

BAB III

METODOLOGI PENELITIAN

1.1. Metodologi Penelitian

Tahapan yang dilakukan pada penelitian ini diantaranya :

- a) Melakukan identifikasi masalah yang ada kemudian menganalisanya untuk menemukan kemungkinan-kemungkinan solusi yang dapat dilakukan.
- b) Melakukan studi literatur mengenai resiko kredit macet, serta metode-metode klasifikasi yang akan digunakan, dari berbagai sumber seperti buku ataupun jurnal penelitian.
- c) Melakukan konsultasi kepada dosen pembimbing atau dosen lain yang ahli di bidang data mining mengenai rancangan yang akan dibuat serta penyelesaian yang dilakukan.
- d) Mengumpulkan data penelitian yang dibutuhkan yaitu berupa data transaksi yang berasal dari KSPPS. BMT Adil Berkah Sejahtera.
- e) Membuat rancangan sistem penelitian yang terdiri dari yaitu *preprocessing* dataset, pembagian data latih dan data uji, pelatihan dan pengujian data menggunakan *Naive Bayes* dan algoritma C4.5.
- f) Mengimplementasikan rancangan sistem serta menganalisa hasil.

1.2. Alat dan Bahan

Adapun alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

3.2.1. Hardware

Kebutuhan perangkat keras (*hardware*) atau PC yang digunakan memiliki spesifikasi sebagai berikut.

- a. Kapasitas Processor : 2.20 GHz, 7 Core
- b. RAM : 8 Gb

- c. VGA : GTX 1650 Ti, 8 Gb
- d. *Motherboard* : B450 Chipset
- e. *Storage* : Sata 60 Gb

3.2.2. *Software*

Kebutuhan perangkat lunak (*software*) yang digunakan :

- a. Sistem Operasi Windows 11 Home
- b. Aplikasi RapidMiner, *Graphical User Interface (GUI)*

1.3. Teknik Pengambilan Data

Metode pengumpulan data merupakan suatu hal yang penting dalam penelitian dan merupakan strategi atau cara yang digunakan oleh peneliti dalam mengumpulkan data yang diperlukan dalam penelitiannya. Metode pengumpulan data yang digunakan dalam penelitian ini yaitu :

1.3.1. Studi Lapangan (*Field Research*)

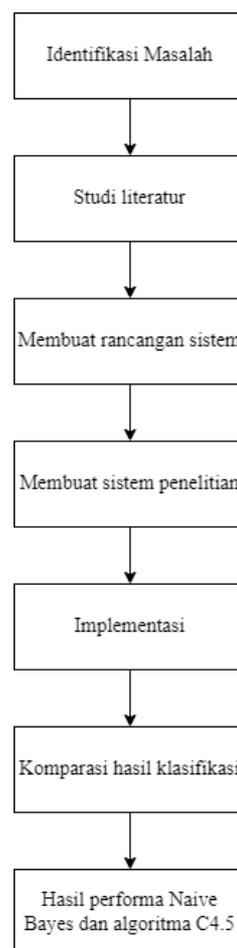
Pengamatan atau observasi adalah suatu metode atau cara untuk memperoleh data dengan cara mengamati proses kegiatan yang sedang berlangsung. Pengamatan dapat dilakukan dengan partisipasi ataupun nonpartisipasi (Sudaryono 2015). Berdasarkan pengertian di atas observasi merupakan suatu pengamatan yang dilakukan secara langsung untuk mengumpulkan data yang digunakan dalam penelitian. Adapun teknik pengumpulan data dan informasi yang dilakukan pada saat studi lapangan pada titik daratan yang akan dianalisa adalah Pengamatan Langsung (*Observation*) yaitu Pengumpulan data yang dilakukan penulis pada saat pengamatan langsung pada data nasabah yang terdapat pada KSPPS. BMT Adil Berkah Sejahtera.

1.3.2. Tinjauan Pustaka (*Research Library*)

Tinjauan pustaka dilakukan dengan cara membaca, mengutip dan membuat catatan yang bersumber pada bahan-bahan pustaka yang mendukung dan berkaitan dengan penelitian dalam hal ini mengenai data mining Algoritma C4.5, dan *Naive Bayes*.

1.4. Alur Penelitian

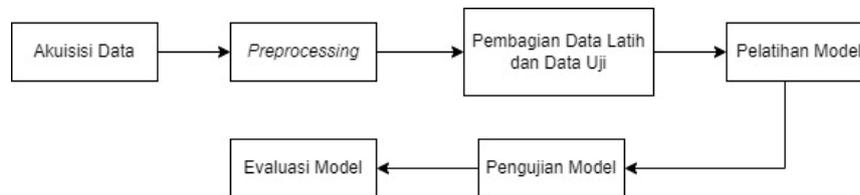
Pada proses pelaksanaan penelitian, terlebih dahulu dilakukan identifikasi masalah melalui observasi dan studi literatur untuk menganalisa masalah dan menemukan kemungkinan-kemungkinan solusi yang dapat dilakukan. Setelah diperoleh kemungkinan solusi yang dapat dilakukan, langkah selanjutnya adalah pengumpulan data penelitian yang diperlukan. Setelah diperoleh data penelitian, langkah selanjutnya adalah membuat rancangan sistem penelitian dan implementasi sistem dimana masing-masing metode akan diimplementasikan secara terpisah untuk dilihat hasil klasifikasinya. Dari hasil klasifikasi kedua metode, kemudian dilakukan komparasi performa kedua metode melalui perhitungan akurasi, *precision*, *recall*, *AUC score*, dan *F1-score*. Alur rancangan penelitian yang akan dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3.1 Skema lengkap alur penelitian

1.5. Gambaran Umum Sistem

Secara keseluruhan, sistem klasifikasi kelayakan nasabah kredit terdiri dari beberapa tahapan yang digambarkan pada Gambar 3.3 berikut.



Gambar 3.2 Rancangan implementasi sistem

Pada Gambar 3.2 digambarkan penelitian yang dibuat dimulai dari proses akuisisi data hingga evaluasi model yang dibuat untuk menghitung nilai akurasi model. Langkah awal yang akan dilakukan pada penelitian ini adalah mengumpulkan dataset yang akan digunakan atau data mentah yang berasal dari data transaksi nasabah milik KSPPS. BMT Adil Berkah Sejahtera. Data kemudian diolah terlebih dahulu melalui proses *preprocessing*. Pada proses ini dilakukan penyaringan data, termasuk data *null* atau data kosong (*NaN*), serta mengubah beberapa nilai yang dinotasikan dengan tanda “,” sebagai ribuan. Proses selanjutnya yaitu menentukan atribut *predictor* serta hasil atau target klasifikasi. Setelah data siap, data kemudian dibagi menjadi data latih dan data uji. Data latih akan digunakan pada saat pelatihan model klasifikasi sedangkan data uji adalah data yang akan digunakan pada proses pengujian model. Setelah dilakukan pengujian terhadap model, dilakukan perhitungan akurasi menggunakan *confusion matrix*.

1.6. Karakteristik Data

Set data (data set/himpunan data) merupakan kumpulan objek dan atributnya. Nama lain dari objek yang sering digunakan diantaranya *record*, *point*, *vector*, *pattern*, *event*, *observation*, *case*, *sample*, *instance*, *entitas*. Objek digambarkan dengan sejumlah atribut yang menerangkan sifat atau karakteristik dari objek tersebut. Data set yang digunakan pada penelitian ini bersumber dari data Anggota/nasabah KSPPS. BMT Adil Berkah Sejahtera. Total data yang terdapat pada dataset ini yaitu 1427 data observasi dengan 8 atribut, yang telah selesai masa kreditnya baik yang macet ataupun yang lancar.

1.7. *Preprocessing*

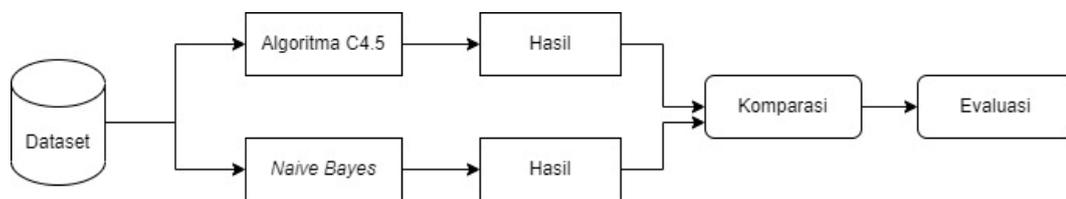
Pada proses pengolahan data mentah menjadi data siap, data terlebih dahulu dibersihkan melalui proses *data cleaning* dengan menghapus data *null* serta data duplikasi. Setelah itu, dilakukan pemilihan atribut mana yang akan dipakai sebagai atribut *predictor* dan menghapus data yang tidak terpakai, seperti nomor rekening dan nama. Target atau kelas yang digunakan pada penelitian ini adalah macet dan lancar yang digunakan untuk indikasi kelayakan nasabah. Setelah ditentukan atribut mana yang akan digunakan, proses dilanjutkan dengan menentukan atribut fitur dan label atau target. Tabel atribut yang akan digunakan dapat dilihat pada Lampiran 1.

1.8. Pembagian Data Latih dan Data Uji

Setelah dilakukan *preprocessing* hingga data mentah menjadi data siap, dilakukan pembagian data latih dan data uji dilakukan menggunakan metode *K-Fold Cross Validation*, data dibagi menjadi 10 subset dimana tiap subsetnya akan dibagi menjadi data latih dan data uji. Pada setiap subset, dilakukan *training* terhadap data latih kemudian dilakukan *testing* terhadap data uji. Proses *training* dan *testing* akan dilakukan terus berulang hingga sepuluh kali, menyesuaikan jumlah subset atau k .

1.9. Klasifikasi

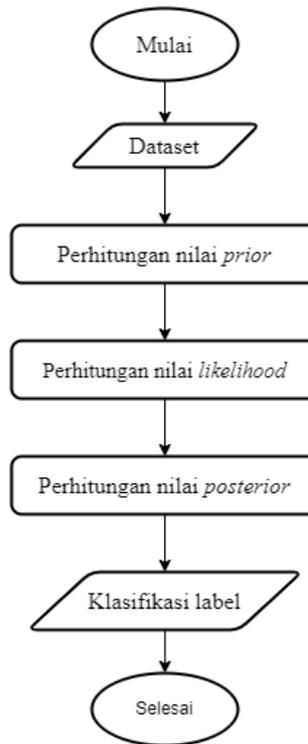
Pada penelitian ini, model klasifikasi yang dibuat menggunakan dua buah metode, yaitu *Naive Bayes* dan algoritma C4.5. Proses klasifikasi, perangkat lunak yang digunakan untuk menganalisis adalah RapidMiner Studio. Proses dilakukan satu per satu terlebih dahulu secara terpisah, yang pertama menggunakan *Naive Bayes* hingga selesai, kemudian dilanjutkan pelatihan menggunakan algoritma C4.5. Hasil akurasi kedua metode kemudian dibandingkan untuk melihat metode mana yang memiliki performa lebih baik. Skema proses perbandingan klasifikasi menggunakan metode *Naive Bayes* dengan algoritma C4.5 dapat dilihat pada Gambar 3.3.



Gambar 3.3 Skema penelitian antara Algoritma C4.5 dan Naive Bayes

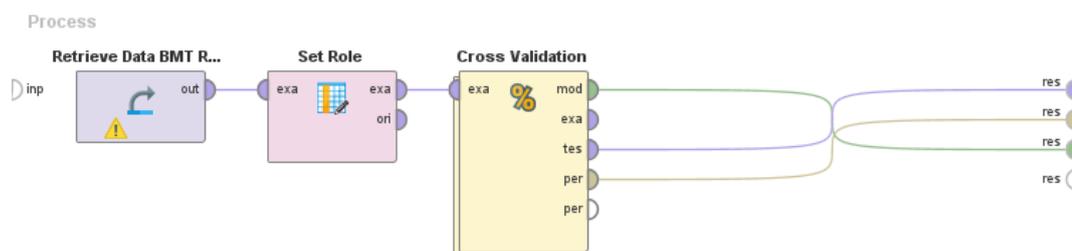
1.9.1. Naive Bayes

Pada implementasi model, setelah data dataset siap, data latih akan diproses menggunakan Naive Bayes untuk dihitung probabilitas tiap kejadian macet atau lancar berdasarkan *prior*, *likelihood*, dan *evidence* yang diketahui seperti pada Persamaan Bab II. Atribut fitur terdiri dari masa, jenis pembiayaan, pokok, *margin*, dan *plafond*, yang dilambangkan sebagai X_1 , X_2 , X_3 , X_4 , dan X_5 sedangkan target atau label macet/lancar dilambangkan dengan Y . Atribut fitur digunakan untuk mencari nilai *likelihood* atau probabilitas kejadian X jika diketahui Y . Nilai prior sendiri dihitung dengan mencari peluang kejadian Y , yaitu besar peluang antara macet dan lancar. Setelah diperoleh nilai *likelihood*, *prior*, dan *evidence* kemudian dicari nilai posterior atau nilai probabilitasnya apakah mendekati nilai macet atau lancar. Untuk nilai *likelihood* pada atribut numerikal, nilai dicari menggunakan distribusi normal dengan terlebih dahulu dicari nilai rata-rata dan standar deviasi dari masing-masing atribut yang bertipe numerikal. *Flowchart* proses Naive Bayes dapat dilihat pada Gambar 3.4.

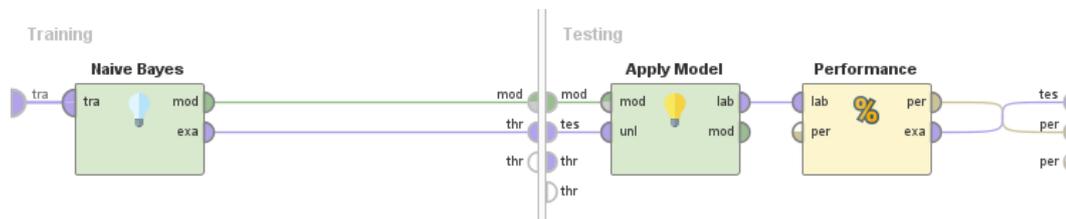


Gambar 3.4 *Flowchart* Naive Bayes

Pada penelitian yang dilakukan, klasifikasi dilakukan menggunakan aplikasi RapidMiner. Dataset bertipe Excel terlebih dahulu diimpor ke dalam lembar kerja pada aplikasi. Dataset kemudian dibaca dan ditentukan *role* atau peran dari masing-masing atribut menggunakan operator *Set Role* pada RapidMiner. Setelah ditentukan peran masing-masing atribut kemudian data dibagi menjadi data latih dan data uji menggunakan operator *cross validation* dengan metode *stratified* serta menggunakan subset atau *k* berjumlah sepuluh. Implementasi pada aplikasi RapidMiner dapat dilihat pada Gambar 3.5 dan Gambar 3.6.



Gambar 3.5 Implementasi Naive Bayes (1)



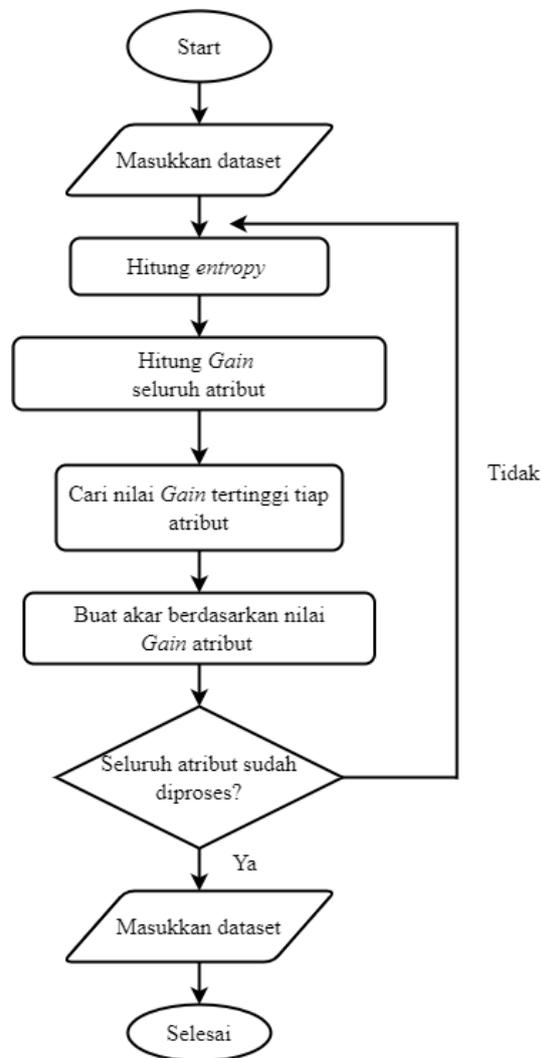
Gambar 3.6 Implementasi Naive Bayes (2)

Pada Gambar 3.6 operator *cross validation* juga berisi operator Naive Bayes sebagai metode yang digunakan untuk melatih model dan operator *Apply Model* digunakan untuk mengimplementasikan model pada proses *testing* untuk memprediksi label dan melihat hasil pengujian model terhadap dataset serta operator *Performance* digunakan untuk mengukur performa model. Hasil klasifikasi ditunjukkan menggunakan akurasi serta nilai AUC atau *Area Under ROC Curve*.

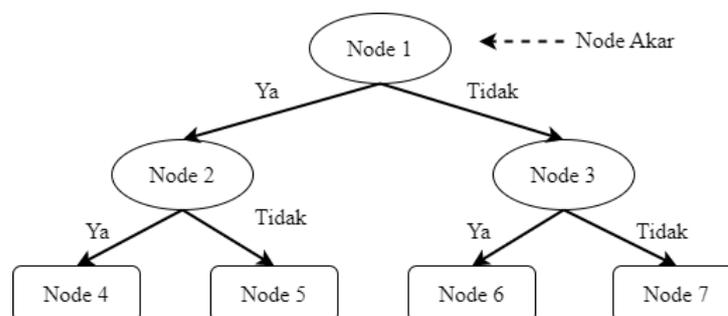
1.9.2. Algoritma C4.5

Algoritma C4.5 merupakan metode yang menggunakan pohon keputusan dalam menentukan atau memprediksi suatu nilai atau label berdasarkan aturan-aturan dalam *tree*. Pada algoritma C4.5 sebelum *rule* terbentuk, terlebih dahulu dibangun *tree* atau pohon yang terdiri dari *root node* atau node akar, *internal node*, serta *leaf*. Pembentukan node tersebut didasarkan pada nilai *Gain* dari masing-masing atribut. Atribut yang memiliki nilai tertinggi akan dijadikan sebagai node pertama atau node akar. Proses pembentukan node tersebut berulang hingga semua atribut terpartisi. Perhitungan *Gain* sendiri dilakukan dengan menghitung *entropy* terlebih dahulu. Dari nilai *entropy* kemudian dicari nilai *Information Gain* serta *Gain*. Karena atribut terdiri dari atribut numerikal dan kategorikal maka penentuan node akar dilakukan dengan menghitung *Gain Ratio*. Setelah semua terpartisi dan terbentuk *tree*, maka dilakukan klasifikasi terhadap dataset untuk menentukan label.

Alur proses algoritma C4.5 dapat dilihat pada Gambar 3.6, sedangkan skema node pada tree dapat dilihat pada Gambar 3.7.

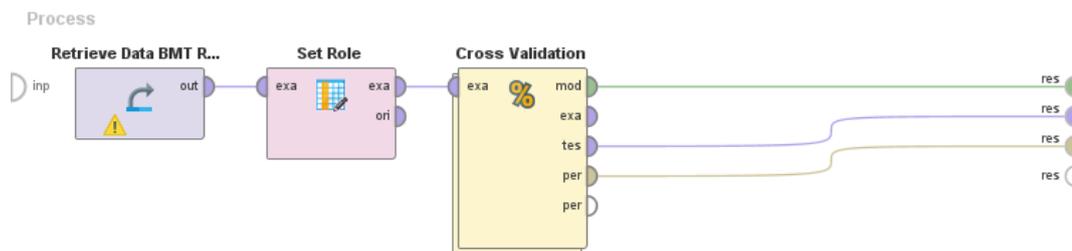


Gambar 3.7 Flowchart algoritma C4.5

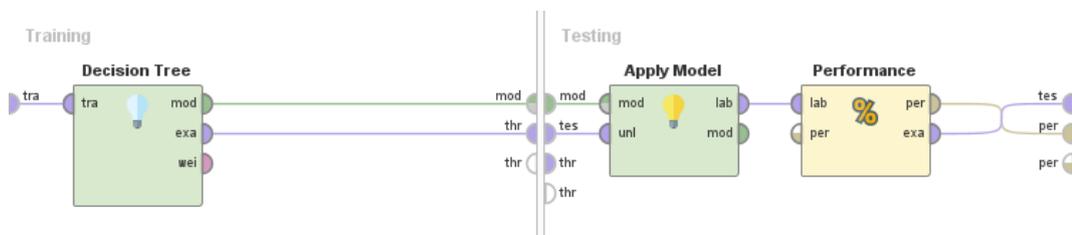


Gambar 3.8 Node pada algoritma C4.5

Proses klasifikasi algoritma C4.5 sendiri diimplementasikan menggunakan aplikasi RapidMiner. Sebagian besar proses klasifikasi menggunakan algoritma C4.5 sama dengan Naive Bayes, hanya berbeda beberapa operator saja. Implementasi algoritma C4.5 dapat dilihat pada Gambar 3.9 dan Gambar 3.10.



Gambar 3.9 Implementasi algoritma C4.5 (1)



Gambar 3.10 Implementasi algoritma C4.5 (2)

Pada Gambar 3.9, terdapat data masukan yang berupa tabel dalam Excel, operator *Set Role* yang digunakan untuk menentukan atribut serta label atau target, serta *cross validation* yang digunakan untuk membagi data latih dan data uji dimana pada model ini juga digunakan *fold* atau *k* sebanyak lima, sepuluh, dan lima belas untuk melihat jumlah *folds* yang memiliki akurasi terbaik. Pengujian ini juga diimplementasikan pada klasifikasi menggunakan Naive Bayes. Sedangkan Gambar 3.10 digambarkan implementasi algoritma C4.5 di dalam *cross validation*. Operator *Apply Model* dan *Performance* juga digunakan untuk mengimplementasikan model pada proses testing serta untuk menguji performa model dalam mengklasifikasikan dataset. Hasil dari *Performance* merupakan akurasi serta grafik dan nilai AUC.

1.10. Evaluasi Model

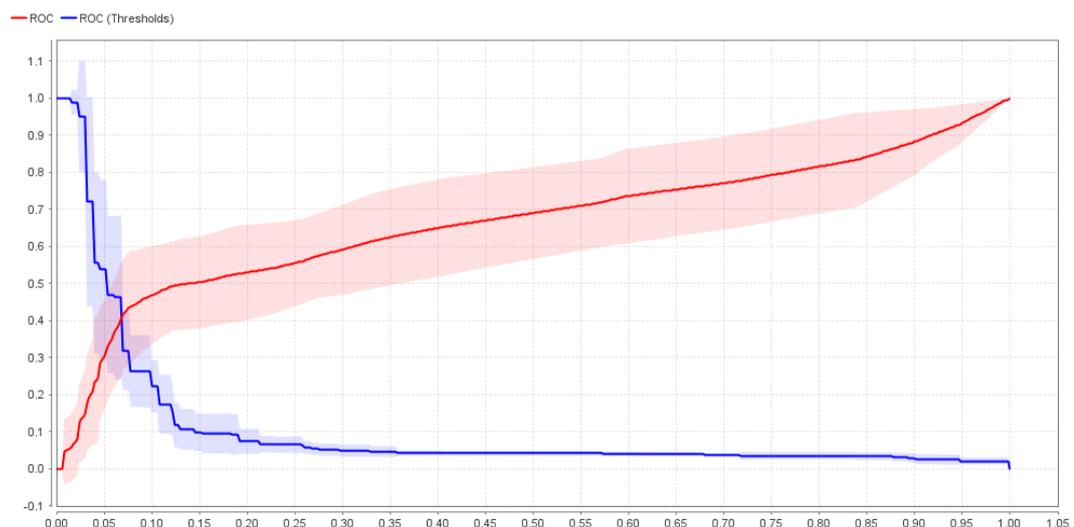
Untuk menentukan performa kedua model dalam mengklasifikasikan data dilakukan perhitungan akurasi menggunakan *confusion matrix*. Pada perhitungan akurasi, hasil pengujian akan dibandingkan dengan klasifikasi sebenarnya menggunakan tabel *confusion matrix*, sehingga diperoleh nilai TP, FP, TN, dan FN. Dari hasil perhitungan tersebut kemudian dihitung akurasinya menggunakan rumus pada Persamaan (2.7) pada bab dua. Contoh hasil *confusion matrix* yang dihasilkan dapat dilihat pada Gambar 3.11.

accuracy: 89.90% +/- 1.05% (micro average: 89.90%)

	true Lancar	true Macet	class precision
pred. Lancar	1276	109	92.13%
pred. Macet	35	6	14.63%
class recall	97.33%	5.22%	

Gambar 3.11 Contoh hasil perhitungan akurasi pada RapidMiner

Selain nilai akurasi menggunakan *confusion matrix* juga diperoleh grafik ROC dari proses pengujian. Contoh grafik ROC dapat dilihat pada Gambar 3.12.



Gambar 3.12 Contoh grafik ROC pada RapidMiner

1.11. Tabel Hasil Komparasi

Dari hasil evaluasi model, dihasilkan akurasi dari kedua model. Nilai akurasi kedua model tersebut kemudian dibandingkan untuk melihat model mana yang menghasilkan akurasi lebih tinggi. Tabel perbandingan yang akan digunakan dapat dilihat pada Tabel 3.1.

Tabel 3.1 Tabel hasil komparasi

No	Metode yang digunakan	Akurasi	AUC	<i>F1-score</i>
1	Algoritma C4.5			
2	Naïve Bayes			

