3. Use Case Problem

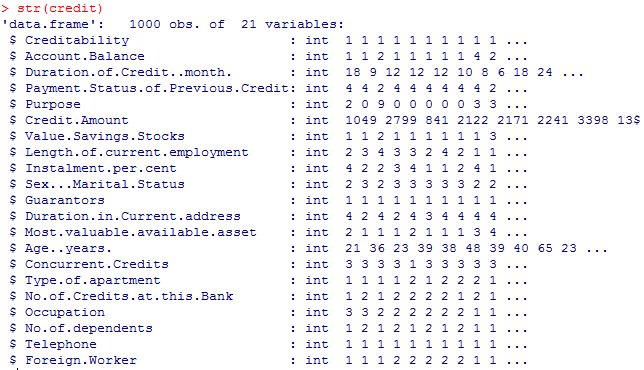
a. Masalah Credit Scoring

data yang diberikan adalah data kredit dari nasabah Jerman. Adapun metode yang dipakai untuk menganalisis data yaitu Decision Tree serta langkah-langkah untuk menyelesaikan masalah sebagai berikut :

1. Load data dari file Germany\_credit

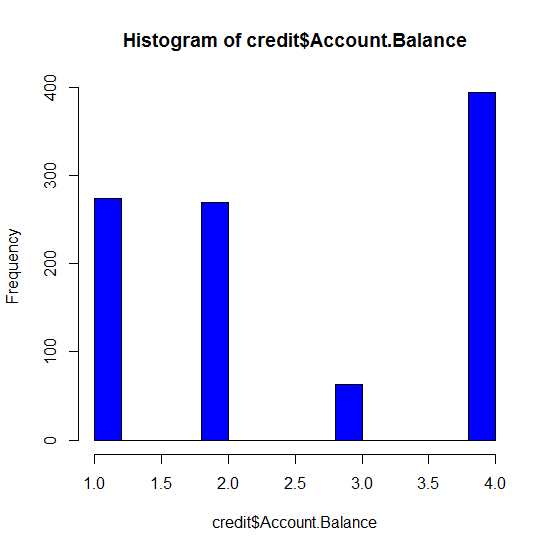


1. Kemudian perincikan data yaitu dengan cara



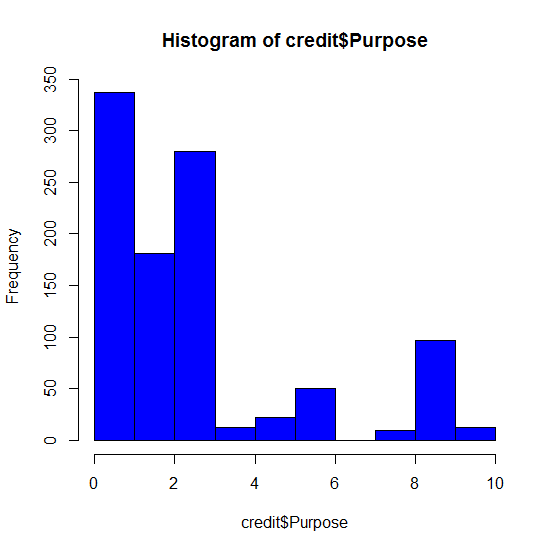
1. Eksplorasi data dengan Histogram



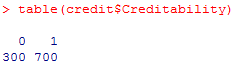


Dari Histogram diatas yang memiliki nilai tertinggi untuk frekuensinya diperoleh oleh nasabah yang terkategori dalam bilangan 4 yaitu nilai akunnya yang berada diatas sama dengan 200 DM dan terendah oleh bilangan 3 yaitu nilai akunnya berada diantara 0 DM dan dibawah 200 DM.

Selanjutnya tujuan kredit yang digunakan ditunjukan dengan histogram sebagai berikut.



Artinya tujuan kredit dengan frekuensi tertinggi dilakukan oleh nasabah yang terkategori di bilangan 1 yaitu kredit untuk mobil baru kemudian dilanjutkan dengan bilangan ke 3 yaitu untuk kebutuhan furniture. Setelah itu kemudian lihat data untuk kemampuan kredit yang layak sebagai berikut.



Maksudnya yaitu 70% kredit diberikan masih layak dan nasabah masih mempunyai kemampuan untuk membayar sedangkan 30% kredit yang diberikan tidak layak dan memiliki potensi untuk kredit macet.

1. Lakukan Training Data dengan cara sebagai berikut.



Dengan training data yang diambil dari row 1 sampai 900 untuk training secara acak agar pengambilan data lebih baik serta 901 sampai 1000 untuk testing data.

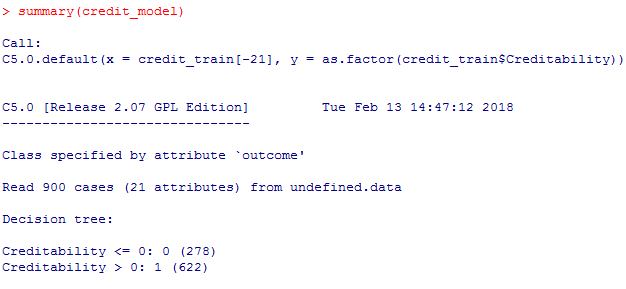
1. Terapkan Metode Decision Tree

Dengan memanggil package library (C50) untuk menerapkan metode.

Selanjutnya pakai fungsi untuk menerapkan metode sebagai berikut.



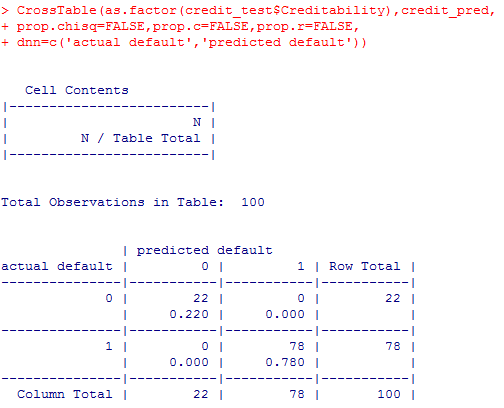
Setelah itu lakukan summary



Selanjutnya fungsi yang telah di training digabungkan dengan fungsi testing



Setelah itu dengan library (gmodels) untuk memprediksi klasifikasi



Dari hasil kesimpulan diatas artinya 78% klasifikasi kredit berada pada kondisi nasabah mampu membayar secara baik dan 22% berada pada posisi kegagalan membayar kredit (kredit macet).

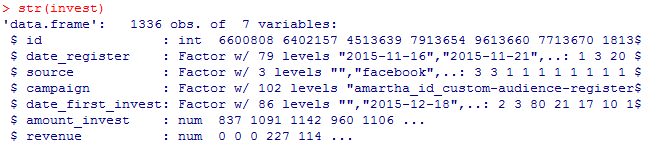
b. Masalah Investor Data

Data yang diberikan adalah data investor untuk mencari hal yang berpengaruh dalam investasi. Adapun metode yang dipakai dalam menganalis data yaitu Metode Regresi Linear. Selanjutnya, langkah-langkah yang diambil sebagai berikut.

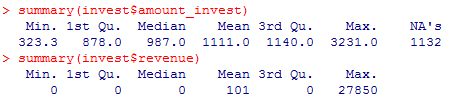
1. Load Data Investor sebagai berikut.



1. Perinci Data Investor sebagai berikut.

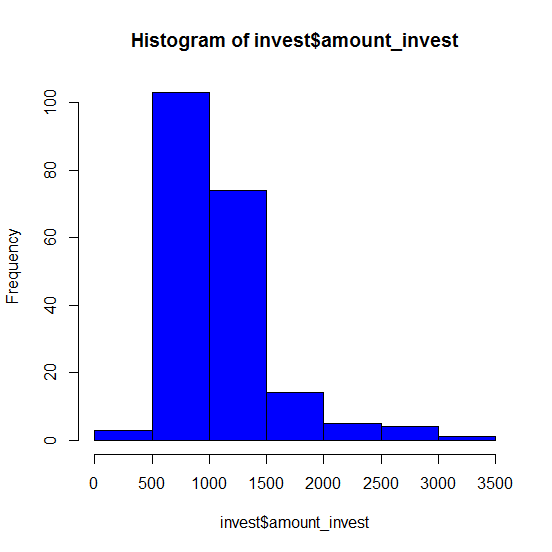


1. Eksplorasi Data Investor sebagai berikut.



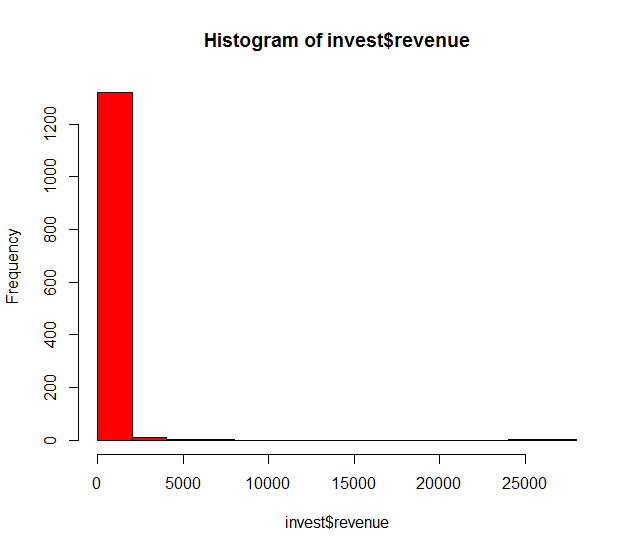
Sehingga diperoleh informasi bahwa rata-rata jumlah investasi yaitu 1111 dan maksimal jumlah yang diinvestasikan yaitu 3231 sedangkan rata-rata revenue yang diperoleh 101 dan revenue maksimal yang diperoleh yaitu 27850.

Selanjutnya kita bisa eksplorasi data dengan histogram untuk jumlah investasi sebagai berikut.



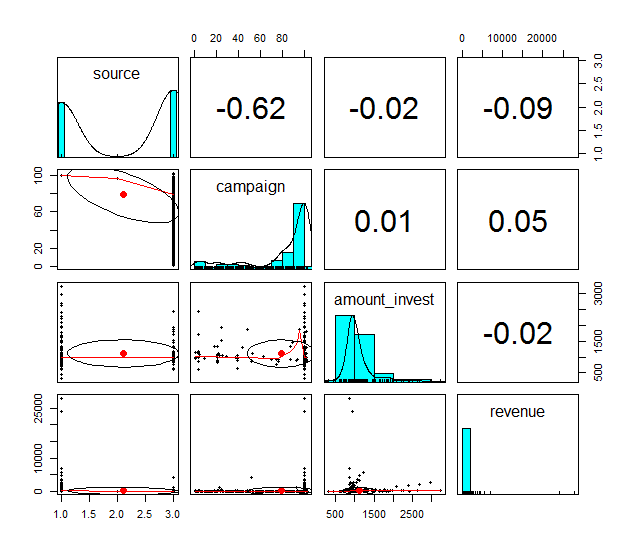
Dari informasi diatas dapat diketahui bahwa investor banyak menginvestasikan pada kisaran 500-1000 dan 1000-1500.

Selanjutnya untuk revenue yang diperoleh dapat ditunjukan dengan histogram sebagai berikut.



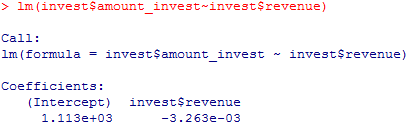
Dari informasi diatas dapat diketahui bahwa revenue yang diperoleh oleh mayoritas investor masih berada di kisaran 0-5000.

Setelah itu, cari korelasi antar variabel sebelum menerapkan metode Regresi yaitu dengan memakai library (psych).

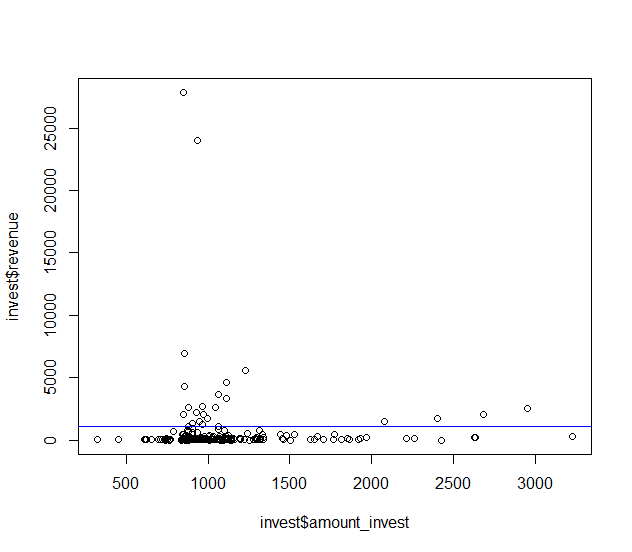


Dari informasi diatas kita dapat menarik pemahaman bahwa korelasi yang terendah yaitu korelasi source dengan campaign ternyata korelasinya sangat rendah. Dan korelasi yang tertinggi yaitu korelasi antara campaign dan revenue artinya campaign ini mempunyai pengaruh cukup kuat untuk menghasilkan revenue yang lebih baik serta jumlah investor.

1. Dengan menerapkan metode regresi maka diperoleh sebagai berikut



Lalu selanjutnya dengan plot regresi diperoleh sebagai berikut.



Dari informasi diatas dapat diketahui bahwa hasil regresi akurasinya tidak begitu dekat dengan data aktual karena korelasi yang rendah.