

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai proses, hasil dan pembahasan penelitian. Sebagai alat bantu pengolahan data, digunakan software RapidMiner. Adapun proses, hasil dan pembahasan dibagi dalam beberapa subbab yakni Pengolahan dan distribusi Data, Implementasi algoritma pada RapidMiner.

4.1 Hasil

Setelah melakukan analisa terhadap perancangan dengan tahapan data mining untuk mengklasifikasi nasabah pengajuan kredit pinjaman pada koperasi Gentiaras Pringsewu Lampung menggunakan algoritma KNN, Naive Bayes, dan Logistic Reggresion. Analisa ini berakhir dengan melakukan proses data mining yang sesungguhnya, maka hasil yang dicapai oleh peneliti merupakan untuk mengetahui klasifikasi calon nasabah baru yang mengajukan pinjaman kredit berdasarkan data nasabah yang digunakan per tanggal 9 september 2021 di koperasi Gentiaras Pringsewu Lampung. Peneliti menggunakan aplikasi RapidMiner dalam melakukan teknik klasifikasi data mining, yang menghasilkan informasi para calon nasabah yang mengajukan pinjaman kredit.

4.1.1 Persiapan Data

A. Data Selection

Pada penelitian ini menggunakan data nasabah koperasi Gentiaras Pringsewu Lampung dan dibagi berdasarkan atribut-atribut. Atribut yang dimaksud terdiri dari lima atribut yaitu nama, nilai pinjaman, jangka waktu, jumlah tanggungan dan predikat, status dan pendidikan terakhir. Dari kelima atribut tersebut menjadi penentu apakah seorang nasabah dapat mengajukan kredit pinjaman dan atribut tersebut dapat dijadikan sebagai acuan oleh pihak koperasi untuk memutuskan apakah permohonan kredit dari seorang nasabah dapat disetujui atau tidak. Adapun atribut yang digunakan dalam penentuan klasifikasi calon nasabah dapat dilihat pada tabel 4.1

Tabel 4.1 Data Atribut yang Digunakan

No	Atribut	Detail Penggunaan
1.	Nama	Digunakan
2.	Nilai Pinjaman	Digunakan
3.	Jangka waktu	Digunakan
4.	Jumlah tanggungan	Digunakan
5.	Predikat	Digunakan
6.	Status	Digunakan
7.	Pendidikan terakhir	Digunakan
8.	Jaminan	Digunakan

B. Integrasi Data (Data Integration)

Tahapan integrasi data adalah tahapan dimana data dijadikan dalam satu kesatuan data untuk proses data mining. Hasil integrasi data dapat dilihat pada tabel 4.2 :

Tabel 4.2 Integrasi Data

No	Nama	Nilai Pinjaman	Jangka waktu	Jlm tanggungan	Status	Pendidikan	Jaminan	Predikat
1	Maria Magdalena Dian Ekawati	24,000,000	25-37	2	Kawin	Strata 1	bpkb	Lancar
2	Legimin	7,000,000	26	2	Kawin	Strata 1	bpkb	Lancar
3	Priwatini (sibuhar jaminan)	55,000,000	12	2	Kawin	Strata 1	bpkb	Tidak Lancar
4	Gregorius Bayu Kristiawan	8,000,000	25-37	2	Kawin	Strata 1	Sertifikat rumah	Lancar
5	Muryanto	57,000,000	18	2	Kawin	Strata 1	Sertifikat Tanah	Tidak Lancar
....
501	M.said	4,000,000	10	2	Kawin	Strata 1	bpkb	Tidak Lancar

C. Preprocessing

Tahap preprocessing atau pembersihan data bertujuan untuk menghilangkan data noise (data yang tidak relevan / berhubungan langsung dengan tujuan akhir proses data mining). Proses cleaning mencakup antara lain membuang duplikasi data , memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Proses cleaning pada data anggota pinjaman dapat dilihat pada tabel 4.3.

Tabel 4.3 Data Preprocessing

No	Nama	Nilai Pinjaman	Jangka waktu	Jlm tanggungan	Status	Pendidikan	Jaminan	Predikat
1	Maria Magdalena Dian Ekawati	24,000,000	25-37	2	Kawin	Strata 1	bpkb	Lancar
2	Legimin	7,000,000	26	2	Kawin	Strata 1	bpkb	Lancar
3	Priwatini (sibuhar jaminan)	55,000,000	12	2	Kawin	Strata 1	bpkb	Tidak Lancar
4	Gregorius Bayu Kristiawan	8,000,000	25-37	2	Kawin	Strata 1	Sertifikat rumah	Lancar
5	Muryanto	57,000,000	18	2	Kawin	Strata 1	Sertifikat Tanah	Tidak Lancar
....
501	M.said	4,000,000	10	2	Kawin	Strata 1	bpkb	Tidak Lancar

4.1.2 Implementasi RapidMiner

Tujuan utama dalam penelitian ini untuk mengklasifikasi calon nasabah kredit pinjaman dan untuk mengetahui tingkat akurasi dari algoritma KNN, Naive Bayes dan Logistic Regression dalam klasifikasi nasabah kredit pinjaman di koperasi Gentiaras Lampung. Pengujian untuk klasifikasi data nasabah menggunakan software RapidMiner.

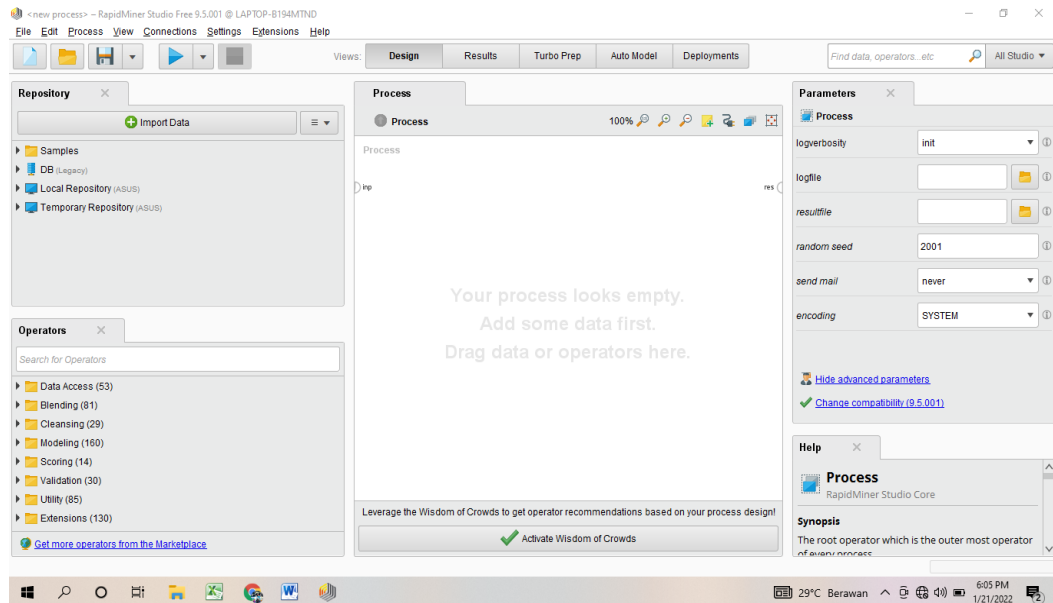
4.1.2.1 Pengenalan RapidMiner

Berikut merupakan tampilan awal ketika aplikasi rapidminer ketika dijalankan.



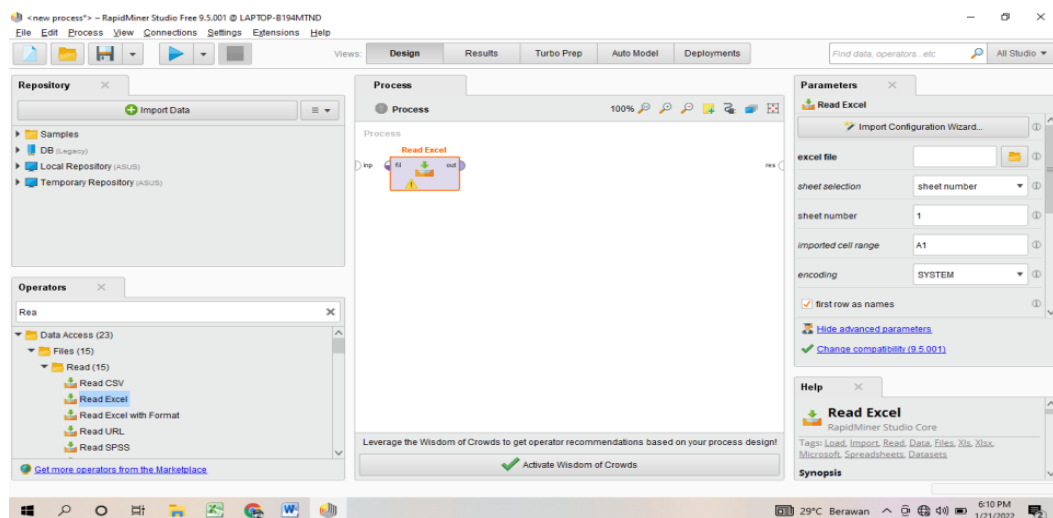
Gambar 4.1 Tampilan Awal RapidMiner

Setelah muncul tampilan logo dan informasi software RapidMiner maka muncul tampilan layar membuat lembar kerja baru atau membuka file yang sudah ada. Proses selanjutnya pengambilan data dibutuhkan operator read excel dengan dilakukan drag dan drop kedalam view process sehingga operator read excel tampil dalam view process seperti Gambar 4. Selanjutnya klik Import Configuration Wizard Gambar 4.2



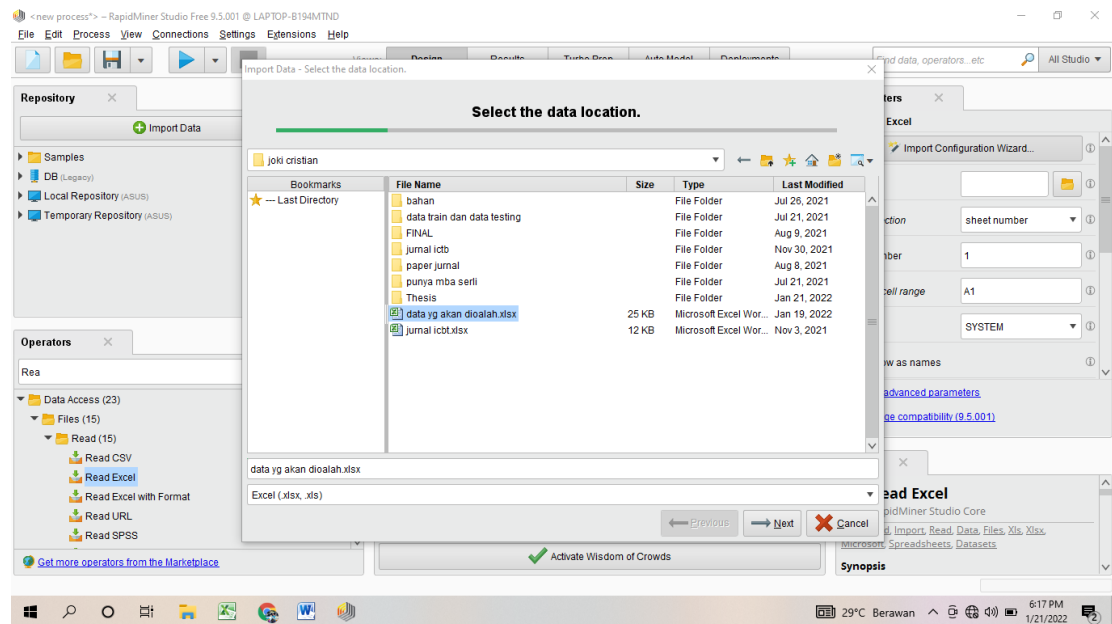
Gambar 4.2 Halaman Lembar Kerja RapidMiner

Proses selanjutnya pengambilan data dibutuhkan operator read excel dengan dilakukan drag dan drop kedalam view process sehingga operator read excel tampil dalam view process seperti Gambar 4.3. Selanjutnya klik Import Configuration Wizard.



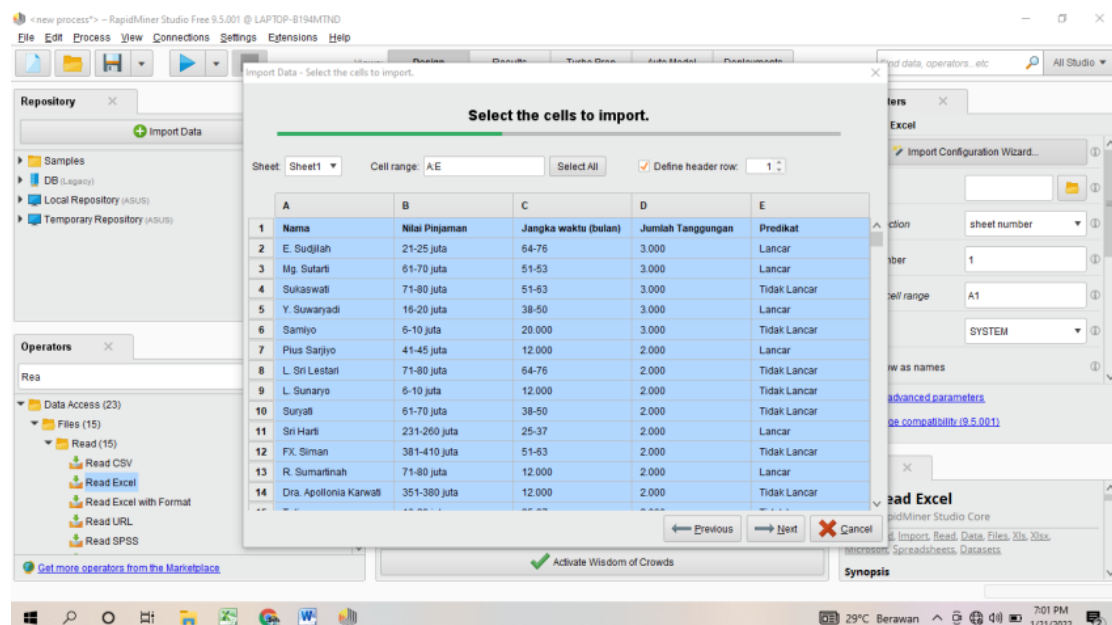
Gambar 4.3 import Configuration Wizard

Setelah klik import configuration wizard maka selanjutnya form data import wizard tahap 1 untuk memilih tempat lokasi file yang digunakan penelitian seperti Gambar 4.4.

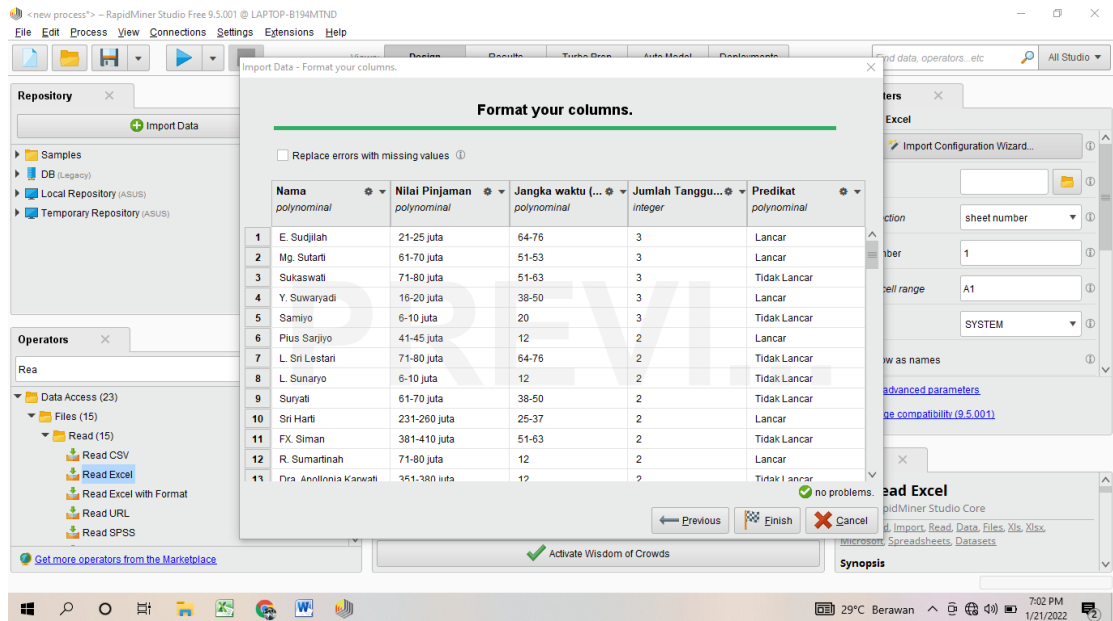


Gambar 4.4 Alur Proses Import Data

Data yang telah dipilih selanjutnya yaitu klik Next kemudian akan muncul form data import wizard tahap 2 seperti Gambar 4.5 dan Gambar 4.6

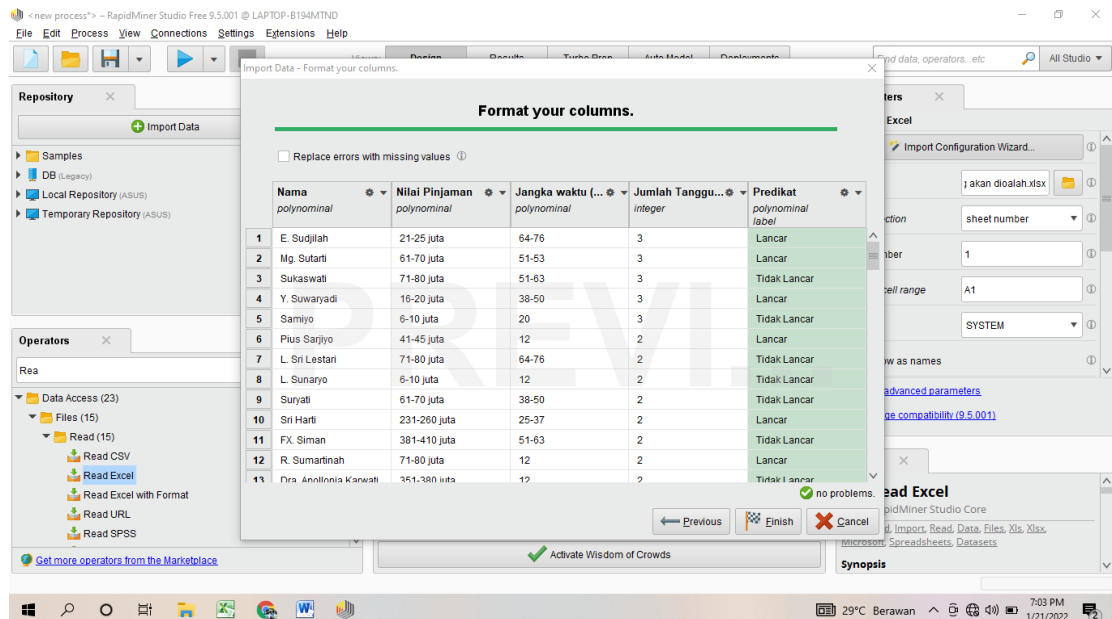


Gambar 4.5 Alur Proses Import Data



Gambar 4.6 Alur Proses Import Data

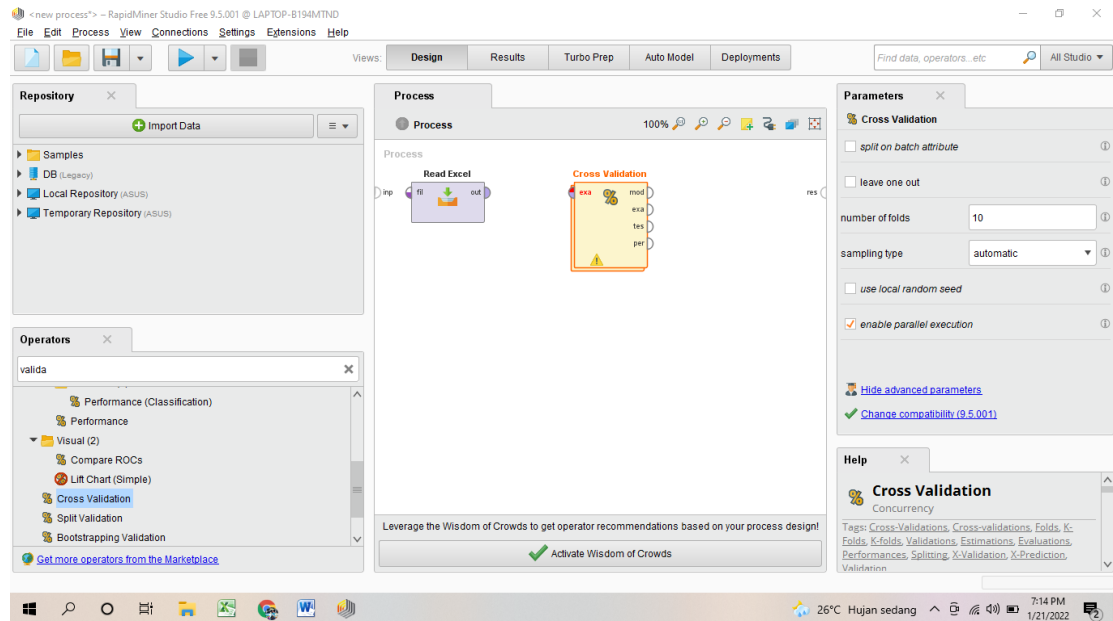
Tahap 3 tidak melakukan apapun maka langsung klik next untuk menuju tahap 4 yang akan muncul form data import wizard tahap 4 seperti Gambar 4.7.



Gambar 4.7 Alur Proses Import Data

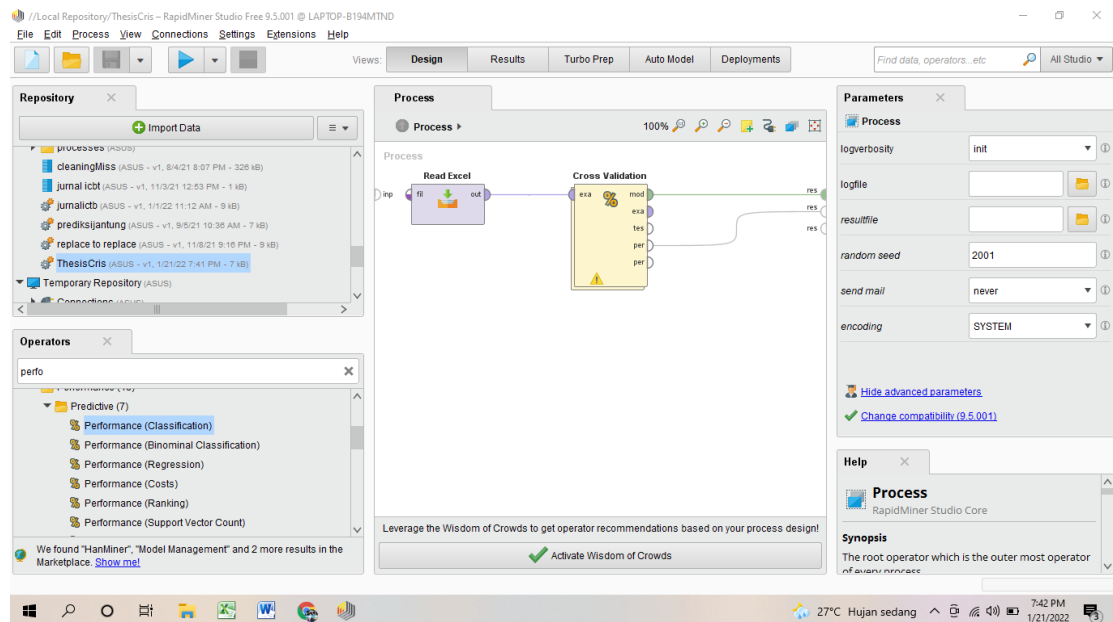
Tahap 4 seperti gambar di atas perlu diperhatikan sebelum langkah ke tahap selanjutnya pilih salah satu atribut label atau target karena pada klasifikasi tentu ada atribut menjadi label atau target maka yang dipilih sebagai label adalah atribut keputusan dan selanjutnya maka klik finish.

Data yang telah selesai diimport maka langkah selanjutnya drag and drop Validation.



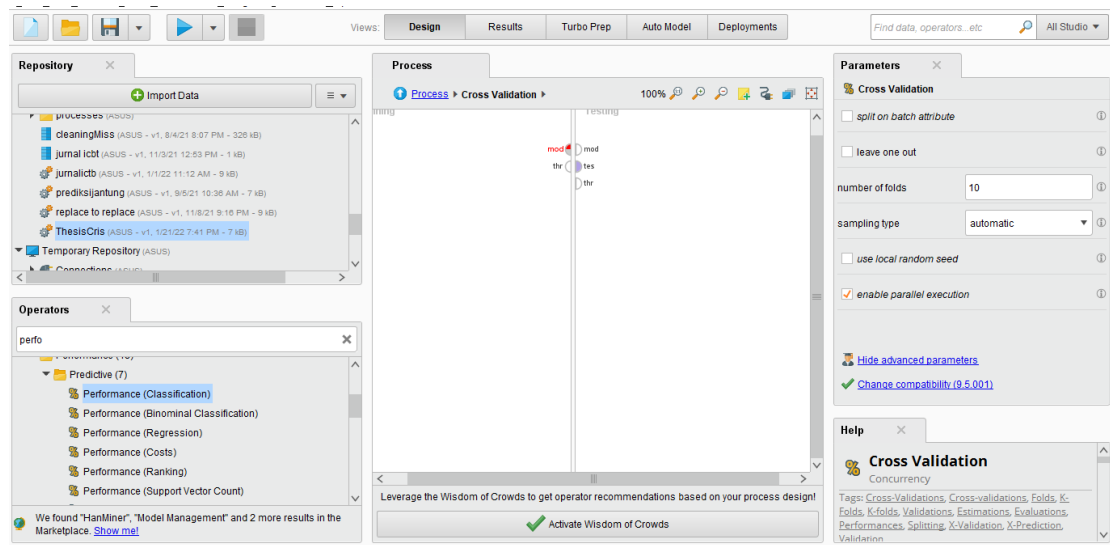
Gambar 4.8 Operator Read Excel dan Validation

Validation ini dapat memvalidasi data yang modelnya algoritma dan fungsi Validation dapat memaksimalkan nilai akurasi pengolahan data. Langkah selanjutnya hubungkan operator read excel dengan Validation dengan menarik garis tabel read excel ke operator Validation dan menarik garis lagi operator Validation ke result di sisi kanan seperti Gambar 4.9



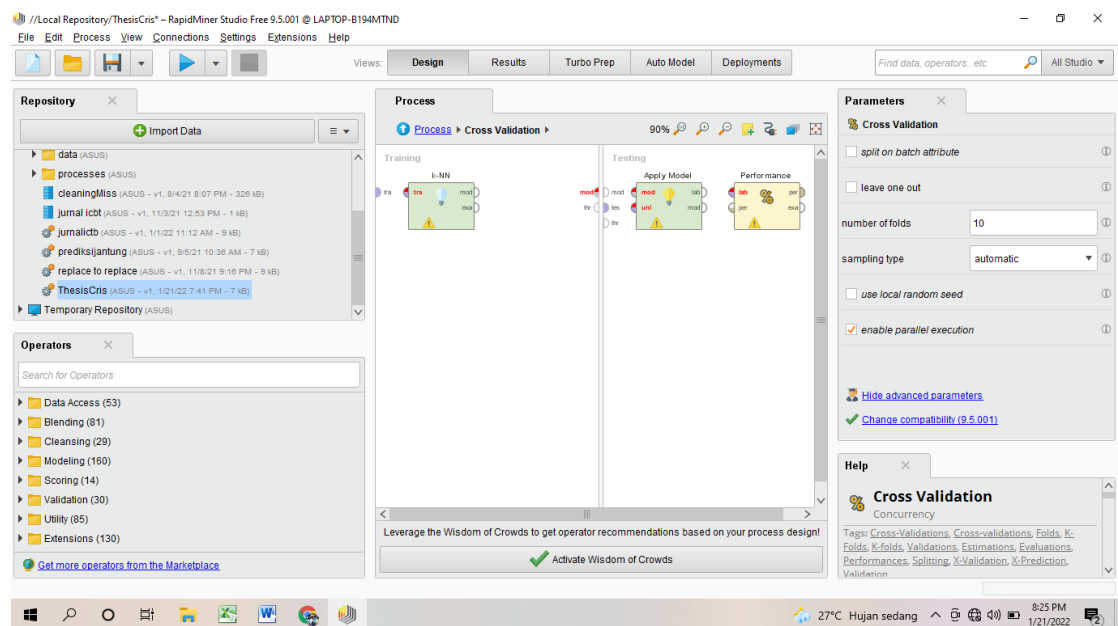
Gambar 4.9 Menghubungkan Read Excel dengan Operator Validation

Operator Validation memiliki port input yaitu, training ExampleSet sebagai port input untuk memperkirakan ExampleSet dalam melatih sebuah model. ExampleSet juga dapat menguji semua model dan memiliki port output sebagai berikut:



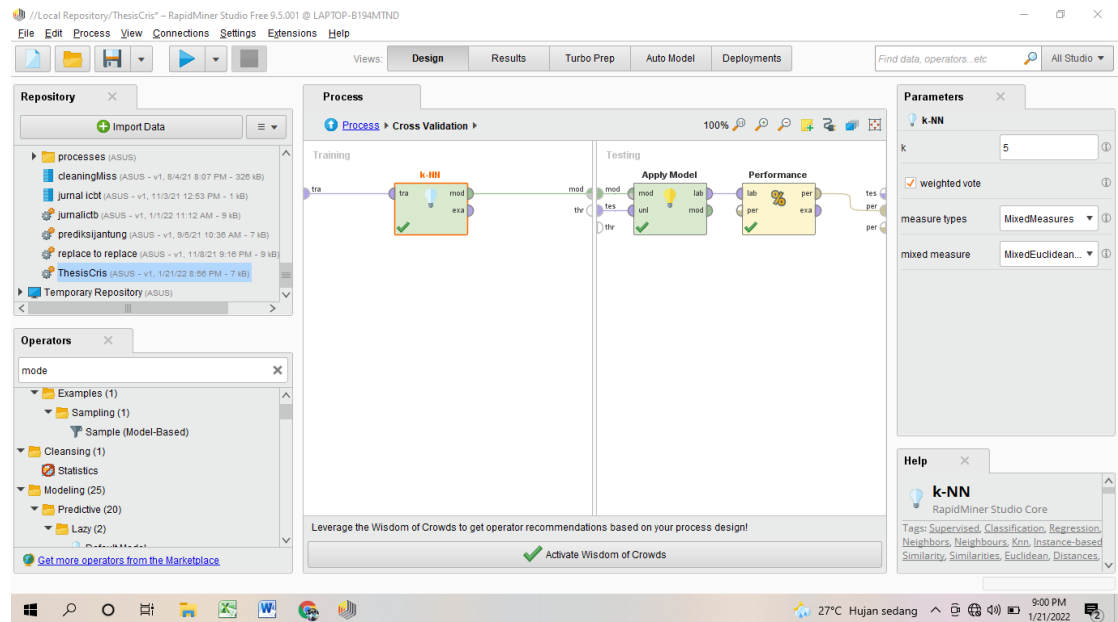
Gambar 4.10 Validation

Tampilan port out otomatis muncul form seperti gambar di atas maka selanjutnya drag dan drop algoritma yang akan diuji, karena penelitian ini menggunakan Naive Bayes maka drag dan drop Naive Bayes pada box training dan pada box testing drag dan drop apply model dan performance.



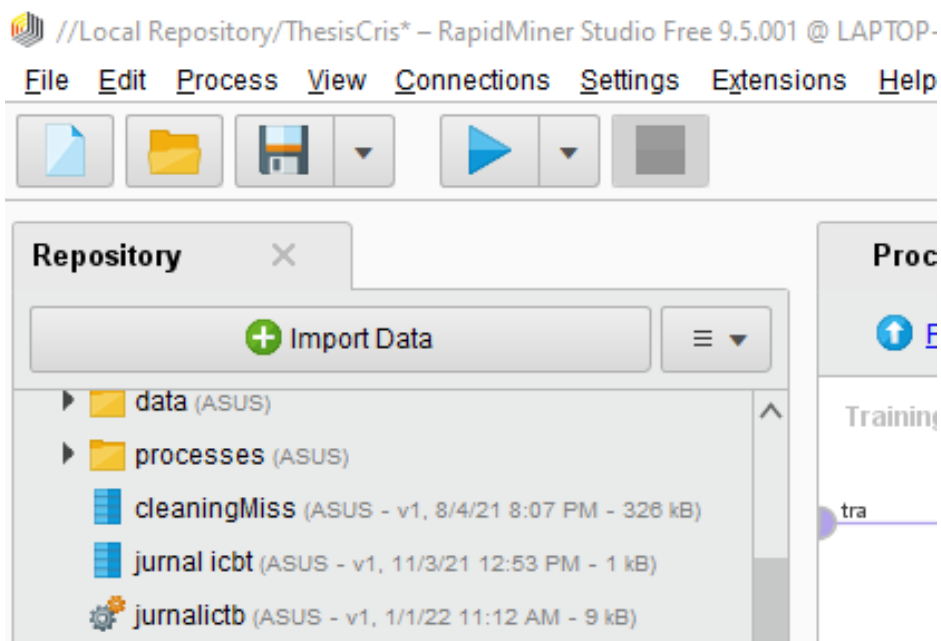
Gambar 4.11 Pemodelan Validation

Langkah selanjutnya hubungkan operator KNN dengan Apply Model dan Performance dengan menarik garis tabel KNN ke operator Apply Model dan menarik garis lagi operator Performance ke result di sisi kanan seperti Gambar 4.12

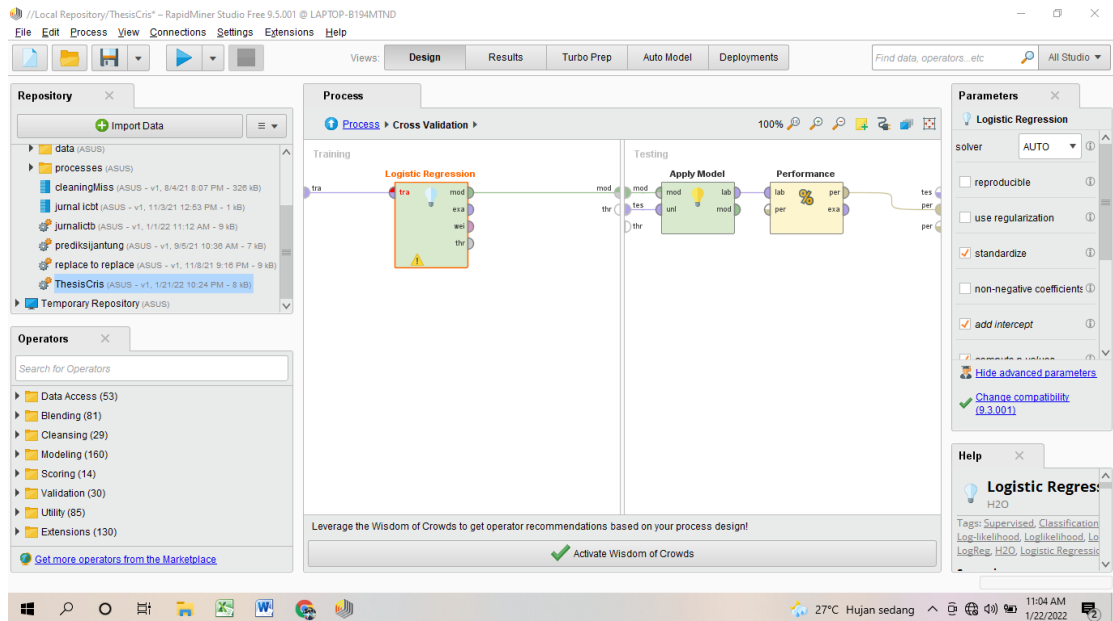


Gambar 4.12 Susunan Operator KNN, Apply Model, dan Performance

Langkah terakhir adalah klik ikon Run pada toolbar untuk menampilkan hasilnya, seperti pada Gambar 4.13. Tunggu beberapa saat komputer untuk menyelesaikan perhitungannya, kemudian lakukan hal yang sama dengan menggunakan algoritma Naive Bayes dan Logistic regression seperti pada gambar 4. Dan 4.



Gambar 4.13 Ikon Run



Gambar 4.14 Susunan Operator Logistic Regression, Apply Model, dan Performance

4.1.3 Implementasi RapidMiner Dataset Koperasi.

Tahapan selanjutnya yaitu melakukan uji coba dataset nasabah kredit pinjaman koperasi Gentiaras Lampung dengan data yang telah disiapkan. Jumlah dataset koperasi yang digunakan sebanyak 501 data nasabah. Dataset yang digunakan dapat dilihat pada tabel 4.4:

Tabel 4.4 Dataset Nasabah

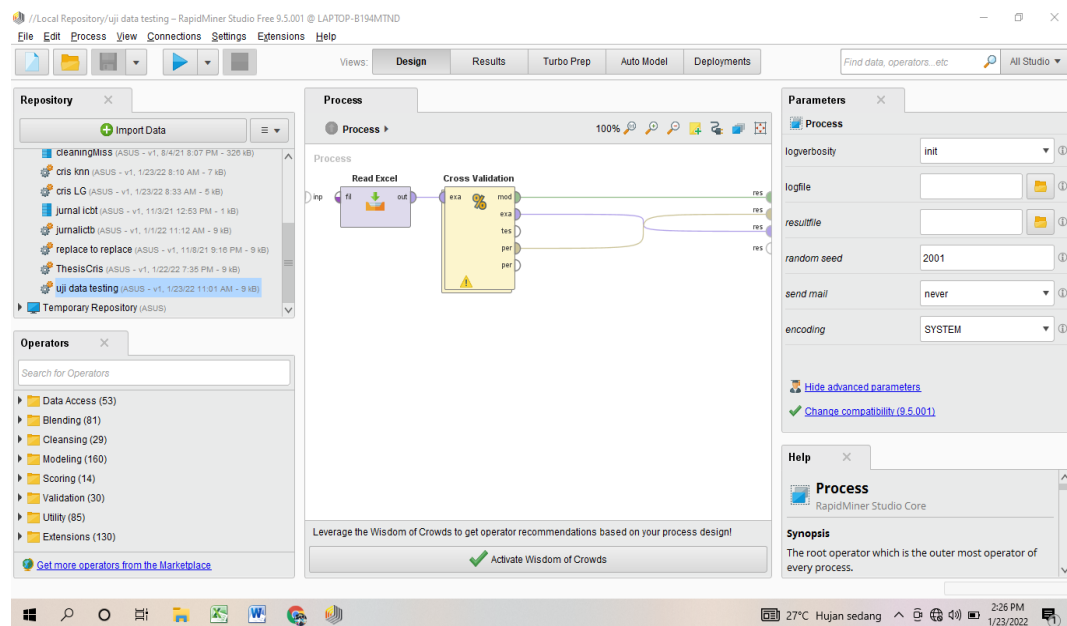
No	Nama	Nilai Pinjaman	Jangka waktu	Jlm tanggungan	Status	Pendidikan	Jaminan	Predikat
1	Maria Magdalena Dian Ekawati	24,000,000	25-37	2	Kawin	Strata 1	bpkb	Lancar
2	Legimin	7,000,000	26	2	Kawin	Strata 1	bpkb	Lancar
3	Priwatini (sibuhar jaminan)	55,000,000	12	2	Kawin	Strata 1	bpkb	Tidak Lancar
4	Gregorius Bayu Kristiawan	8,000,000	25-37	2	Kawin	Strata 1	Sertifikat rumah	Lancar
5	Muryanto	57,000,000	18	2	Kawin	Strata 1	Sertifikat Tanah	Tidak Lancar
....
501	M.said	4,000,000	10	2	Kawin	Strata 1	bpkb	Tidak Lancar

4.1.4 Uji Performa Dataset

Setelah data testing mendapatkan label data yang telah dilakukan pada tahap sebelumnya, maka langkah selanjutnya yaitu menguji performa akurasi, presisi dan recall terhadap algoritma KNN, Naïve Bayes dan Logistic Regression. Pada bagian ini akan dipaparkan mengenai implementasi dari algoritma yang akan dibandingkan yaitu KNN, Logistic Regression dan Naïve Bayes menggunakan Software RapidMiner untuk mendapatkan analisa klasifikasi kredit Lancar dan Tidak Lancar.

4.1.4.1 Proses Pemodelan

Proses ini menggunakan k fold cross validation dimana metode k fold cross validation digunakan dalam memperkirakan kesalahan prediksi untuk evaluasi kinerja model. Metode ini memiliki prinsip dasar membagi keseluruhan data menjadi data training dan data testing. . Data dibagi sebanyak k bagian dengan metode k fold cross validation lalu dilakukan analisis diskriman dengan mencari model diskriminan, untuk melakukan pengujian model setelah sebelumnya data telah dibaca dengan operator read excel, seperti pada gambar 4.15



Gambar 4.15 Read Excel dan Validation Data Training dan Testing

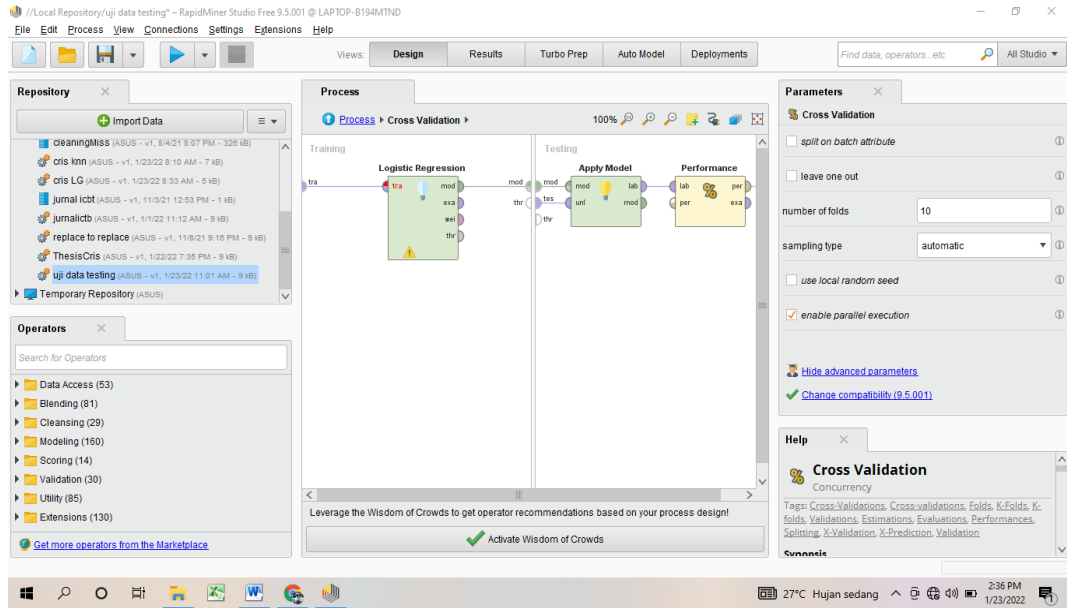
Pada gambar diatas dilakukan proses pengujian model yang telah dibaca menggunakan operator read excel, data yang telah dibaca tersebut kemudian

dimasukkan kedalam operator cross validation, dalam penelitian ini cross validation yang digunakan 2 sampai 10 fold validation untuk mengetahui fold terbaik. K fold cross validation digunakan untuk mengestimasi kesalahan prediksi dalam mengevaluasi kinerja model. Data dibagi menjadi himpunan bagian k berjumlah hampir sama. Model dalam klasifikasi dilatih dan diuji sebanyak k. Disetiap pengulangan, salah satu himpunan bagian akan digunakan sebagai data training dan data testing. Langkah-langkah dari k fold cross validation yaitu:

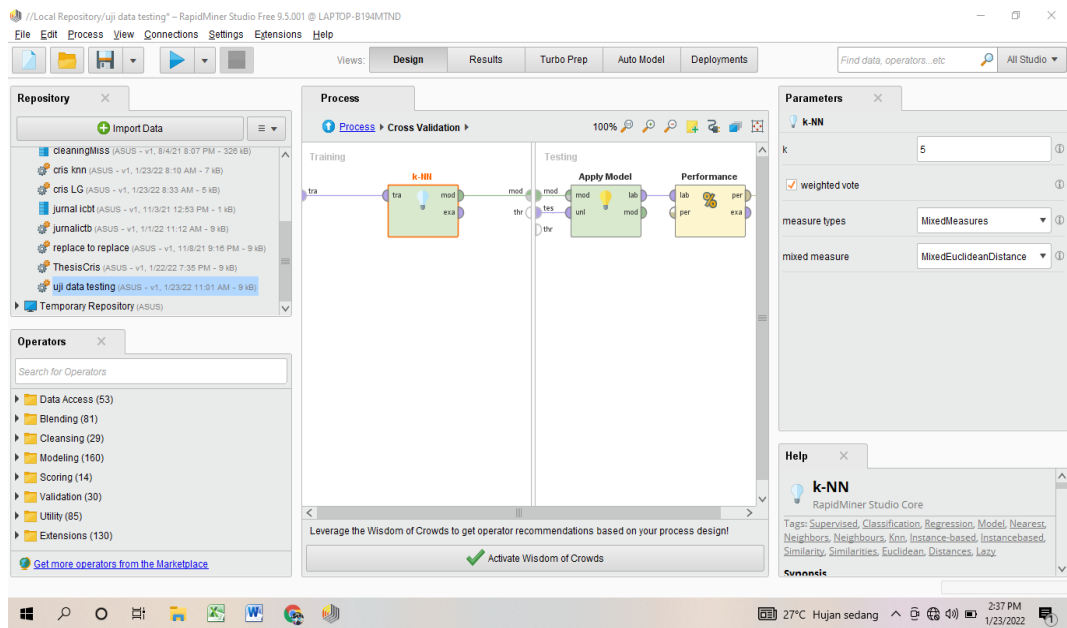
1. Total data dibagi menjadi k bagian.
2. Fold ke-1 adalah ketika bagian ke-1 menjadi data uji (testing data) dan sisanya menjadi data latih (training data). kemudian, hitung akurasi atau kesamaan atau kedekatan suatu hasil pengukuran dengan angka atau data yang sebenarnya berdasarkan porsi data tersebut.
3. Fold ke-2 adalah ketika bagian ke-2 menjadi data uji (testing data) dan sisanya menjadi data latih (training data). kemudian hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya hingga mencapai fold ke-k. Hitung rata-rata akurasi dari k buah akurasi diatas. Rata-rata akurasi ini menjadi akurasi final.

4.1.4.2 Proses Pemodelan Dataset

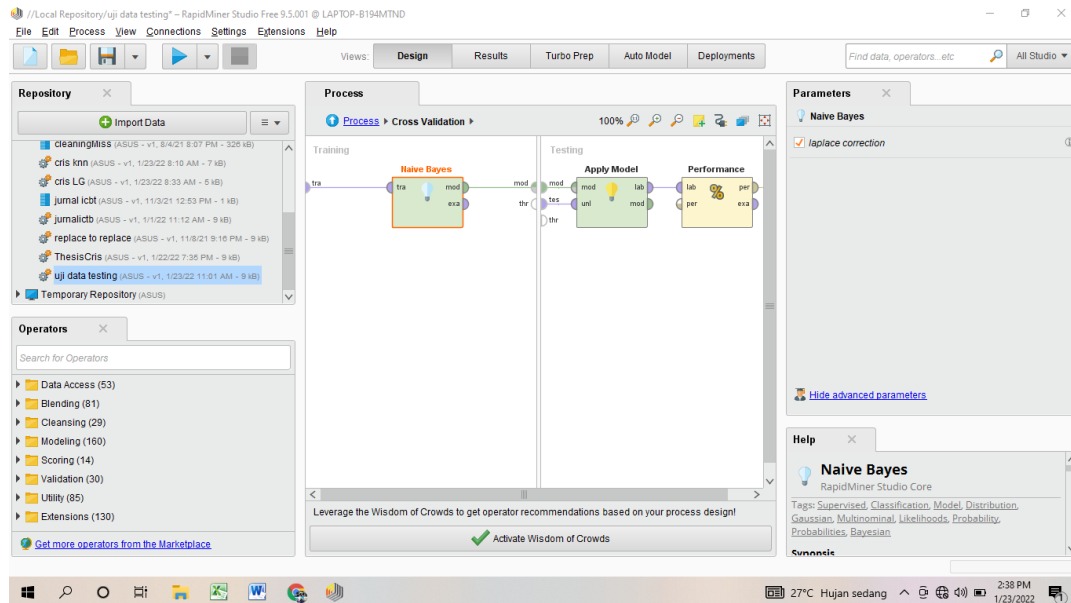
Proses selanjutnya yang dilakukan adalah proses training, proses ini dilakukan di dalam operator cross validation dengan menggunakan algoritma KNN, Naïve Bayes dan Logistic Regression, dengan ketiga algoritma ini maka data akan dibagi menjadi 2, bagian pertama digunakan sebagai data training dan bagian kedua digunakan sebagai data testing, seperti pada gambar 4.16, gambar 4.17 dan gambar 4.18 dibawah ini.



Gambar 4.16 Pemodelan Algoritma Logistic Regression



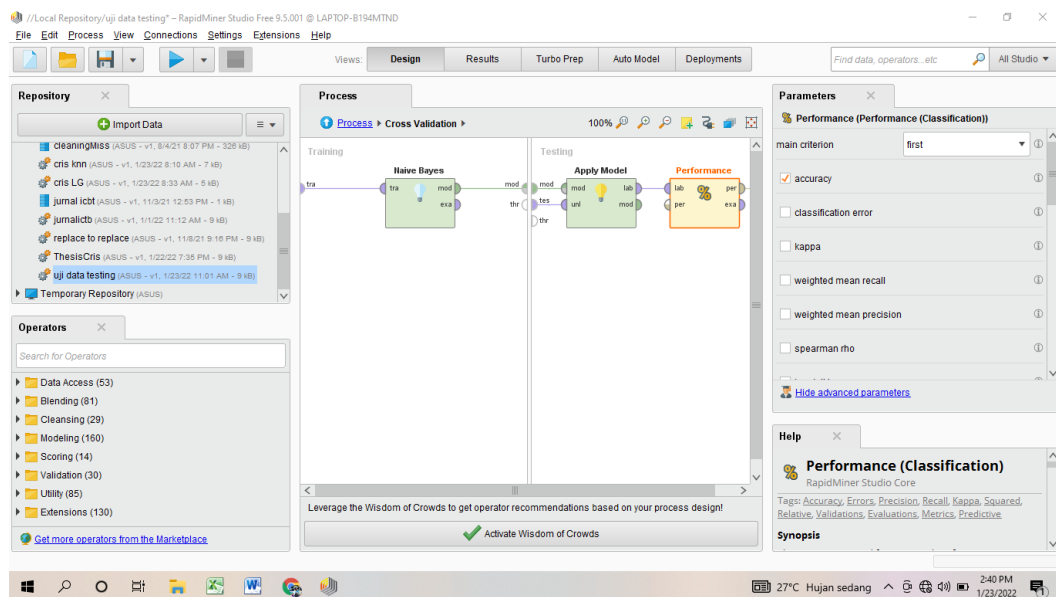
Gambar 4.17 Pemodelan Algoritma KNN



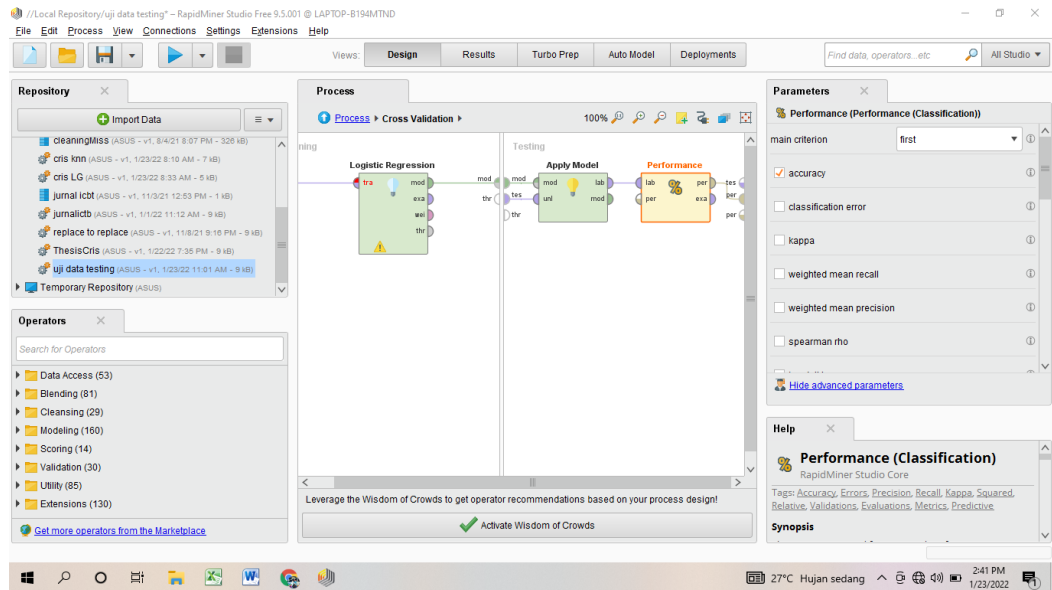
Gambar 4.18 Pemodelan Algoritma Naïve Bayes

4.1.4.3 Proses Pemodelan Performance Data

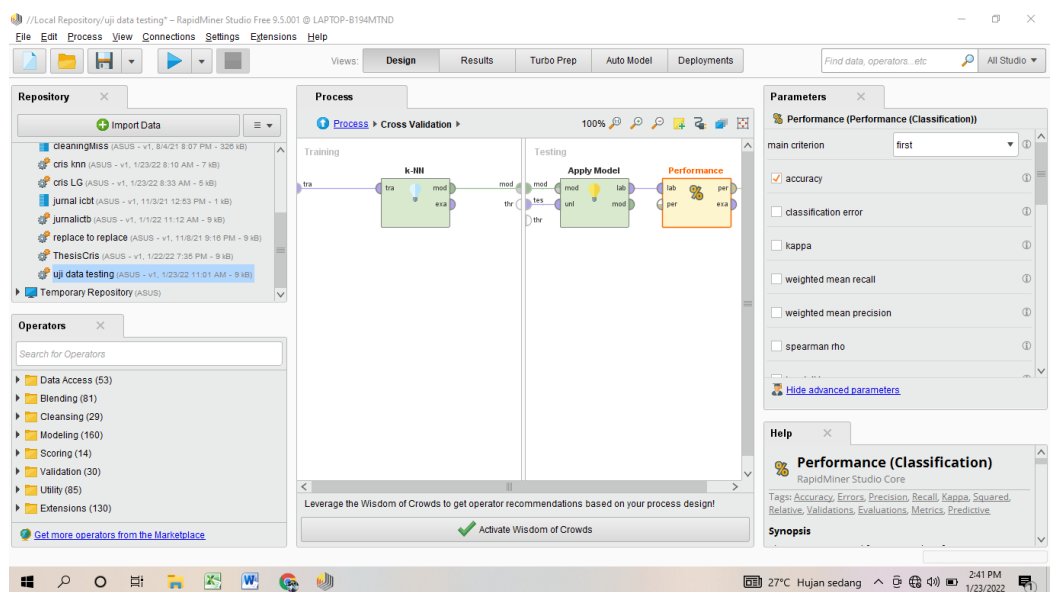
Tahapan terakhir dari proses ini setelah dilakukan training adalah testing terhadap dataset menggunakan fungsi backpropagation yang telah dilakukan pada data training sebelumnya, proses testing ini dilakukan dengan cara insert apply model dan performance dengan main criterion accuracy seperti gambar dibawah ini :



Gambar 4.19 Pemodelan Performance Naïve Bayes



Gambar 4.20 Pemodelan Performance Logistic Regression



Gambar 4.21 Pemodelan Performance KNN

4.1.4.4 Hasil Pemodelan Cross Validation

Hasil dari pemodelan cross validation ketiga algoritma tersebut dengan menggunakan nomor fold 2,3,4,5,6,7,8,9 dan 10 terhadap data nasabah koperasi Gentiaras Pringsewu Lampung yang mengajukan kredit pinjaman dapat kita lihat pada tabel 4.5 dibawah ini.

Tabel 4.5 Hasil Akurasi Pemodelan Cross Validation

Fold	Naïve Bayes	KNN	Logistic Regression
2	77.04%	70.05%	50.32%
3	77.64%	76.05%	63.27%
4	76.05%	72.87%	52.52%
5	76.44%	75.25%	68.07%
6	76.86%	76.64%	66.07%
7	76.43%	75.82%	67.87%
8	76.45%	75.85%	64.30%
9	77.46%	76.25%	67.09%
10	76.85%	75.84%	68.06%

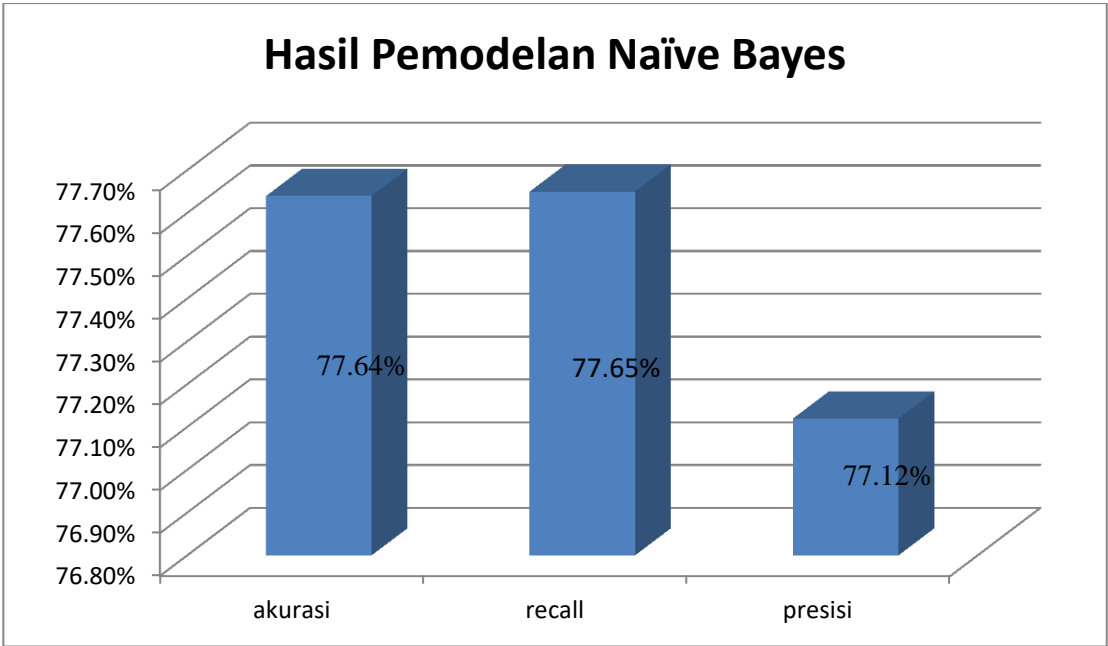
Tabel 4.6 Hasil Recall Pemodelan Cross Validation

Fold	Naïve Bayes	KNN	Logistic Regression
2	76.99%	67.18%	50.34%
3	77.65%	74.66%	66.77%
4	75.85%	71.85%	66.77%
5	76.29%	74.47%	57.99%
6	76.82%	75.85%	70.36%
7	76.53%	75.36%	68.49%
8	76.22%	75.71%	70.18%
9	77.38%	75.42%	67.37%
10	76.66%	75.48%	70.31%

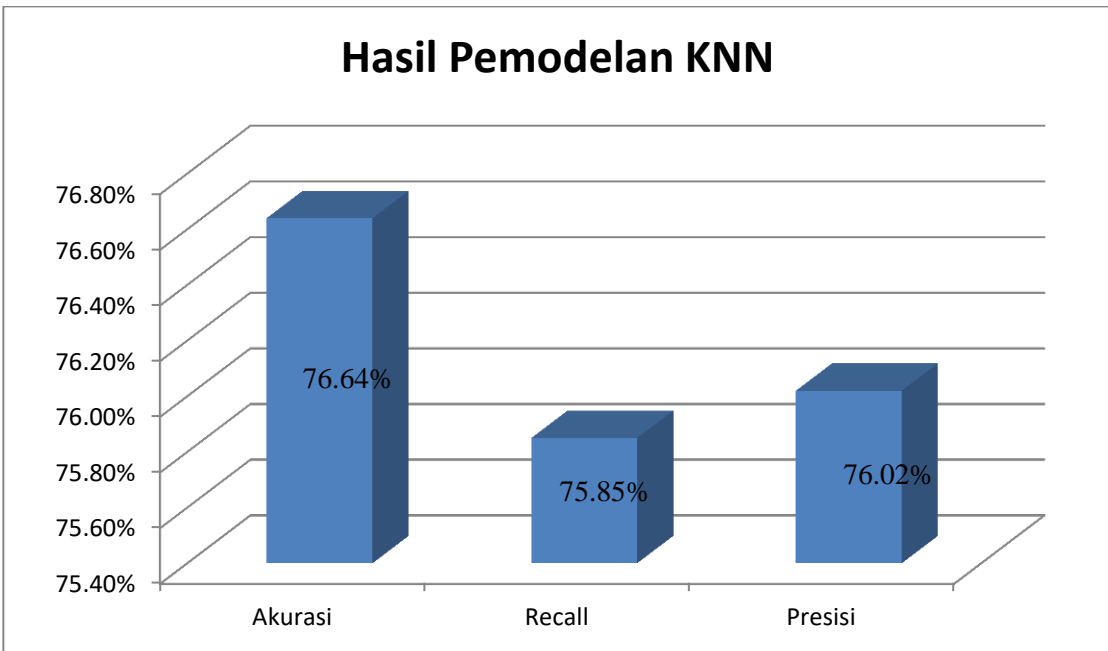
Tabel 4.7 Hasil Presisi Pemodelan Cross Validation

Fold	Naïve Bayes	KNN	Logistic Regression
2	76.42%	69.44%	56.29%
3	77.12%	75.49%	68.41%
4	75.48%	72.34%	68.41%
5	76.10%	74.47%	54.34%
6	76.41%	76.02%	70.99%
7	76.01%	75.39%	69.37%
8	75.94%	75.31%	70.37%
9	77.03%	75.68%	64.74%
10	76.78%	75.42%	71.13%

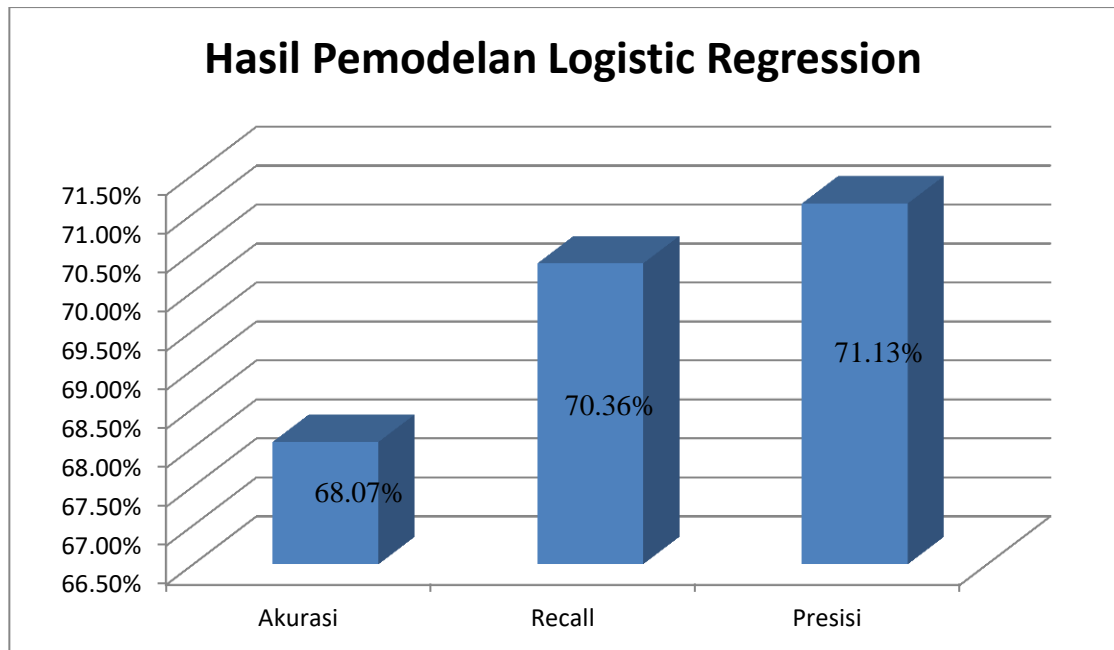
Dari hasil pemodelan cross validation yang ditunjukkan pada tabel 4.5, tabel 4.6 dan tabel 4.7 maka dapat diketahui hasil yang paling baik dari pemodelan cross validation pada algoritma naïve bayes, knn dan logistic regression dengan menerapkan uji fold 2 hingga 10, yang divisualisasikan pada gambar 4.22, gambar 4.23 dan 4.24 di bawah ini.



Gambar 4.22 Visualisasi Hasil Pemodelan Naïve Bayes



Gambar 4.23 Visualisasi Hasil Pemodelan KNN

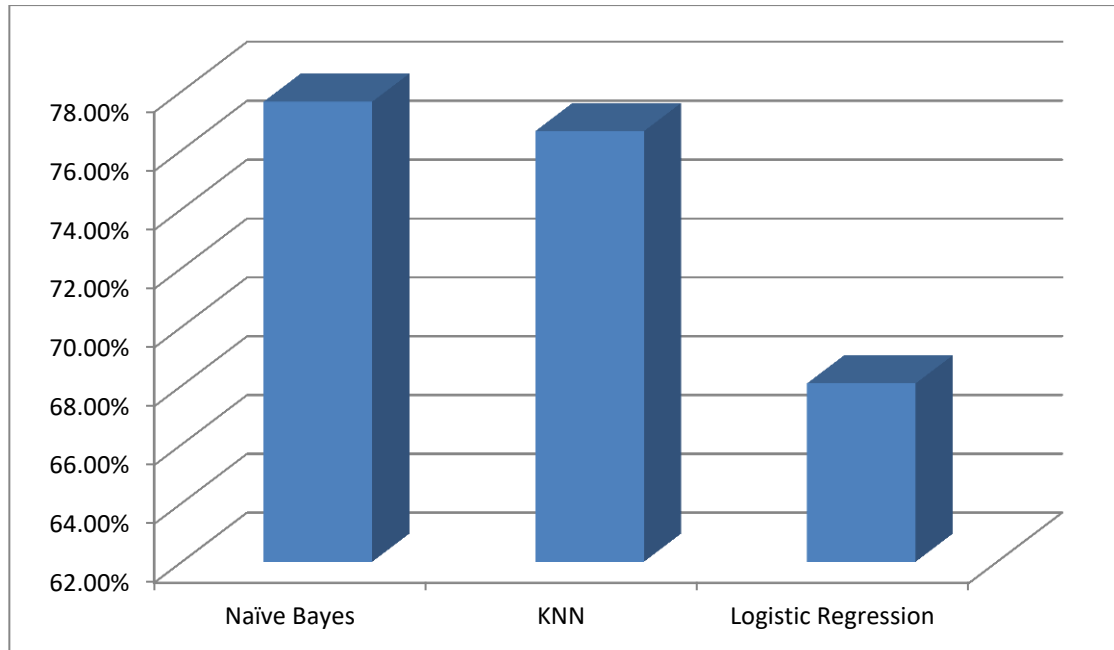


Gambar 4.24 Visualisasi Hasil Pemodelan Logistic Regression

4.2. Pembahasan

Penelitian yang telah dilakukan didapatkan hasil sebuah pola informasi dalam menggunakan proses data mining untuk mengklasifikasi para calon nasabah kredit pinjaman pada koperasi Gentiaras Pringsewu Lampung. Penelitian ini menghasilkan suatu pola informasi yang sesuai dengan tujuan data mining yaitu pola data training dan data testing untuk mengklasifikasi para calon nasabah koperasi dari setiap atribut yang menggunakan data training dan data testing untuk mendapatkan informasi baru, apakah data para calon nasabah termasuk kedalam nasabah yang Lancar atau Tidak Lancar. Proses klasifikasi menggunakan algoritma Naïve Bayes, KNN dan Logistic Regression.

Setelah proses pengklasifikasian telah selesai selanjutnya dilakukan proses pencarian tingkat akurasi dari ketiga algoritma tersebut. Proses klasifikasi dan pencarian nilai akurasi menggunakan software rapidminer. Dari proses pencarian nilai akurasi dari algoritma Naïve Bayes, KNN dan Logistic Regression maka diperoleh nilai akurasi Naïve Bayes sebesar : 77.64%, Logistic Regression sebesar: 68.07% dan KNN sebesar 76.64%. Visualisasi perbandingan akurasi ketiga algoritma tersebut dapat dilihat pada Gambar 4.25.



Gambar 4.25 Perbandingan Nilai Akurasi

Nilai akurasi 47.92 membuktikan bahwa algoritma Logistic Regression kurang baik digunakan untuk melakukan klasifikasi, nilai akurasi 77.64% membuktikan bahwa algoritma Naïve Bayes cukup baik digunakan untuk melakukan klasifikasi dan nilai akurasi 77.64% membuktikan bahwa algoritma KNN sangat baik digunakan untuk melakukan klasifikasi calon nasabah kredit pinjaman koperasi Gentaras Pringsewu Lampung.