

## **DECISION TREE MENGGUNAKAN ALGORITMA C4.5 UNTUK ANALISA KELAYAKAN PEMBERIAN KREDIT**

Diana Yusuf<sup>1</sup>, Saeful Bahri<sup>2</sup>, Anggita Larasati<sup>3</sup> (\*)

<sup>1</sup>ITB Ahmad Dahlan, Jakarta

<sup>2</sup>ITB Ahmad Dahlan, Jakarta

<sup>3</sup>ITB Ahmad Dahlan, Jakarta

---

### **Abstract**

*Data mining is a new technology that has been successfully applied in many fields. Many problems are solved by data mining algorithms as a supporter. Classification is one of the methods contained in data mining and not a few researchers who use the classification methods in solving problems. Credit analysis will be done with the digging data against existing data customer credits based on attribute with data mining algorithm. The algorithm C4.5 is itself a group of decision tree algorithm. This algorithm has input in the form of training and samples. Data mining technique used to classify loans with algorithm C4.5. Analysis and processing of data use applied tools of RapidMiner.*

---

**Kata Kunci:** *Decision Tree, Algorithm C 4.5, Credit Analysis, Classification*

Juli – Desember 2021, Vol 2 (2) : hlm 97-106  
©2021 Institut Teknologi dan Bisnis Ahmad Dahlan.  
All rights reserved.

---

(\*) Korespondensi: [dianayusuf01@gmail.com](mailto:dianayusuf01@gmail.com) (Diana Yusuf), [mr.saeful.bahri@gmail.com](mailto:mr.saeful.bahri@gmail.com) (Saeful Bahri), [larassati.anggita@gmail.com](mailto:larassati.anggita@gmail.com) (Anggita Larasati),



## PENDAHULUAN

Kredit merupakan salah satu sumber utama pemasukan bagi suatu bank, meskipun sebagai sumber utama penghasilan, kredit juga sekaligus menjadi sumber resiko bisnis terbesar. Kredit bermasalah atau sering disebut dengan kredit macet adalah permasalahan yang telah melekat dalam perbankan, karena kredit macet dalam jumlah besar dapat mengganggu sendi perekonomian dan juga bisa menggerus dana operasional bank tersebut. Dan masih banyak dampak dan kerugian lainnya yang disebabkan oleh kredit macet ini jika tidak bisa dikendalikan dengan baik.

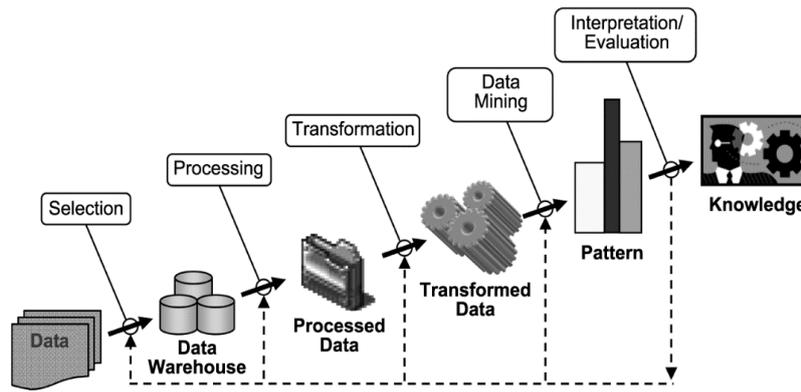
Algoritma C 4.5 dan Decision Tree telah banyak digunakan dalam berbagai penelitian untuk menganalisa kelayakan pemberian kredit terhadap beberapa lembaga pembiayaan. Tidak hanya itu dalam penelitian (Gu dan Zhou, 2012) mereka membahas tentang bagaimana melakukan analisis dari setiap ujian para siswa mulai dari semester pertama beserta dengan parameter-parameter pendukung lainnya, tujuannya ialah untuk memprediksi berapa skor yang akan didapatkan untuk masuk perguruan tinggi. Penelitian yang mereka lakukan tersebut diharapkan dapat membantu perbaikan proses mengajar pada sekolah menengah sehingga kualitas mengajar menjadi lebih baik. Dalam penelitian lain yang dilakukan oleh (Adi Sucipto, 2015) Algoritma C 4.5 digunakan dalam memprediksi kredit macet melalui perilaku nasabah pada sebuah koperasi simpan pinjam.

Tingginya persentase resiko kredit macet yang saat ini dihadapi oleh lembaga pembiayaan atau bank, menjadi alasan untuk menemukan cara meminimalisir permasalahan ini terus terjadi. Perlu adanya perbaikan pola atau analisa pemberian kredit kepada calon nasabah pada suatu bank. Perlu dilihat dari berbagai parameter yang ada sehingga didapatkan hasil analisa yang kredit tersebut layak untuk diberikan.

Melakukan analisis terhadap data mentah secara manual dan menemukan informasi yang benar dalam sebuah perusahaan merupakan proses yang cukup sulit. Akan tetapi teknik data mining secara otomatis dapat mendeteksi pola yang relevan dari kumpulan data mentah tersebut menggunakan algoritma yang ada. Dalam data mining dikenal juga dengan metode pohon keputusan atau *Decision Tree*, metode ini adalah salah satu metode terbaik dan sering digunakan dalam kasus-kasus data mining.

## METODE

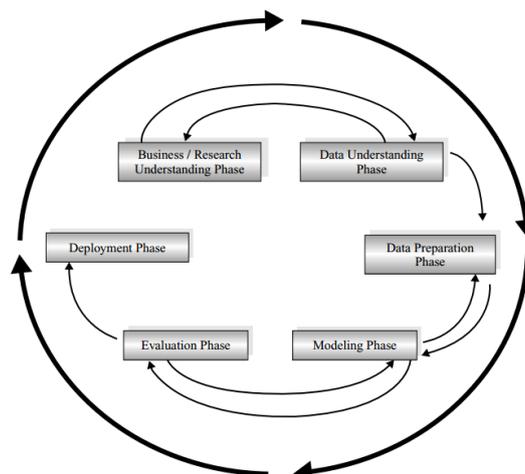
*Knowledge Discovery in Database* (KDD) ialah segala bentuk kegiatan yang didalamnya meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar. KDD juga dapat didefinisikan sebagai metode untuk memperoleh pengetahuan dari basis data yang dibentuk.



Gambar 1. Tahapan dalam *Knowledge Discovery in Database* (KDD)

Pada Gambar 1 dapat dilihat bahwa sebenarnya Data Mining merupakan salah satu tahapan yang ada dalam KDD. Sehingga Data Mining bisa didefinisikan sebagai sebuah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisa dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Dalam sebuah penelitian disebutkan bahwa Data Mining ialah proses untuk mengekstrak potensi, pengetahuan yang berharga (model atau aturan) dari dalam database, gudang, atau media penyimpanan informasi lainnya. Data Mining juga mengadopsi ilmu statistik, database, dan teknologi kecerdasan buatan dalam proses pengembangan diri (Jiang Wei-Li, 2011).

Metode penelitian yang digunakan ialah CRISP-DM yang mana merupakan gabungan beberapa industri yang menyediakan standar proses Data Mining sebagai salah satu strategi pemecah masalah dari ruang lingkup bisnis atau unit penelitian. Sebuah proyek Data Mining mempunyai siklus hidup yang terbagi dalam beberapa tahapan, dimana keseluruhan tahapan tersebut berurutan dan bersifat adaptif. Artinya tahap berikutnya dalam urutan bergantung pada keluaran dari tahap sebelumnya.



Gambar 2. Proses Data Mining menurut CRISP-DM

Tahapan dalam Data Mining antara lain :

1. *Bussiness Phase*

Fase ini ialah fase dimana mulai menentukan tujuan proyek dan kebutuhan-kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan. Menterjemahkan tujuan dan batasan menjadi formula dari permasalahan Data Mining serta menyiapkan strategi awal untuk mencapai tujuan

2. *Data Understanding Phase*

Pada fase ini dilakukan proses pengumpulan data kemudian menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal. Serta melakukan evaluasi terhadap kualitas data dan jika dibutuhkan bisa pilih sebagian kecil group data yang mungkin mengandung pola-pola dari permasalahan.

3. *Data Preparation Phase*

Siapkan data-data dari awal yakni kumpulan data yang akan digunakan untuk keseluruhan tahapan berikutnya. Pada tahap ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif. Kemudian pilih kasus dan variabel yang ingin dianalisis serta sesuai analisis yang akan dilakukan. Setelah itu lakukan perubahan pada variabel jika memang dibutuhkan.

4. *Modeling Phase*

Fase ini kita bisa memilih dan mengaplikasikan teknik pemodelan yang sesuai, kalibrasi aturan model untuk mengoptimalkan hasil, dan perlu juga diperhatikan bahwa beberapa teknik Data Mining memungkinkan untuk digunakan pada permasalahan yang sama. Tidak hanya itu jika memang dibutuhkan proses dapat kembali ke tahap pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik Data Mining tertentu.

Klasifikasi merupakan salah satu teknik Data Mining yang cukup sering digunakan untuk mengolah data menjadi informasi dan memperoleh pengetahuan baik dalam unit bisnis atau unit penelitian. Teknik klasifikasi adalah suatu pekerjaan menilai objek data untuk memasukkan ke dalam kelas dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang akan dilakukan, pertama pembangunan model sebagai *prototype* untuk disimpan sebagai memori. Kedua, penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya.

Setiap teknik Data Mining menggunakan suatu algoritma pembelajaran, begitu juga dengan teknik klasifikasi terdapat algoritma untuk mendapatkan suatu model yang paling memenuhi antara himpunan atribut dan label kelas dalam data masukan. Tujuan dari algoritma pembelajaran ialah untuk membangun model secara umum yang berkemampuan baik yakni model yang dapat memprediksi label kelas dari *record* yang tidak diketahui kelas sebelumnya dengan lebih akurat. Terdapat beberapa model-model yang dikembangkan oleh peneliti untuk menyelesaikan kasus-kasus klasifikasi salah satunya yang akan dibahas ialah Pohon Keputusan/*Decision Tree*.

*Decision Tree* ialah cara yang paling potensial untuk menentukan aturan dalam klasifikasi Data Mining. Level teknik ini lebih tinggi dari jaringan saraf tiruan, metode statistik multivariant seperti mampu menganalisis data secara cepat, memiliki presensi yang tinggi dan mudah digunakan untuk membangun suatu model. Konsep dasar pohon keputusan ini adalah mengubah data menjadi pohon keputusan dengan aturan-aturannya. Atribut yang dipilih akan menghasilkan partisi dengan data yang lebih seragam dan dapat menghasilkan pohon keputusan yang sederhana dengan jumlah perulangan yang sedikit. Sebuah pohon keputusan terdiri dari sekumpulan aturan yang bertujuan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil dan lebih homogen dengan memperhatikan variabel tujuan.

Proses pada *Decision Tree* ialah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule tersebut. Ada beberapa algoritma yang dapat dipakai dalam membentuk *Decision Tree* seperti ID3, CART, dan algoritma C 4.5.

Algoritma C 4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan seperti yang sudah disebutkan sebelumnya. Secara umum algoritma C 4.5 antara lain :

1. Memilih atribut sebagai akar
2. Membuat cabang untuk tiap-tiap nilai
3. Membagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang mempunyai kelas yang sama

Beberapa pengembangan algoritma C 4.5 dalam membangun pohon keputusan sebagai berikut :

1. Menyiapkan data training. Data ini bisa diambil data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain tertinggi akan menjadi akar pertama. Sebelum memperoleh nilai gain dari atribut, terlebih dahulu dicari nilai entropy. Rumus mencari nilai etropy yakni :

$$Entropy (S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

3. Kemudian hitung nilai gain menggunakan rumus berikut :

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i)$$

## HASIL DAN PEMBAHASAN

Data yang diambil ialah data debitur di PT BPR Lubuk Raya Mandiri, data ini merupakan database penting perusahaan karena terdiri dari kumpulan fakta-fakta sejarah kredit debitur dan termasuk juga didalamnya sistem pembayaran kredit dari debitur-debitur tersebut. Dari sekian banyak data yang diperoleh, dilakukan analisa

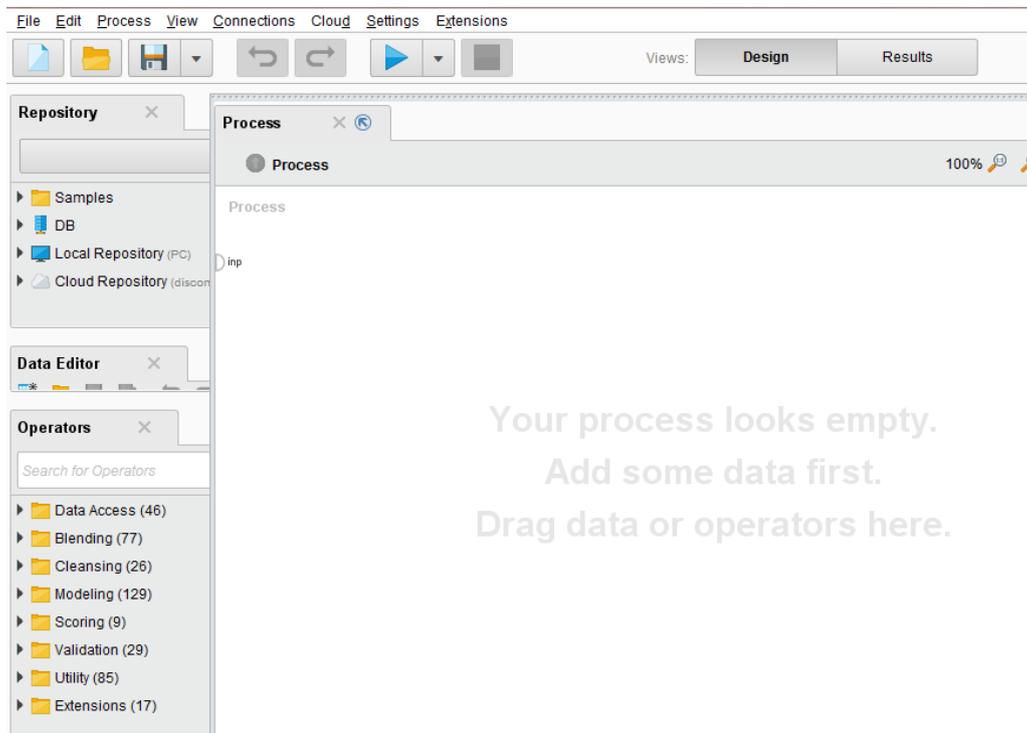
sehingga diperoleh data-data yang mempunyai keterkaitan dengan penelitian yang akan dilakukan.

Data mentah yang ada kemudian dilakukan *pra-process* yakni *cleaning data* dan *selection data*, sehingga diperoleh format data yang akan diolah menjadi sebuah model pohon keputusan menggunakan algoritma C 4.5.

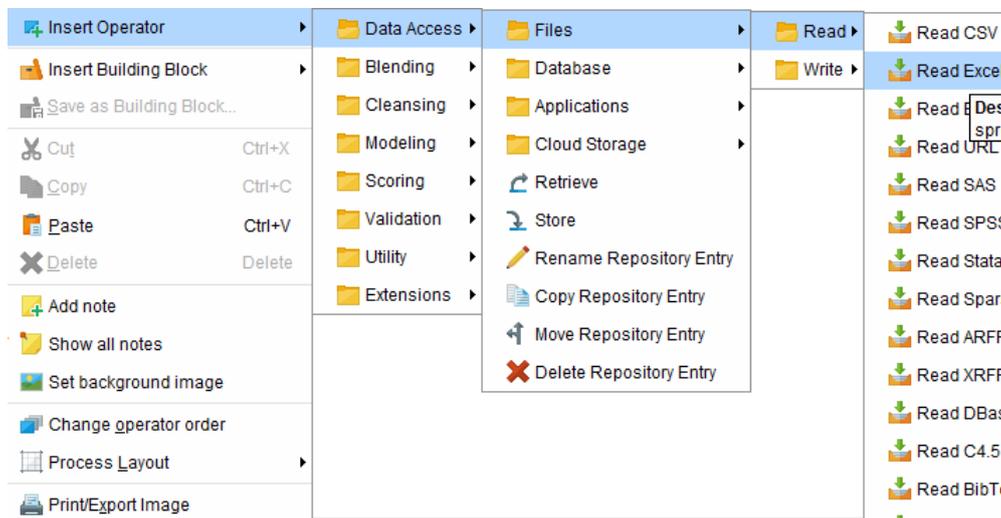
Tabel 1. Dataset

Nama	Pekerjaan	Penghasilan	Kepemilikan Rumah	Lingkungan	Kelayakan
Herlina	Karyawan Tetap	> 3X Angsuran	Pribadi	Bermasalah	Layak
Yori Naldi	Karyawan Tetap	> 3X Angsuran	Pribadi	Baik	Layak
Monalisa	Wiraswasta	> 3X Angsuran	Kontrak	Bermasalah	Layak
Eldawati	Honoror	> 3X Angsuran	Pribadi	Baik	Layak
Jero Fili	Wiraswasta	< 3X Angsuran	Pribadi	Bermasalah	Layak
Depy Saputra	Wiraswasta	< 3X Angsuran	Kontrak	Baik	Layak
Guspar	Honoror	> 3X Angsuran	Pribadi	Baik	Layak
Riri Anggrani	Wiraswasta	< 3X Angsuran	Kontrak	Baik	Layak
Wiwin Febriyan	Honoror	< 3X Angsuran	Pribadi	Baik	Layak
Alimar T	Wiraswasta	> 3X Angsuran	Kontrak	Bermasalah	Tidak Layak
Hernita	Honoror	< 3X Angsuran	Kontrak	Baik	Tidak Layak
Rafliza Tanjung	Karyawan Tetap	< 3X Angsuran	Pribadi	Bermasalah	Tidak Layak
Nia Alami	Karyawan Tetap	< 3X Angsuran	Kontrak	Bermasalah	Tidak Layak
Anton Sujarwo	Wiraswasta	< 3X Angsuran	Pribadi	Bermasalah	Tidak Layak
Ricky Resta	Karyawan Tetap	> 3X Angsuran	Pribadi	Bermasalah	Layak
PT Ayah Ibu Transport	Karyawan Tetap	> 3X Angsuran	Pribadi	Baik	Layak
Rahma Dewita	Wiraswasta	> 3X Angsuran	Kontrak	Bermasalah	Layak
Anggi Saputra	Honoror	> 3X Angsuran	Pribadi	Baik	Layak

Proses input data nilai pada *RapidMiner* dimulai dengan membuat project baru dari menu *NEW Project*. Dalam project inilah proses pengerjaan klasifikasi data akan dilakukan.

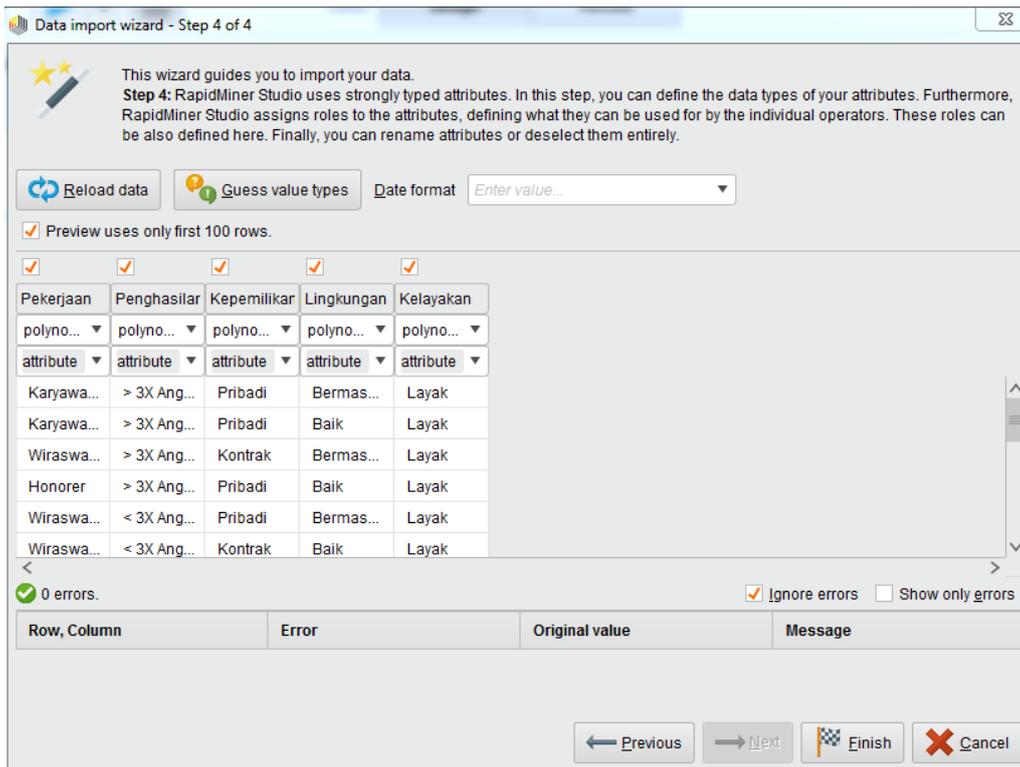


Gambar 3. Halaman Process *RapidMiner*



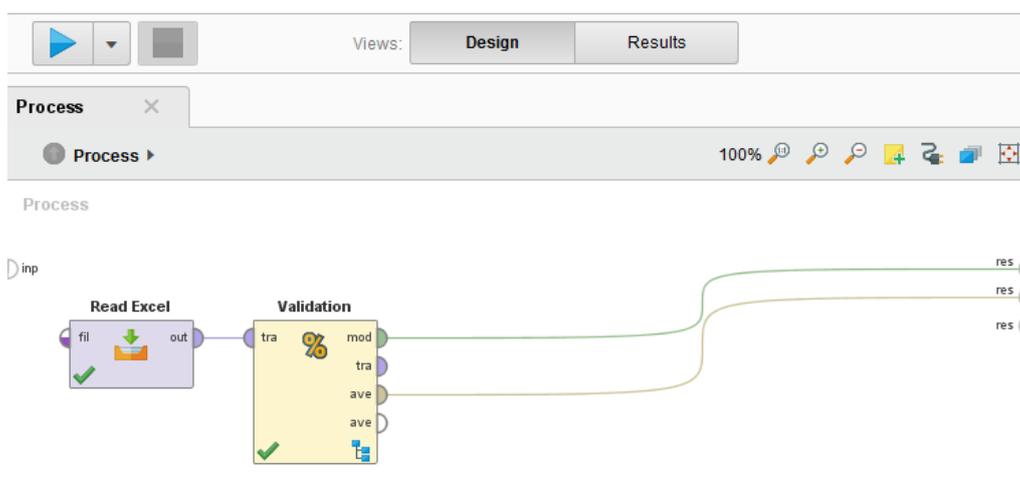
Gambar 4. Pembuatan Operator Baru pada *RapidMiner*

Setelah selesai membuat operator baru dilanjutkan dengan proses Data Import Wizard, pada tahap ini akan dipilih sheet yang berisikan sumber data pada Microsoft Excel yang digunakan. Dalam tahapan ini juga menentukan tipe data dan atribut yang akan digunakan, *RapidMiner* biasanya akan memberikan tipe data dan atribut secara otomatis. Namun apabila kita merasa tipe data yang diberikan oleh *RapidMiner* tidak cocok, maka kita bisa mengubahnya seperti tampilan berikut ini :

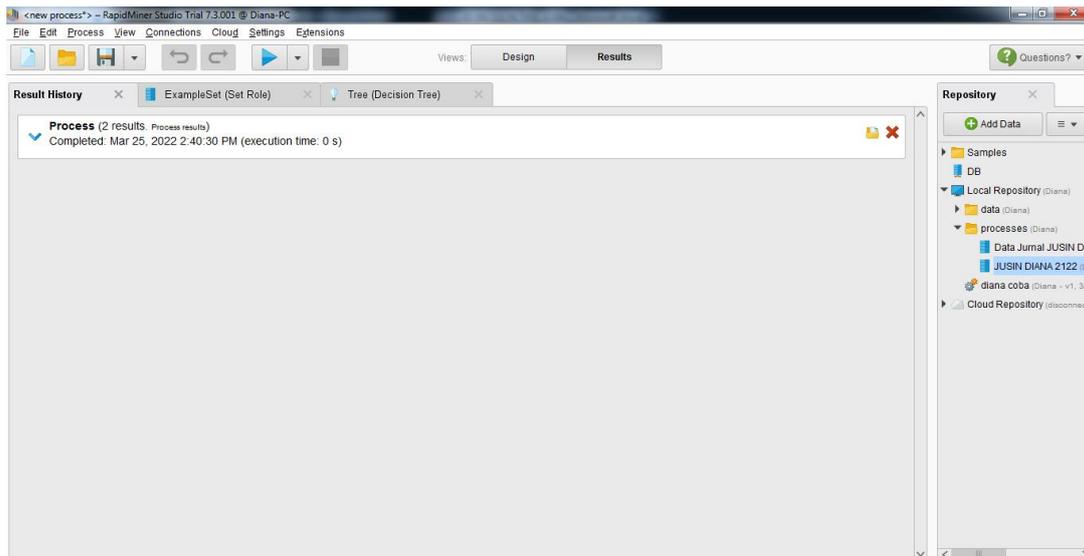


Gambar 5. Penentuan Tipe Data dan Atribut

Tahap import data selesai dilakukan, maka operator yang sudah dibuat akan berisi data siap oleh yang telah di import langsung dari file excel. Tahapan selanjutnya yang harus dilakukan untuk memperoleh pohon keputusan dari kasus ini ialah dengan membuat dan menghubungkan berbagai operator ke *main process*. Mulai dari operator *Cross Validation* sampai dengan operator *decision tree* yang siap untuk di proses oleh *RapidMiner*.

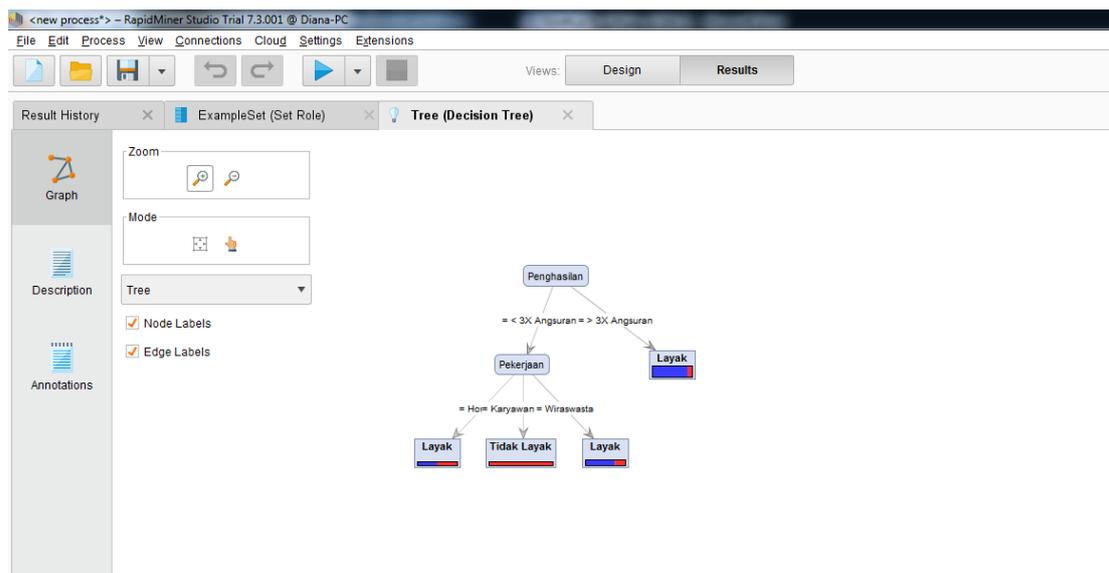


Gambar 6. Tampilan Data Siap di Proses



Gambar 7. Tampilan *Running Data*

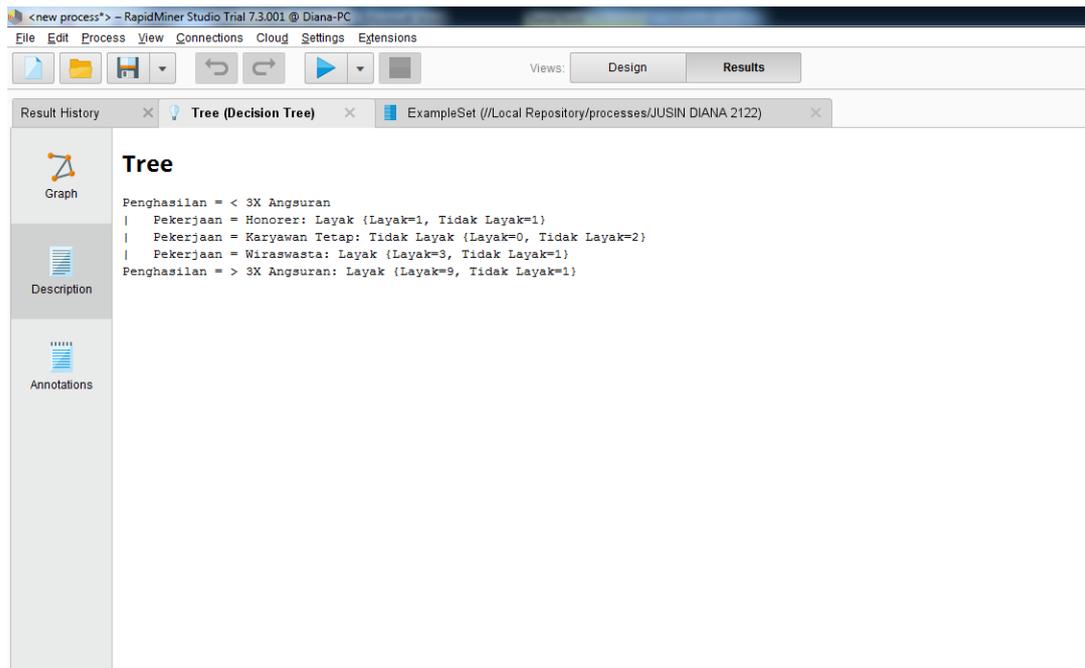
Setelah dilakukan pengujian data dengan Algoritma C 4.5 *RapidMiner* maka terbentuklah sebuah pohon keputusan seperti pada gambar dibawah ini.



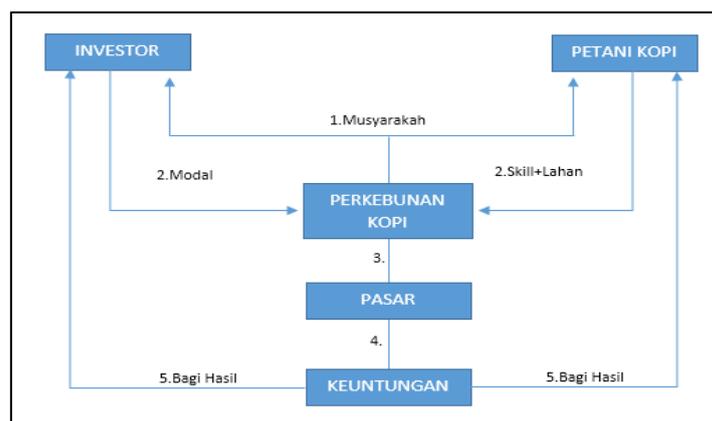
Gambar 8. *Decision Tree*

Berdasarkan hasil pengujian menggunakan *RapidMiner* terhadap keseluruhan data bahwa dalam melakukan analisa kelayakan kredit pada PT BPR Lubuk Raya Mandiri dapat dilihat dari penghasilan debitur. Karena terlihat pada *decision tree* yang terbentuk akan tertinggi adalah penghasilan.

Selain itu dari hasil pengujian juga diperoleh rule yang bisa menjadi landasan perusahaan untuk membuat kebijakan terkait kredit kedepannya. Rule yang dihasilkan ialah seperti berikut ini :



Gambar 9. Rule yang Terbentuk



Sumber: analisis data, 2016

Gambar 1. Model Pembiayaan Syariah Petani Kopi

## KESIMPULAN

Berdasarkan proses yang dilakukan kesimpulan yang dapat diambil dari penelitian ini antara lain sebagai berikut :

1. Algoritma C 4.5 dengan decision tree bisa digunakan untuk menganalisa kelayakan kredit calon debitur.
2. Rule yang dihasilkan oleh decision tree bisa digunakan sebagai landasan pembuatan kebijakan kredit di PT BPR Lubuk Raya Mandiri.
3. Hasil dari decision tree yang dihasilkan tidak hanya berupa keputusan layak atau tidak layak. Tetapi juga digunakan untuk membuat berbagai kebijakan pemberian kredit berdasarkan pola yang dihasilkan.

## DAFTAR PUSTAKA

- Wei-Li, Jiang. 2011. *Research and Application of Credit Score Based on Decison Tree Model*. ICAIC 2011, Part I, CCIS 224, pp. 493-501.
- Iriandi, Nandang dan Nia Nuraeni. 2016. *Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta*. Vol. II No. 1.
- Rusito, et all. 2016. *Implementasi Metode Decison Tree dan Algoritma C4.5 untuk Klasifikasi Data Nasabah Bank*. Infokam Nomor I Th. XII.
- Pudjo, Prabowo., dkk. 2013. *Penerapan Data Mining dengan Matlab*.
- Susanti. 2014. *Klasifikasi Kredit Menggunakan Metode Decison Tree Pada Nasabah PD BPR BKK Gabus*.
- Sathyadevan, Shiju., and Remya R. Nair. *Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest*. Volume I.
- Tampubolon, Kennedi., et all. 2013. *Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan*. Volume I. ISSN :2339-210X.
- Gu, Ping., and Qi Zhou. 2012. *Student Performance Prediction Based on Improved C4.5 Decision Tree Algorithm*. ECICE 2012, AISC 146, pp. 1-8
- Sucipto, Adi. 2015. *Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam dengan Menggunakan Metode Algoritma Klasifikasi C4.5*. Volume 6 No 1.
- Kim, DaEun. 2006. *Minimizing Structural Risk on Decision Tree Classification*. SCI 16, 241-260.