**Laporan Akhir Tugas MK. Data Mining (KOM332), Semester Genap 2018/2019**

**Prediksi Kelayakan Pemohon Kredit Dengan Metode Klasifikasi Menggunakan Algoritma Decision Tree**

Yusuf Ibadurrahman Assidiq (G64160002), Rheisa Gusmendasari (G64160035) , Vito Melvin Maratur Octavian Ompusunggu (G64160032),[[1]](#footnote-1)\*

Kelompok: 11, Kelas Paralel: 1

**Abstrak**

Bank adalah lembaga keuangan yang dapat memberikan kredit. Kredit adalah penyediaan uang atau tagihan berdasarkan kesepakatan pinjam meminjam antara pihak bank dengan pihak lain dengan ketentuan-ketentuan tertentu. Salah satu permasalahan dari program kredit pinjaman adalah nasabah yang lalai dalam pembayaran angsuran. Untuk itu, pihak bank harus dapat menyeleksi calon nasabah yang tepat untuk meminimalisir kerugian. Penggunaan algoritma *decision tree* dapat memudahkan pihak bank untuk menentukan calon nasabah yang berkemungkinan kecil untuk menimbulkan kerugian bagi bank. Data yang digunakan adalah *Statlog German Credit* dari UCI Repositories dengan distribusi data yang tidak merata. Akurasi yang didapatkan adalah sebesar 71.85 % menggunakan *teknik simplified random sampling* dan sebesar 74.09% menggunakan teknik *stratified random sampling..*

Kata kunci : *Decision Tree*, klasifikasi, kredit, prediksi.

**PENDAHULUAN**

**Latar Belakang**

Bank adalah lembaga keuangan yang dapat memberikan kredit. Kredit adalah penyediaan uang atau tagihan berdasarkan kesepakatan pinjam meminjam antara pihak bank dengan pihak lain yang mewajibkan pihak peminjam untuk melakukan pembayaran dengan jumlah bunga sebagai imbalan secara periodik sesuai kesepakatan jangka waktu yang telah disepakati. Sebagai lembaga keuangan bank dapat memberikan bantuan kepada nasabah yang membutuhkan pinjaman dengan memberikan kredit pinjaman kepada nasabahnya. Bank akan memberikan kredit pinjaman kepada nasabah yang dianggap mampu melunasi kredit setiap bulan sesuai dengan perjanjian dengan kedua belah pihak.

Namun demikian, terdapat sejumlah permasalahan yang muncul dari program kredit pinjaman. Salah satunya adalah adanya nasabah yang telat membayar angsuran. Oleh karena itu, bank harus dapat memilih dengan benar calon nasabah yang akan meminta kredit pinjaman agar resiko kerugian bank dapat diminimalkan.

Ketika sebuah bank menerima permohonan pinjaman dari seseorang calon nasabah, bank harus membuat keputusan apakah menerima untuk melanjutkan proses pinjaman atau tidak berdasarkan profil pemohon. Profil tersebut dapat digunakan sebagai referensi dalam keputusan pemberian pinjaman kepada calon kreditor tersebut. Dalam pemberian pinjaman, instansi perbankan memiliki dua risiko, yaitu risiko dengan kreditor yang baik dan risiko dengan kreditor yang buruk. Jika kreditur memiliki risiko kredit yang baik dan rutin membayarkan pinjamannya, maka keputusan bank untuk tidak memberi pinjaman membuat bank merugi karena kehilangan kesempatan memperoleh keuntungan. Akan tetapi, jika kreditur memiliki risiko kredit yang buruk, maka keputusan bank untuk memberi pinjaman akan menjadikan bank mengalami kerugian. Hasil keputusan akan lebih buruk jika nasabah yang tidak sesuai kriteria diklasifikasikan dalam nasabah yang layak mendapatkan kredit.

**Tujuan**

Tujuan dari penelitian ini adalah memprediksi kelayakan pemohon kredit menggunakan metode klasifikasi dengan algoritma decision tree pada data yang telah memiliki kelas kelayakan pemohon untuk diberikan pinjaman atau melanjutkan permohonan kredit bagi nasabah dengan kriteria tertentu.

**Ruang Lingkup**

Sumber data yang digunakan adalah german credit card dataset yang berasal dari *UCI Repository* yang memiliki 20 atribut dan 1000 *instance.* Teknik yang digunakan adalah penerapan metode klasifikasi dengan algoritma decision tree. Pemodelan decision tree menggunakan *package party* pada R

**Manfaat**

Manfaat dari penelitian ini adalah hasil dapat digunakan untuk decision support system oleh analis kredit dan lembaga keuangan dalam penentuan kelayakan kredit bagi seorang calon nasabah.

**TINJAUAN PUSTAKA**

**Metode Klasifikasi**

Data mining menurut Han dkk. (2012) adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. Salah satu metode yang dapat diterapkan dalam permohonan kredit, yaitu klasifikasi. Menurut Han J, et al. (2012) Klasifikasi adalah suatu proses yang digunakan untuk menemukan model (atau fungsi) dengan menggambarkan dan membedakan kelas data atau konsep. Terdapat beberapa algoritma klasifikasi data salah satunya yaitu pohon keputusan atau decision tree.

**Decision Tree**

Decision tree merupakan salah satu metode klasifikasi data mining yang populer digunakan karena mudah diinterpretasikan oleh manusia dengan konsep dasar mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (Witten 2011), mudah mengintegrasikan dengan sistem basis data dan memiliki tingkat ketelitian (akurasi) yang baik (Chen 2008). Peranan pohon keputusan sebagai alat bantu pengambilan keputusan (decision support tool) telah dikembangkan oleh manusia untuk membantu mencari dan membuat keputusan masalah dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut.

Decision tree dapat membagi kumpulan data yang besar menjadi himpunan- himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan (rules), sehingga anggota himpunan hasil menjadi mirip satu dengan lainnya dengan memperhatikan pada variabel tujuannya (Berry 2004).

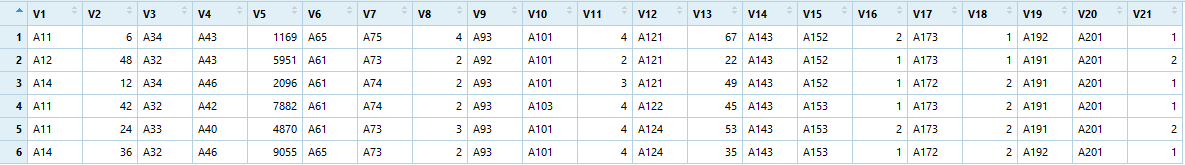
**Kelayakan Pemohon Kredit**

Penilaian yang tepat atas kelayakan pemohon kredit memungkinkan lembaga keuangan meningkatkan volume kredit namun juga meminimalkan kemungkinan kerugian. Data pemohon kredit yang besar dapat diolah menjadi informasi yang dapat digunakan sebagai pendukung keputusan dalam menentukan permohonan kredit. Pengolahan data tersebut termasuk dalam bidang data mining.

**METODE**

**Data**

Data yang akan diproses diambil dari UCI Repository dan berjudul German Credit Card. Data merepresentasikan kategori yang diperlukan untuk memilih nasabah yang layak diberikan kredit oleh bank. Data terdiri dari 20 atribut dan 1 atribut kelas serta data memiliki 1000 records. Tipe data berbentuk kategorik terdiri dari 13 atribut dan numerik terdiri dari 7 atribut serta memiliki satu kelas yang berbentuk kategorik. Gambaran data dapat dilihat pada gambar 1.



Gambar 1 Gambaran data secara umum

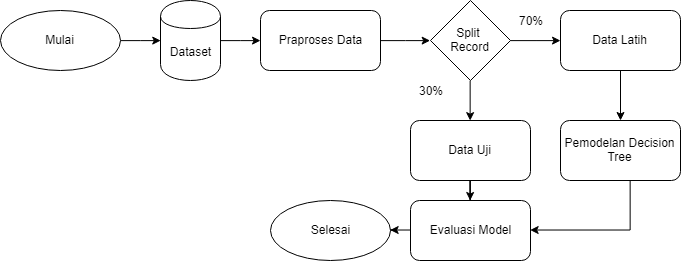
Atribut didominasi oleh tipe data kategorik dibandingkan dengan tipe data numerik. Penjelasan atribut yang dimiliki oleh data german credit card dapat dilihat pada table 1.

Tabel 1 Penjelasan atribut

|  |  |  |  |
| --- | --- | --- | --- |
| NO | ATRIBUT | PENJELASAN | Tipe Data |
| 1. | *Checking. Account* | Status dan batas maksimum kepemilikan akun sebelumnya | Kategorik |
| 2. | *Month* | Durasi pembayaran angsuran yang dilakukan dalam satuan bulan | Numerik |
| 3. | *Credit.History* | Riwayat pembayaran kredit sebelumnya | Kategorik |
| 4. | *Purpose* | Tujuan pengajuan kredit | Kategorik |
| 5. | *Credit.Amount* | Jumlah nominal kredit yang diajukan | Numerik |
| 6. | *Saving.Accounts* | Jumlah nominal akun tersimpan yang dimiliki | Kategorik |
| 7. | *Present* | Durasi pekerjaan yang dijalani saat ini (tahun) | Kategorik |
| 8. | *Installment.Rate* | Rencana tingkat kredit yang diajukan | Numerik |
| 9. | *Status* | Status pernikahan | Kategorik |
| 10. | *Other.Debtors* | Status penjamin | Kategorik |
| 11. | *Residence* | Durasi penempatan tempat tinggal terhitung sampai saat ini. (tahun) | Numerik |
| 12. | *Property* | Aset yang dimiliki saat ini | Kategorik |
| 13. | *Age* | Usia pemohon saat ini | Numerik |
| 14. | *Installment.Plans* | Rencana jenis kredit yang diajukan | Kategorik |
| 15. | *Housing* | Status kepemilikan rumah | Kategorik |
| 16. | *Number.of.Credits* | Jumlah kepemilikan kredit saat ini | Numerik |
| 17. | *Job* | Pengelompokan pekerja berdasarkan keahlian atau jabatan | Kategorik |
| 18. | *Liable* | Jumlah penjamin | Numerik |
| 19. | *Telephone* | Status kepemilikan nomor telepon | Kategorik |
| 20. | *Foreign.Worker* | Status domisili pekerjaan | Kategorik |
| 21. | *Class* | Kelayakan pemohon kredit | Kategorik |

**Tahapan Kegiatan**

Penelitian diawali dengan menyiapkan data yang akan digunakan. Setelah itu dilakukan praproses data dengan melakukan pembersihan dan penyesuaian data. Praproses yang dilakukan adalah penamaan atribut dan penamaan *instance* sesuai dengan penjelasan data pada UCI Repository. Langkah selanjutnya adalah pemisahan *record* dengan pembagian 70% untuk data latih dan 30 % untuk data uji. Selanjutnya pembuatan model decision tree untuk menentukan kelayakan kredit. Setelah itu adalah evaluasi dengan mencari akurasi penerapan algoritma decision tree untuk data tersebut. Tahap knowledge discovery in database (KDD) guna mendapatkan knowledge dalam data set dapat dilihat pada gambar 2.



Gambar 2 Tahapan Kegiatan

**Lingkungan Pengembangan**

Perangkat Keras :

1. Processor AMD FX-7600P Radeon R7

2. Memori 6 GB

3. HDD 1000 GB

4. Layar 15.6 inci

5. Mouse dan Keyboard

Perangkat Lunak :

1. Sistem operasi Windows 10

2. R studio 1.1.463

3. R 3.5.2

**HASIL DAN PEMBAHASAN**

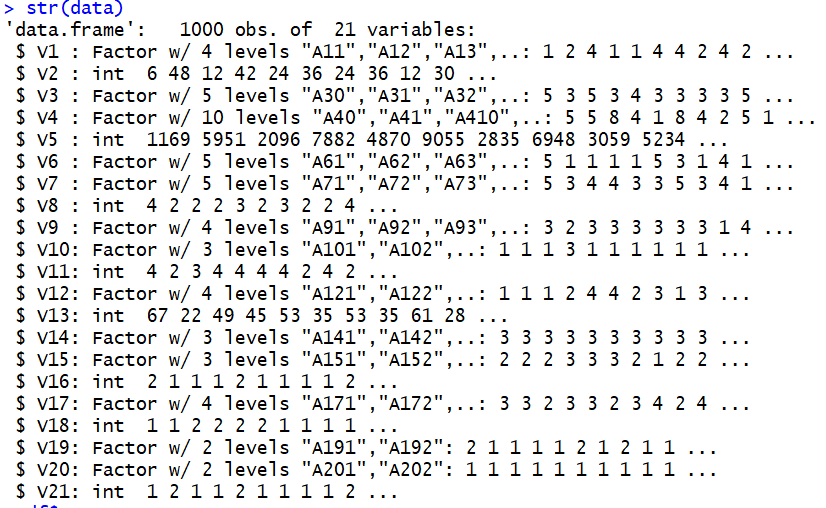
**Praproses Data**

Pada data yang didapatkan nama atribut masih belum didefinisikan sehingga sulit untuk dimengerti. Selain itu, nilai atau data dari tiap atribut masih berbentuk kode yang dijelaskan dalam deskripsi pada UCI Repository. Oleh karena itu, diperlukan praproses data dengan mengganti nama atribut dan data pada atribut sesuai dengan deskripsi/

Sebelum melakukan praproses data, dapat dilihat terlebih dahulu statistika ringkasan dan tipe data dari tiap fitur pada gambar 3 dan gambar 4.

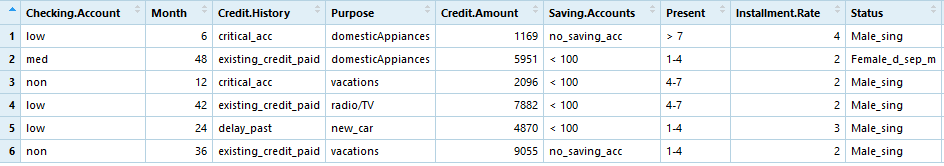
|  |
| --- |
|  |

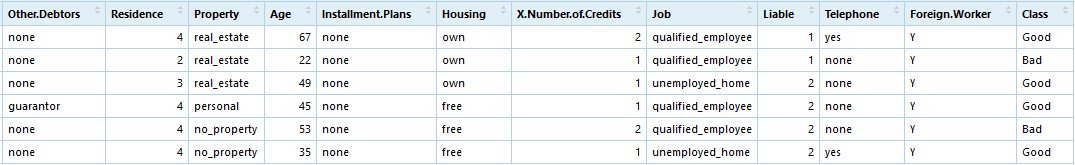
Gambar 3 Statistika Ringkasan



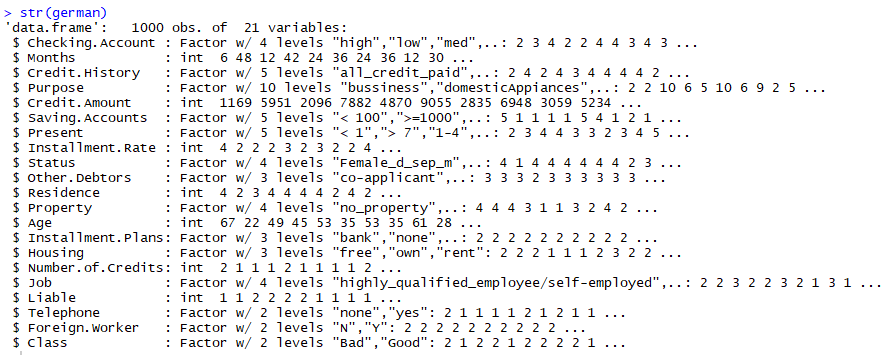
Gambar 4 Ringkasan Tipe Data

Praproses data diawali dengan perubahan nama fitur dan nilai pada tiap fitur berdasarkan informasi atribut yang disertakan dengan data tersebut. Data yang sudah memiliki nama atribut dapat dilihat pada gambar 5.



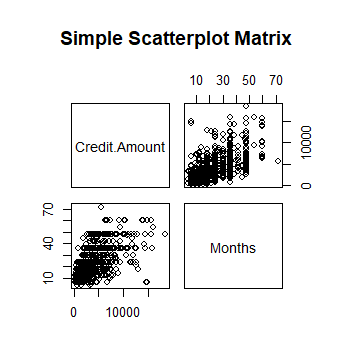


Gambar 5 Data hasil praproses



Gambar 6 Hasil Transformasi Data

Pengukuran korelasi untuk tiap atribut menghasilkan yang korelasi tertinggi dimiliki oleh atribut Credit.Amount dan Months yaitu sebesar 0,62. Sehingga tidak ada data yang perlu direduksi.



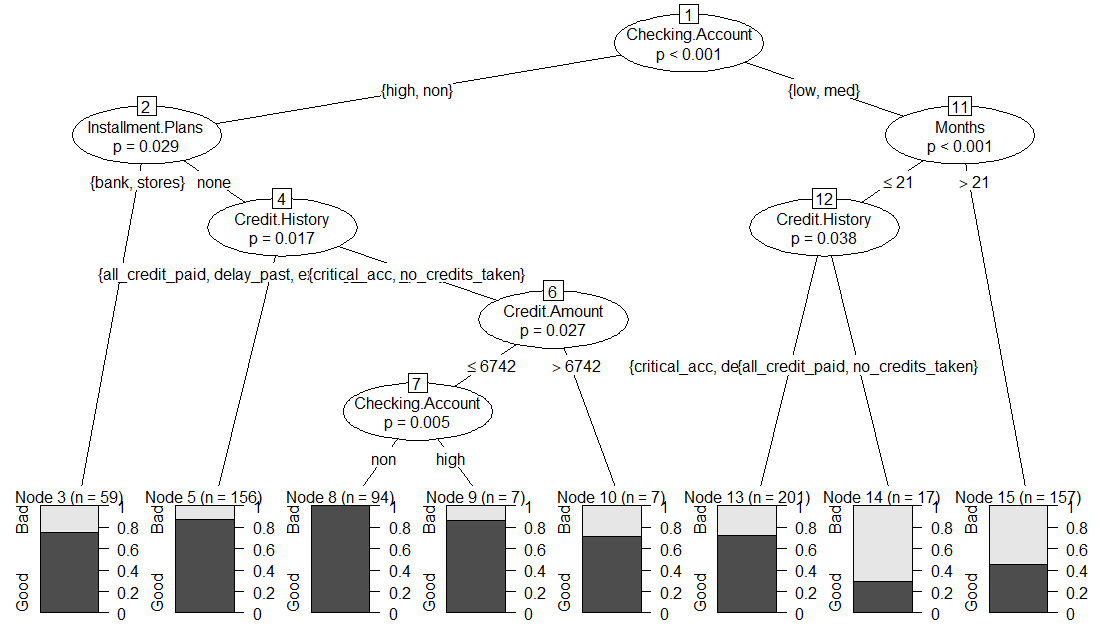
Gambar 7 Scatter Plot dengan korelasi tertinggi

**Pemisahan Data**

Pada tahap pemisahan data dilakukan dua teknik sampling yang berbeda, yaitu dengan menggunakan *simple random sampling* dan *stratified random sampling*. Teknik *Stratified random sampling* memisahkan data berdasarkan kelas dengan proporsi yang sama untuk tiap kelas. Datadipisahkan menjadi data latih dan data uji dengan jumlah 70% data latih dan 30% data uji untuk tiap-tiap teknik sampling.

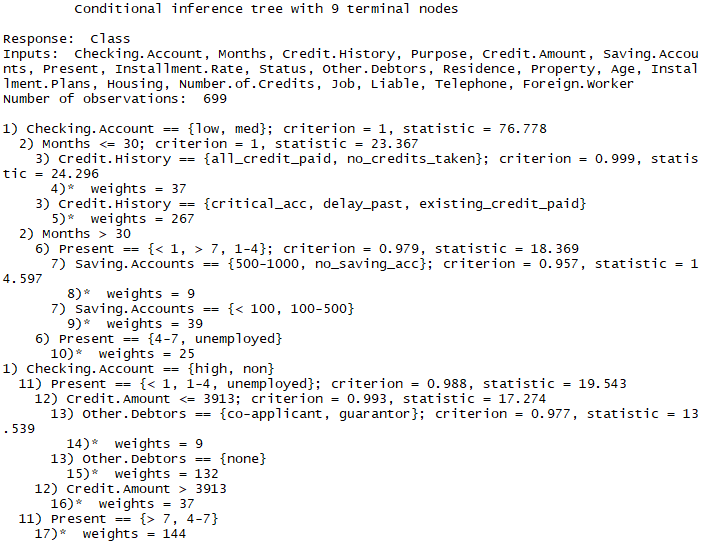
**Pemodelan Decision Tree**

Algoritma *Decision Tree* digunakan dalam membangun model pada dataset ini. Decision Tree digunakan karena sebagian besar atribut pada dataset berupa data kategorik dan untuk data numerik akan dilakukan pemisahan secara langsung dengan package party yang digunakan dalam R studio. Pemodelan decision tree dilakukan pada data latih. Hasil pemisahan data dengan *simple random sampling* dilakukan pemodelan decision tree dan dapat dilihat pada gambar 8.

**

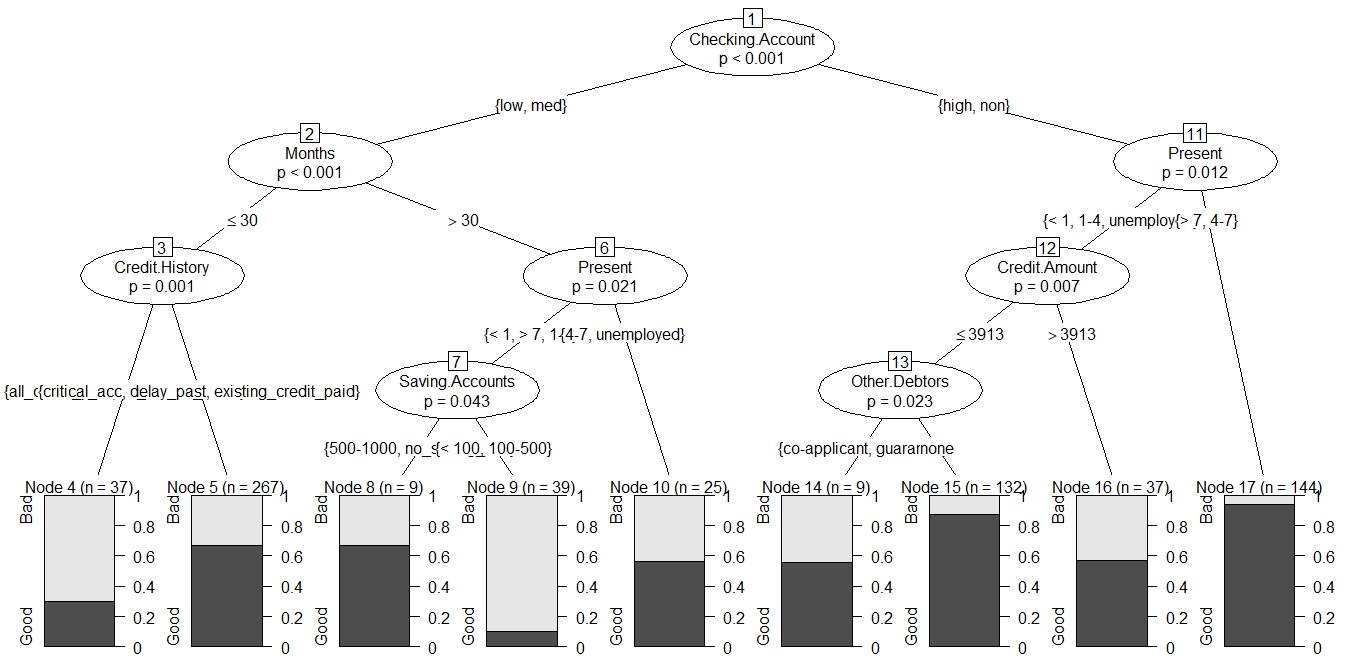
Gambar 8 Decison Tree Simple Random Sampling

Hasil pemodelan decision tree dengan *simple random sampling* dapat dilihat pada gambar 9.



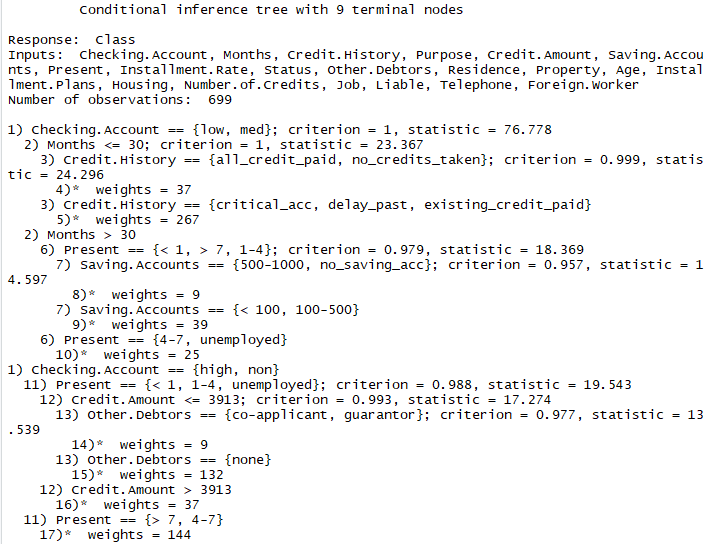
Gambar 9 Pemodelan Decision Tree dengan Simple Random Sampling

Hasil pemisahan data dengan *stratified random sampling* dilakukan pemodelan decision tree dapat dilihat pada gambar 10.

**

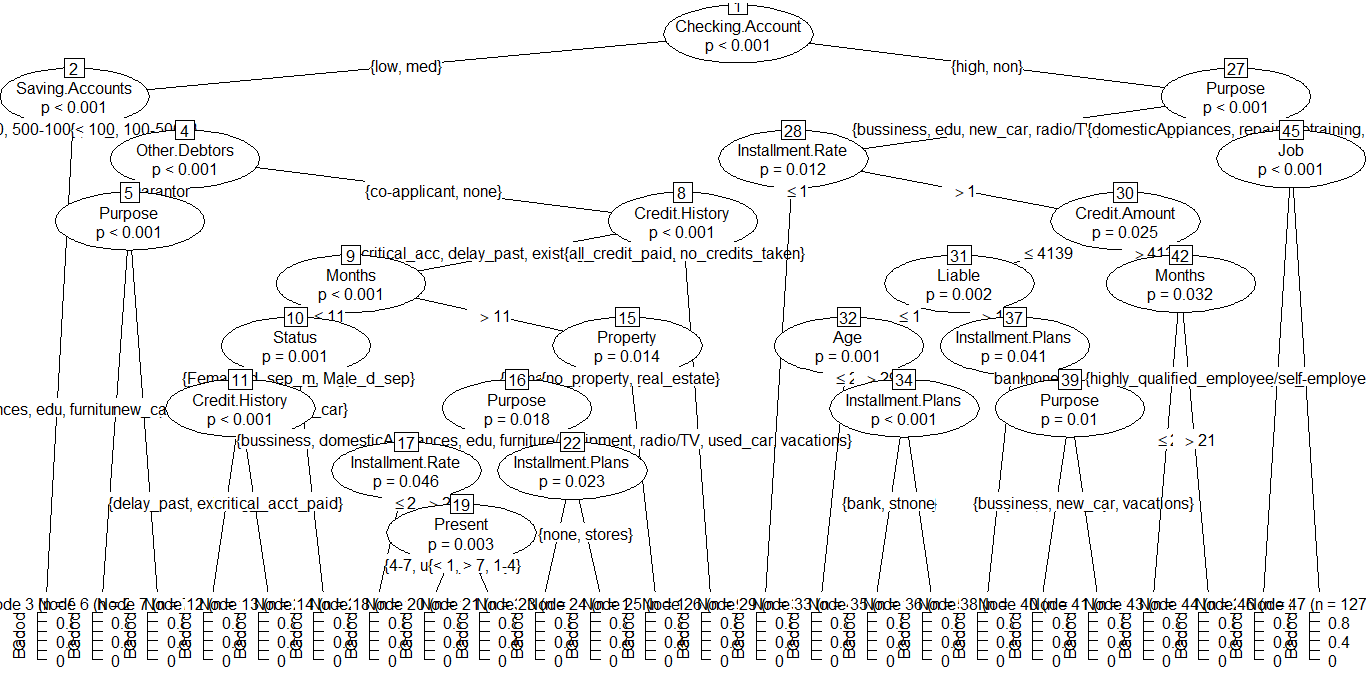
Gambar 10 Decision Tree Dengan Stratified Random Sampling

Hasil pemodelan decision tree dengan *stratified random sampling* dapat dilihat pada gambar 11.



Gambar 11 Decision Tree dengan Stratified Random Sampling

Hasil pemodelan decision tree dengan *stratified random sampling* dapat dilihat pada gambar 12.



Gambar 12 Hasil Pemodelan Decision Tree Over Sampling

**Evaluasi Model**

Berdasarkan hasil pemodelan decision tree pada data latih dilakukan evaluasi dan uji coba pada data uji yang telah diambil. Hasil evaluasi menghasilkan confusion matrix*.* *Hasil* Confusion matrix untuk pemisahan simple random sampling dapat dilihat pada tabel 2.

|  |  |
| --- | --- |
| **https://lh3.googleusercontent.com/MKu0QQOWz6YLC-dL1nD5_k9ELig09CeQAcvab_VAoAIP3twG-8vHlfgSKCSUb_wyBpgX67RQLLuKqCeVowJFoHunhoD-xF7C_1PQLgD4Jo60YKFvBEHSh-GY6WddBRpxFoWUMT5T**  **Data Train** | **https://lh5.googleusercontent.com/ioQsrlZgUGrokHi301WSAvxJeyK2VZCGY_9DU3Mou8iKZao9186GTaBwbKubTCDs8UKYlOf2mCgf4qd6uGBiHSrmLKiWJTz_wtMkyi_H-92r4s1TEldUYBkSON1cwQ2WHDnR4cpW**  **Data Test** |

Tabel 2 Confusion Matrix Decision Tree Simple Random Sampling

Confusion matrix untuk *stratified random sampling* dapat dilihat pada tabel 3.

|  |  |
| --- | --- |
| **https://lh3.googleusercontent.com/Wjo06w1oN0zrBCLcNmQ9PkJtN7cY66tyxy9tN-0ZAhyk6097GHiek_91H3kxtR2a1aCVSVRTmPzuXdh33S1QBEw-5woidzGTpSsT1-gLWJgEHFu7jKaeZ7WjGERjq-mcSFgtRq7x**  **Data Train** | **https://lh3.googleusercontent.com/uhT-P0LUbJPMgY7fiuNDTN02if7wuT526W-EW-H-o0K4FHV-M0TWucOGQhYRTyjExT6ibwvv0jehqHD-3P7Fm0LZo2K3CwEb5_dfLqXInNl6K4qq7H6-SPudzTV2J4YPEem8DyuF**  **Data Test** |

Tabel 3 Confusion Matrix Decision Tree Stratified Random Sampling

Confusion matrix untuk *oversampling* dapat dilihat pada tabel 4.

|  |  |
| --- | --- |
| **https://lh6.googleusercontent.com/cS07EW8o6CTy2BfJ5I9PjqRsVguTZ4y_VKAvfFDfVpRaj0f7uFrlXvPnrBqjZW1pQVI8uI4dgYte3rLiahLbT8aq_8xgpvJt__dWWMcMIh8Nttjsqly-u5qWbe_37tPQscMG0hUq**  **Data Train** | **https://lh5.googleusercontent.com/xxpWP7z1-asQR9HxIgklq_hsBT3mINYMlDnTeA-ACXlibnFUwws0NUnEzCzcEHAdi4CO1kIeo98dQO5cgV8ei147C3gARmHLsDKwPA7zYKPRwbSXpaQf4AZ0mC-Aie2ZoNmwxPpv**  **Data Test** |

Tabel 4 Confusion Matrix Decision Tree Over Sampling

Pemodelan Data German Credit Card menggunakan *Decision Tree* dengan teknik *simple random samplin*g pada data train menghasilkan akurasi sebesar 75.64 %. Sedangkan untuk teknik *stratified random sampling* menghasilkan akurasi sebesar 76.54%. dan untuk teknik *oversampling* menghasilkan akurasi 81.9%.

Pemodelan Data German Credit Card menggunakan *Decision Tree* dengan teknik *simple random samplin*g pada data test menghasilkan akurasi sebesar 71.85 %. Sedangkan untuk teknik *stratified random sampling* menghasilkan akurasi sebesar 74.09%. dan untuk teknik *oversampling* menghasilkan akurasi 70.2%.

**KESIMPULAN DAN SARAN**

**Kesimpulan**

Pemisahan data yang sesuai untuk data german credit card adalah stratified random sampling karena imbalance dari jumlah data antar kelas. Metode klasifikasi dengan algoritme decision tree pada data kelayakan kredit ini menghasilkan akurasi sebesar 74,90%. Penerapan algortime ini sudah cukup baik untuk menentukan kelayakan pemohon kredit.

**Saran**

Saran untuk penelitian selanjutnya adalah penggunaan metode atau model lain pada dataset tersebut agar mendapatkan akurasi yang lebih baik lagi.

**DAFTAR PUSTAKA**

Berry, Michael JA, Gordon SL. 2004. Data Mining Techniques for Marketing, Sales, Customer Relationship Management. Second Edition. Willey Publishing, Inc.

Chen Y, Dai L, Chen XQ. 2008. GATS-C4.5: An Algorithm for Optimizing Features in Flow Classification. EEE Communications Society subject matter experts for publication in the IEEE CCNC 2008 proceedings.hlm 466-470.

Han J, Micheline K, Jian P. 2012. *Data mining: Concepts and Techniques (3 th ed)*. Waltham MA: Elsevier/Morgan Kaufmann.

Witten IH, Frank E, Hall MA. 2011. *Data Mining Practical Machine Learning Tools and Techniques (3rd ed)*. Elsevie. USA

Dataset melalui alamat <https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)>

1. [↑](#footnote-ref-1)