1 s.d. Foreign_Worker_20 dan Variabel Target pada good_bad_21.

4.2. Preprocessing Data

Preprocessing data wajib dilakukan sebelum melakukan pelatihan data, hal tersebut perlu dilakukan hasilnya yang lebih akurat. Pada penelitian ini menggunakan transformasi dan normalisasi.


```
'data.frame': 1000 obs. of 21 variables:
 $ chk_ac_status_1 : Factor w/ 4 levels "A11","A12","A13",...
 1 2 4 1 1 4 4 2 4 2 ...
 $ duration_month_2 : num 6 48 12 42 24 36 24 36 12 30 ...
 $ credit_history_3 : Factor w/ 5 levels "A30","A31","A32",...
 5 3 5 3 4 3 3 3 3 5 ...
 $ purpose_4 : Factor w/ 10 levels "A40","A41","A410",...
 .: 5 5 8 4 1 8 4 2 5 1 ...
 $ credit_amount_5 : num 1169 5951 2096 7882 4870 ...
 $ savings_ac_bond_6 : Factor w/ 5 levels "A61","A62","A63",...
 5 1 1 1 5 3 1 4 1 ...
 $ p_employment_since_7 : Factor w/ 5 levels "A71","A72","A73",...
 5 3 4 4 3 3 5 3 4 1 ...
 $ instalment_pct_8 : num 4 2 2 2 3 2 3 2 2 4 ...
 $ personal_status_9 : Factor w/ 4 levels "A91","A92","A93",...
 3 2 3 3 3 3 3 1 4 ...
 $ other_debtors_or_grantors_10 : Factor w/ 3 levels "A101","A102",...: 1 1
 1 3 1 1 1 1 1 ...
 $ present_residence_since_11 : num 4 2 3 4 4 4 4 2 4 ...
 $ property_type_12 : Factor w/ 4 levels "A121","A122",...: 1 1
 1 2 4 4 2 3 1 3 ...
 $ age_in_yrs_13 : num 67 22 49 45 53 35 53 35 61 28 ...
 $ other_instalment_type_14 : Factor w/ 3 levels "A141","A142",...: 3 3
 3 3 3 3 3 3 3 ...
 $ housing_type_15 : Factor w/ 3 levels "A151","A152",...: 2 2
 2 3 3 3 2 1 2 2 ...
 $ number_cards_this_bank_16 : num 2 1 1 1 2 1 1 1 1 2 ...
 $ job_17 : Factor w/ 4 levels "A171","A172",...: 3 3
 2 3 3 2 3 4 2 4 ...
 $ no_people_liable_for_mntnace_18 : num 1 1 2 2 2 2 1 1 1 1 ...
 $ telephone_19 : Factor w/ 2 levels "A191","A192": 2 1 1
 1 1 2 1 2 1 1 ...
 $ foreign_worker_20 : Factor w/ 2 levels "A201","A202": 1 1 1
 1 1 1 1 1 1 ...
 $ good_bad_21 : Factor w/ 2 levels "Bad","Good": 2 1 2 2
 1 2 2 2 2 1 ...
```

Gambar 8. Struktur Data Hasil Tranformasi

Pada gambar 8 berfungsi untuk mengubah data dengan tipe *numeric* menjadi factor pada data *input* dan tipe *int* menjadi *factor* pada variable target sehingga struktur data yang ada dapat digunakan untuk proses pelatihan sesuai Tabel 13.

Tabel 13. Data Set Hasil Tranformasi

No	V1	V2	V3	V4	V5	...	V21
1	1	6	5	4	1169	...	1
2	2	48	3	4	5951	...	2
3	4	12	5	7	2096	...	1
...
1000	2	45	5	2	4576	...	2

Langkah selanjutnya yaitu tahap normalisasi berdasarkan data pada Tabel 13 Data Set Hasil Tranformasi. Data tersebut akan dilakukan normalisasi dengan metode Min Max secara otomatisasi dengan contoh perhitungan berikut:

- Baris V1, Kolom 1 = $(1 - 1) / (4 - 1) = 0$
- Baris V2, Kolom 1 = $(6 - 4) / (72 - 4) = 0.0294$
- Baris V3, Kolom 1 = $(4 - 0) / (4 - 0) = 1$
- Baris V4, Kolom 1 = $(12 - 2) / (184 - 2) = 0.3334$
- Baris V5, Kolom 1 = $(5 - 1) / (5 - 1) = 0,051$
- Baris V6, Kolom 1 = $(5 - 1) / (5 - 1) = 1$

Tabel 14. Data Set Hasil Normalisasi

No	V1	V2	V3	V4	V5	...	V21
1	0.000	0.0294	1.000	0.333	0.051	...	0.000
2	0.333	0.6471	0.500	0.333	0.314	...	1.000
3	1.000	0.1176	1.000	0.667	0.102	...	0.000
...
1000	1.000	1.0000	1.000	1.000	1.000	...	1.000

Data pada Tabel 14 Data Set Hasil Normalisasi adalah hasil dari proses Normalisasi data dengan menggunakan metode min max sesuai rumus diatas dengan 1000 data.

4.3. Hasil Pengujian

Proses Latih dan Uji menggunakan 1000 data kreditur. Untuk proses pelatihan menggunakan 800 data kreditur atau sebanyak 80% dari total data kreditur, sedangkan untuk pengujian menggunakan 200 data kreditur atau sebanyak 20% dari total data

kreditur. Data yang digunakan untuk pelatihan dan Pengujian dilakukan dengan cara acak menggunakan system R.

Tabel 15. Data Latih

No	V1	V2	V3	V4	V5	V8	...	V21
1	0.500	0.4998	0.556	0.194	0.377	0.455	...	0.500
2	1.000	0.1663	0.556	0.396	0.166	0.455	...	0.500
3	0.500	0.4164	1.000	0.094	0.284	1.000	...	1.000
...
800	0.500	0.6248	1.000	0.396	0.258	1.000	...	1.000

Data Pelatihan menggunakan data pada table 15 Data Latih yang telah berhasil dibuat dengan 21 variable yang terdiri dari 20 Input dan 1 output dengan 800 data kreditur di lakukan secara acak dengan hasil dan setelah data pelatihan tersebut selesai maka akan dilanjutkan pada proses pengujian.

Tabel 16. Data Uji

No	V1	V2	V3	V4	V5	...	V21
1	0.500	0.1996	0.182	0.799	0.077	...	0.500
2	0.500	0.1996	0.591	0.094	0.045	...	1.000
3	0.500	0.3330	0.591	0.194	0.351	...	0.500
...
200	0.250	0.6665	1.000	0.698	0.326	...	1.000

Tabel 17. Persentase Data Pengujian

	Count	Percentage
Bad	40	20
Good	160	80

Data Pengujian menggunakan data pada tabel 16 yang telah berhasil dibuat dengan 21 variable yang terdiri dari 20 Input dan 1 output dengan 200 data kreditur di lakukan secara acak dengan hasil dan persentase terdapat pada tabel 17 dan setelah data pengujian tersebut selesai maka akan dilanjutkan pada proses perhitungan Confusion Matrix dan digambarkan performance dari hasil ROC.

Tabel 18. Confusion Matrix

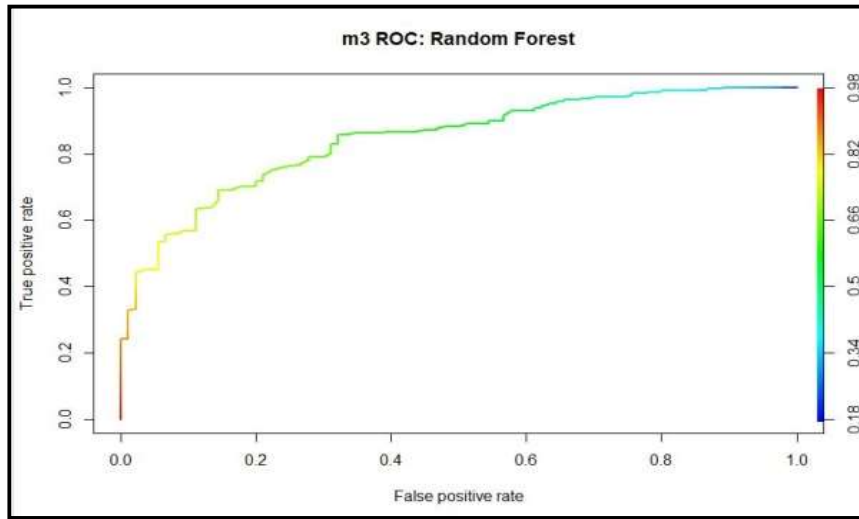
	True Class		
	Positive	Negative	
Predicted Class	Positive	159	33
	Negative	1	7

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{159 + 7}{159 + 7 + 1 + 33} = \frac{166}{200} = 0,83$$

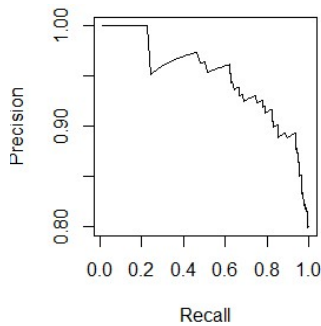
Perhitungan dari tabel 17 didapat dari data pengujian sebanyak 200 Data Kreditur dan mendapatkan perhitungan sebesar 0,83 atau sebesar 83%. Perhitungan *Confusion Matrix* menggambarkan hasil perhitungan dari ROC yang akan dijelaskan selanjutnya.

Berdasarkan Gambar 9 Kurva ROC (*Receiver Operating Characteristics*) diatas berfungsi sebagai alat ukur performance dari classification problem dalam menentukan threshold dari

suatu model. ROC berfungsi untuk menampilkan akurasi dan membandingkan klasifikasi secara visual dan mengekspresikan *confusion matrix*.

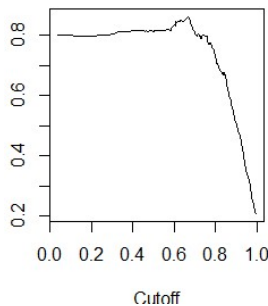


Gambar 9. Kurva ROC



Gambar 10. Precision/Recall Curve

Berdasarkan Gambar 10 Precision/Recall Curve diatas didapatkan informasi gambar kurva yang berfungsi sebagai alat prediksi dari perhitungan confusion matrix berdasarkan data positif dan prediksi positif.



Gambar 11. Accuracy as function of threshold

Perhitungan performa pada penelitian ini menggunakan AUC, AUC menghitung nilai yang berada dibawah kurva ROC dengan persamaan yang telah dijelaskan sebelumnya. Hasil AUC dari model Random Forest memiliki tingkat akurasi sebesar 0,83 atau sebesar 83%.

Tabel 19. Klasifikasi AUC

AUC	Interpretation
1.0 (100%)	Perfect Model

0.9 – 0.99 (90 – 99%)	Excellent Model
0.8 – 0.89 (80 – 89%)	Very Good Model
0.7 – 0.79 (70 – 79%)	Fair Model
0.51 – 0.69 (51 – 69%)	Poor Model
< 0.5 (50%)	Worthless Model

Dengan Metode *Random Forest* terhadap german data credit berdasarkan tabel 19 didapatkan hasil Akurasi sebesar 83% sehingga termasuk pada nilai *Very Good Model* dengan nilai 0.8–0.89 (80-89%).

5. KESIMPULAN

Analisa Kelayakan Kredit Menggunakan General Random Forest (Studi Kasus German Credit Data) berhasil dilakukan. Hasil dari pengujian mendapatkan nilai performa dengan menggunakan AUC mendapatkan nilai sebesar 0,83 (83%) dan termasuk ke dalam *Very Good Model*.

DAFTAR PUSTAKA

- [1] Indriati, Vebian. Zuhroh, Idah. Susilowat, Dwi. 2018. Analisis penyaluran kredit modal kerja pada bank umum di Indonesia. *Jurnal Ilmu Ekonomi*, Vol 2 (3): 529-540.
- [2] Ramelda, Susi. 2017. Pengaruh suku bunga kredit dan produk domestik bruto terhadap penyaluran kredit perbankan bank umum pemerintah di indonesia, *JOM Fekon* Vol 4 (1).
- [3] Safitri, Yuliana. 2015. Analisis Kelayakan Pemberian Kredit Usaha Rakyat (KUR) pada PT. Bank Rakyat Indonesia (Persero) Tbk Unit Air Putih Cabang Samarinda. Fakultas Ekonomi Universitas 17 Agustus 1945 Samarinda.
- [4] Siringoringo, Renniwaty. 2017. Fungsi Intermediasi Perbankan Indonesia (Studi Kasus Bank Umum Konvensional yang Tercatat di BEI Periode 2012-2016), *Jurnal Inspirasi Bisnis dan Manajemen*, Vol 1 (2): 135-144

- [5] Vanissa Wanika Siburian Ika Elvina Mulyana. 2018. Prediksi Harga Ponsel Menggunakan Metode Random Forest. Prosiding Annual Research Seminar 2018 Vol 4 (1).
- [6] Alfaro, Ibnu. Tutupoly, Agasya, Taransa. Suryanto, Ade. 2018. Komparasi algoritma c4.5, naive bayes, dan random forest untuk klasifikasi data kelulusan mahasiswa jakarta
- [7] Hofmann, H. 2000. Statlog (German Credit Dataset) <https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german> diakses tanggal 15 Januari 2020 pukul 13.56 WIB
- [8] S. Pudaruth. 2014. Predicting the price of used cars using machine learning techniques. Int, J, Inf, Comput, Technol. Vol 4 (7). 753–764.
- [9] Rismayanti. 2018. Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus: Fakultas Teknik dan Komputer Universitas Harapan Medan). Vol 2 (1)
- [10] Wahyudin. 2009. Metode Iterative Dischotomizer 3 (ID3) untuk Penyeleksian Penerimaan Mahasiswa Baru, Jurnal Pendidikan Teknologi Informasi dan Komunikasi (PTIK). Vol 1 (2): 5-15.
- [11] Suswanto, Deni. 2016. Analisis Perbandingan Metode Machine Learning Pada Prediksi Khasiat Jamu [Skripsi]. Jawa Barat (ID): Institut Pertanian Bogor.
- [12] Shannon, C.E. 2001. A mathematical theory of communication. Mobile Computing and Communications Review Vol 5 (1).
- [13] Rahayu, Sri, Erna. Wahono, Satria, Romi. Supriyanto, Catur. 2015. Penerapan Metode Average Gain, Threshold Pruning dan Cost Complexity Pruning untuk Split Atribut pada Algoritma C4.5. Journal of Intelligent Systems, Vol 1 (2).
- [14] Nugroho, Sulisty, Yusuf. Emiliyawati, Nova. 2017. Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. Jurnal Teknik Elektro Vol 9 (1)
- [15] Alpha P, A., & Oslan, Y. 2015. Program Bantu Pemilihan Pakaian dan Bahan Batik Bagi Konsumen dengan Pendekatan Decision Tree Studi Kasus : Toko InBATIK. Jurnal EKSIS Vol 8 (1): halaman 37-46
- [16] Han, Jiawei., Micheline Kamber., & Jian Pei. 2012. Data Mining Concepts and Techniques. Elsevier Inc.
- [17] Zefriyenni. Yuliana, Ufi, Ira. 2016. Kebijakan Pemberian Kredit Terhadap Penetapan Jumlah Kredit (Studi Khusus Pada Ued-Sp Amanah Sejahtera Sungai Buluh kecamatan Singingi Hilirkabupaten Kuantan Singingi Propinsi Riau). Penelitian Bidang Komputer Sains dan Pendidikan Informatika Vol 1 (1) : 72-80.
- [18] Bischl, Bernd et.all. 2016. Mlr : Machine Learning in R. *Journal of Machine Learning Research*. 17(2016) : 1-5



Emy Haryatmi menyelesaikan Pendidikan jenjang sarjana jurusan Sistem Komputer di Universitas Gunadarma, Indonesia pada tahun 2000. Pada tahun 2004 berhasil menyelesaikan jenjang master jurusan Telecommunication and Networking di Curtin University of Technology, Perth, Australia. Pada tahun 2016 berhasil menyelesaikan jenjang Doktor Teknologi Informasi di Universitas Gunadarma, Indonesia. Topik penelitian yang dialami adalah bidang telekomunikasi, jaringan, keamanan jaringan, image processing, data mining dan internet of things.

BIODATA PENULIS



Budi Prasajo lahir di Jakarta, Indonesia. Menyelesaikan pendidikan sarjana komputer jurusan Sistem Informasi di Universitas Gunadarma. Memiliki pengalaman bekerja di perusahaan Perbankan di Satuan Kerja Audit Intern BRI periode 2016-2017 dan Periode 2017 sampai dengan Saat ini bekerja di perusahaan Perasuransian di Satuan Kerja Audit Intern BRI Life.