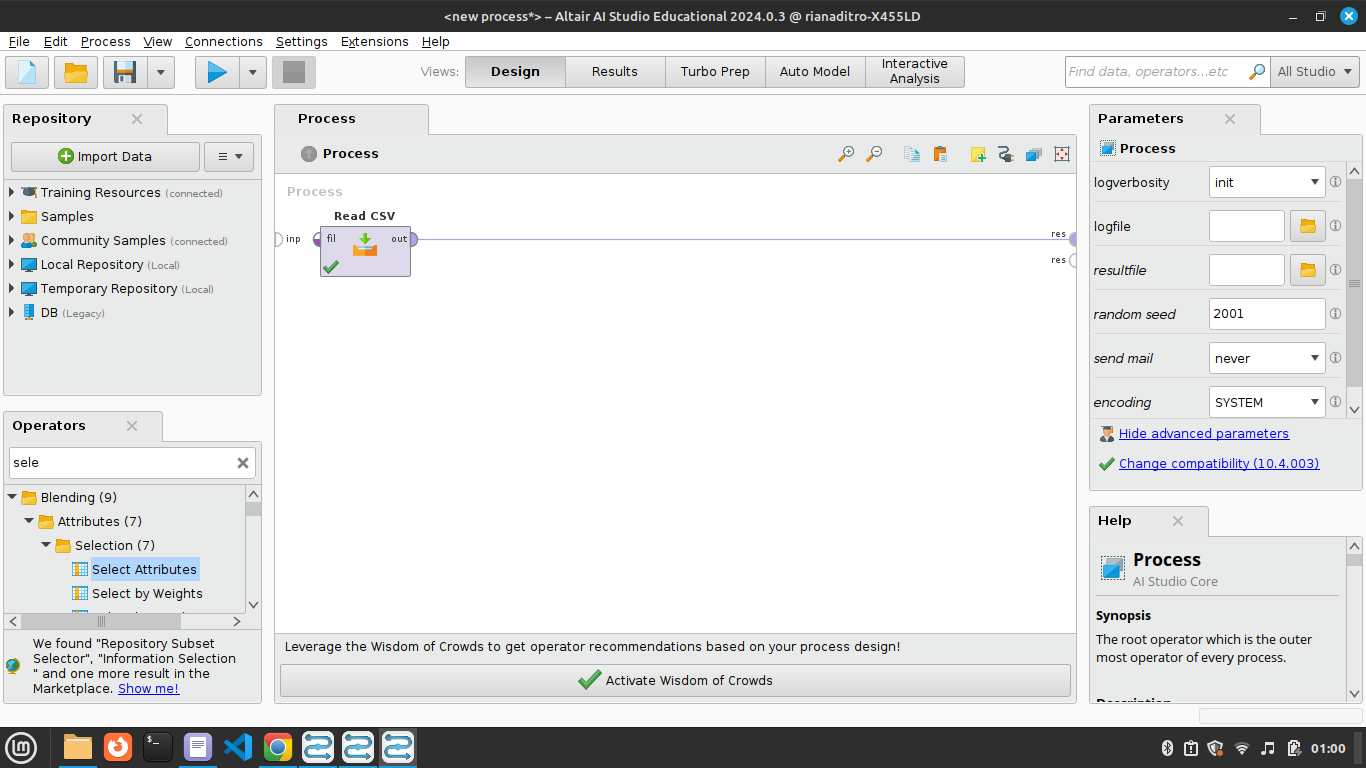
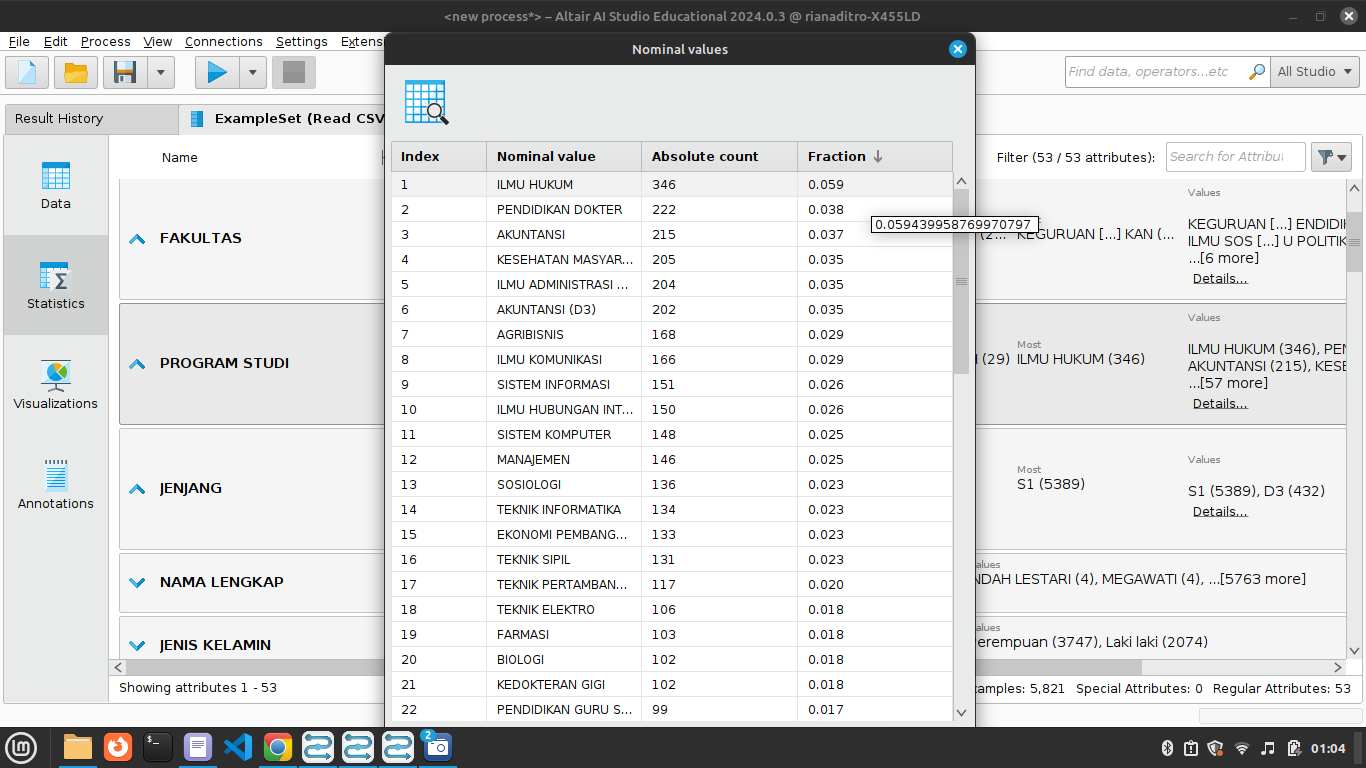
Klasifikasi Masa Tunggu Menggunakan Algoritma Naive Bayes

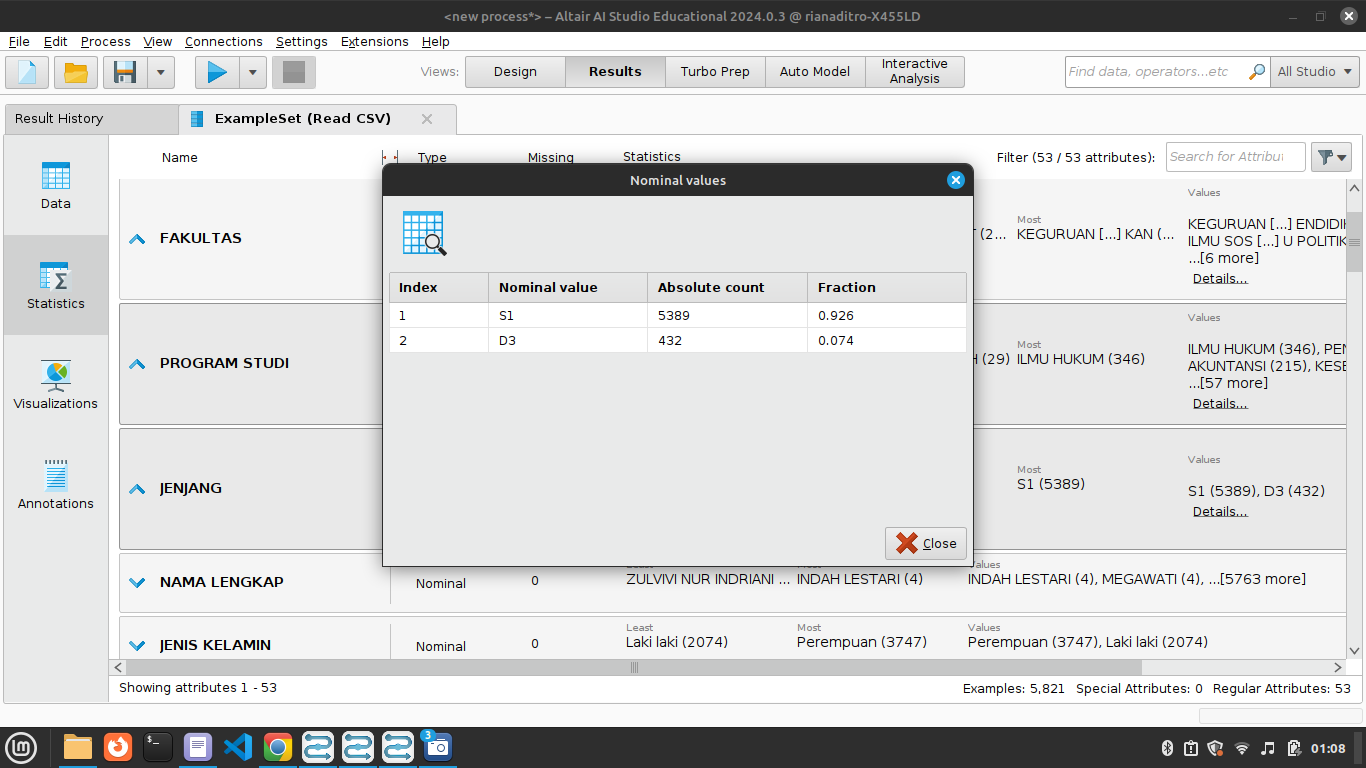
# Exploratory Data Analysis

Tahapan pengenalan data

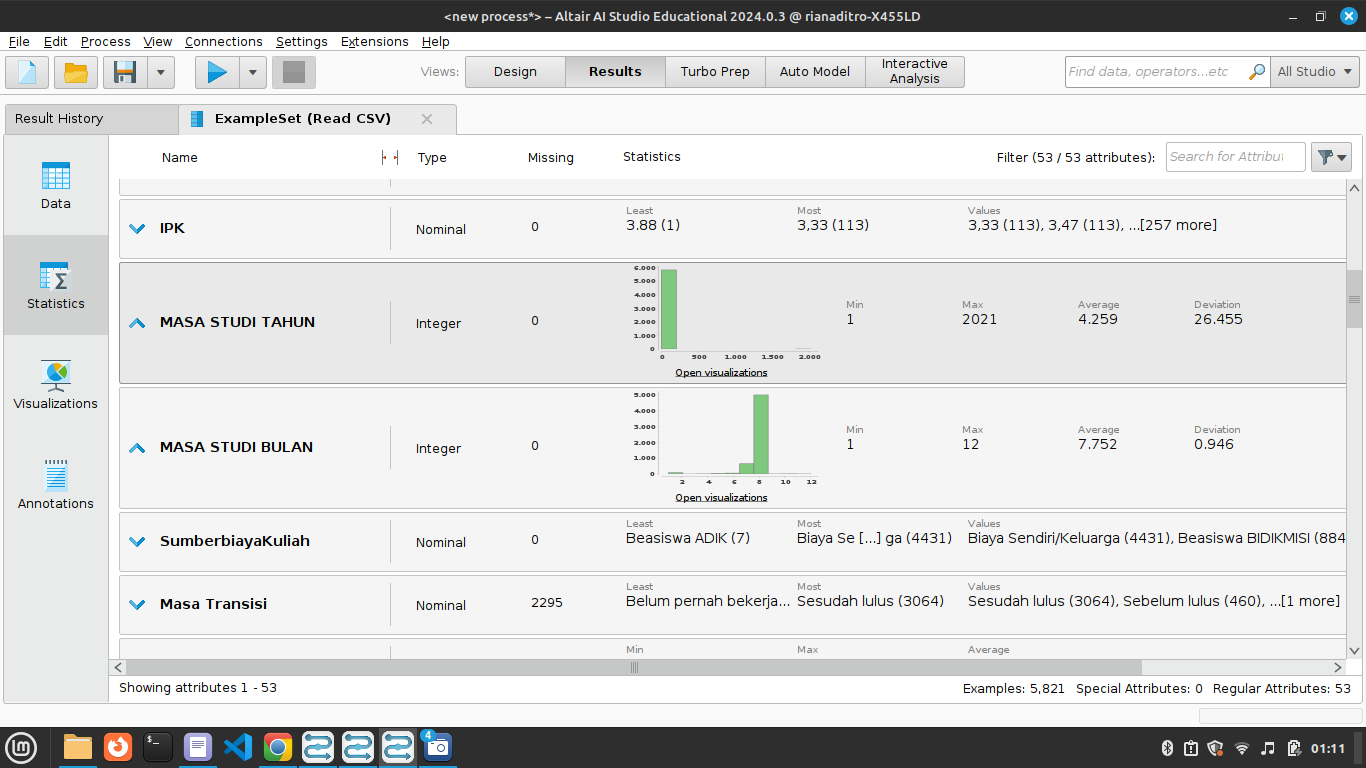




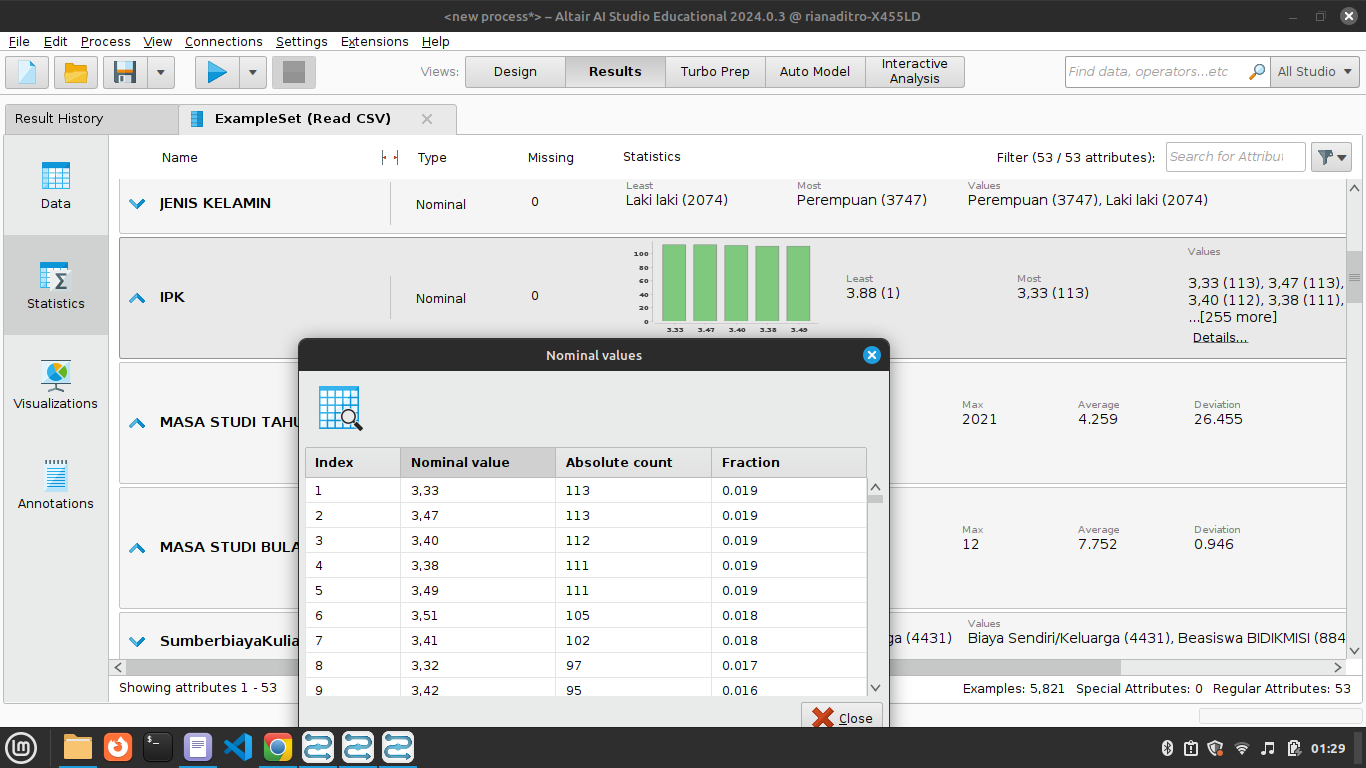
1. Atribut PROGRAM STUDI terlalu beragam, hanya ada 1 yang bernilai diatas 5%, maka atribut ini akan dibuang dan dipilih FAKULTAS sebagai gantinya.



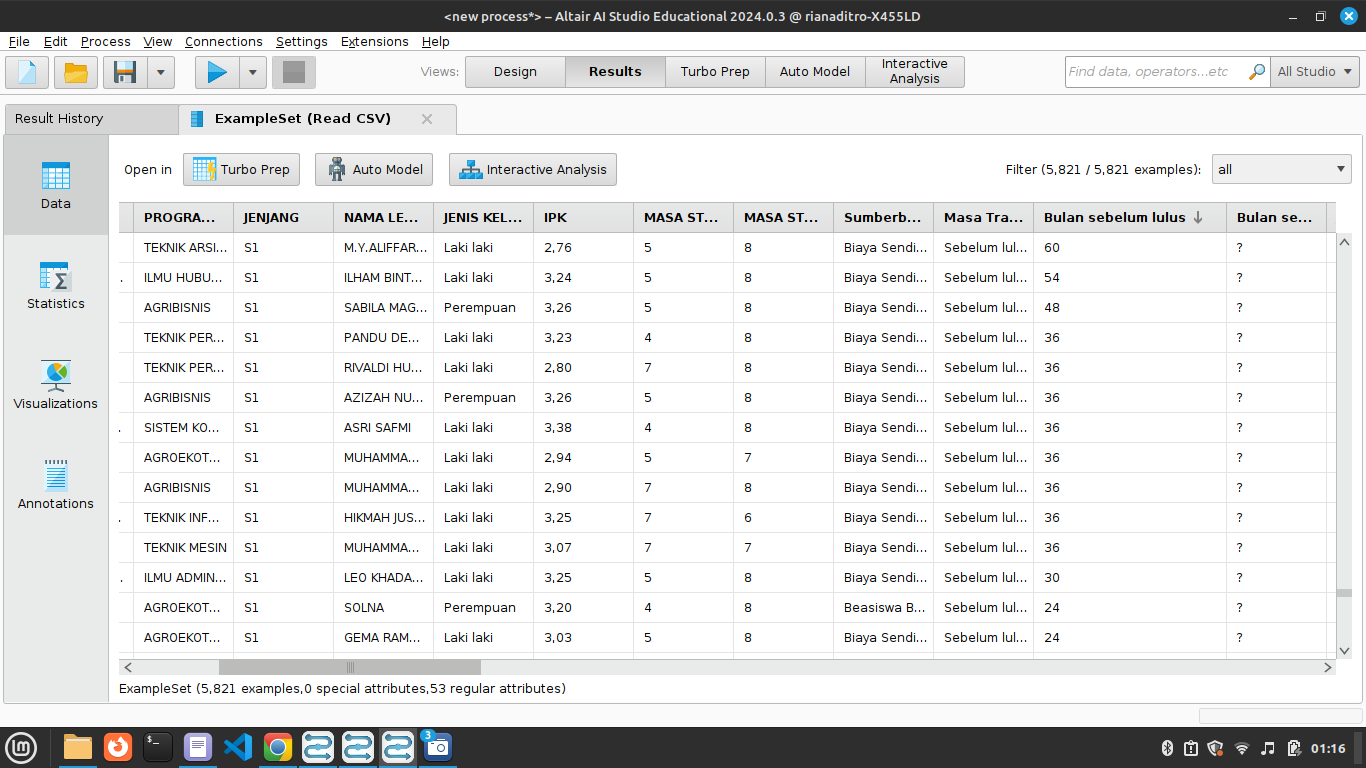
1. Nilai D3 terlalu sedikit, maka pada penelitian ini akan difokuskan ke jenjang S1 saja.



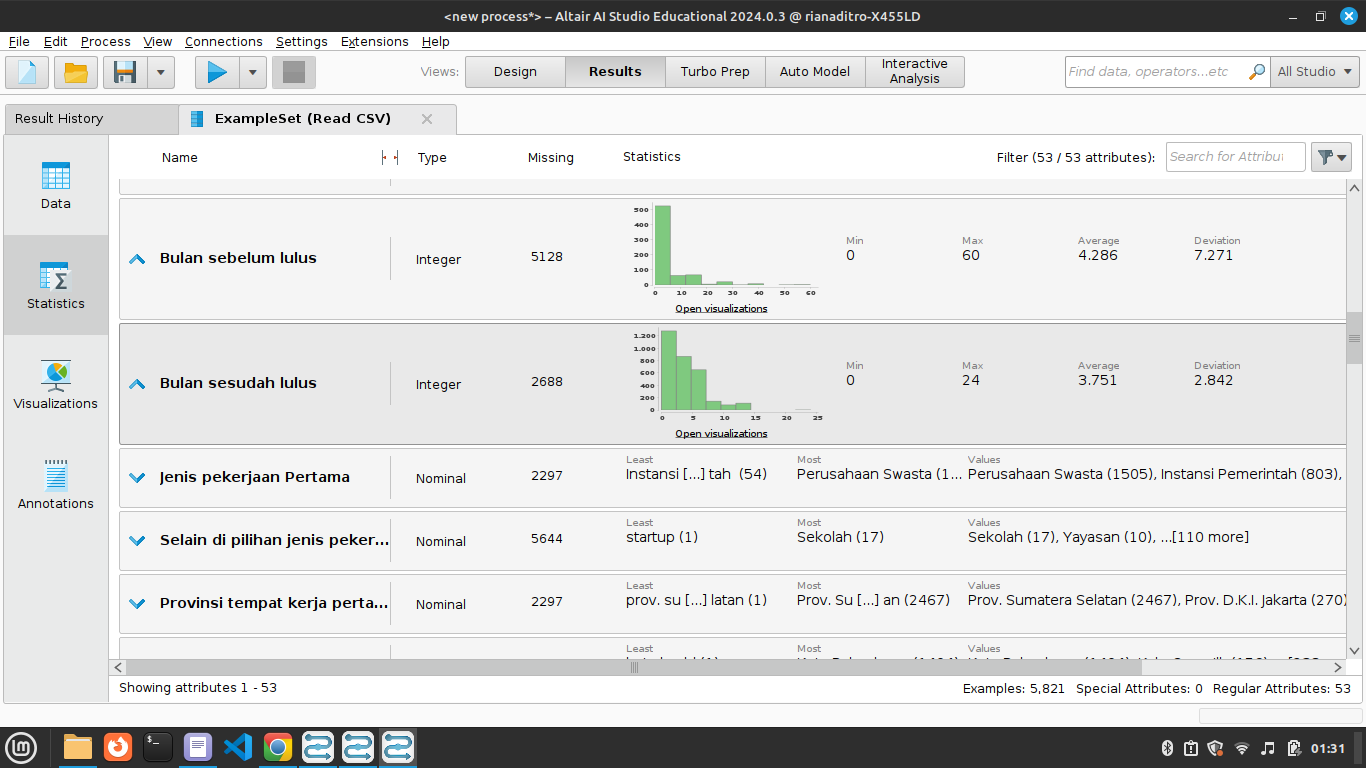
1. MASA STUDI BULAN condong ke satu nilai, tidak perlu diambil. Sebagai gantinya akan diambil MASA STUDI TAHUN. Namun, nilai terendah adalah 1 tahun, untuk jenjang S1 setidaknya harus 2 tahun.



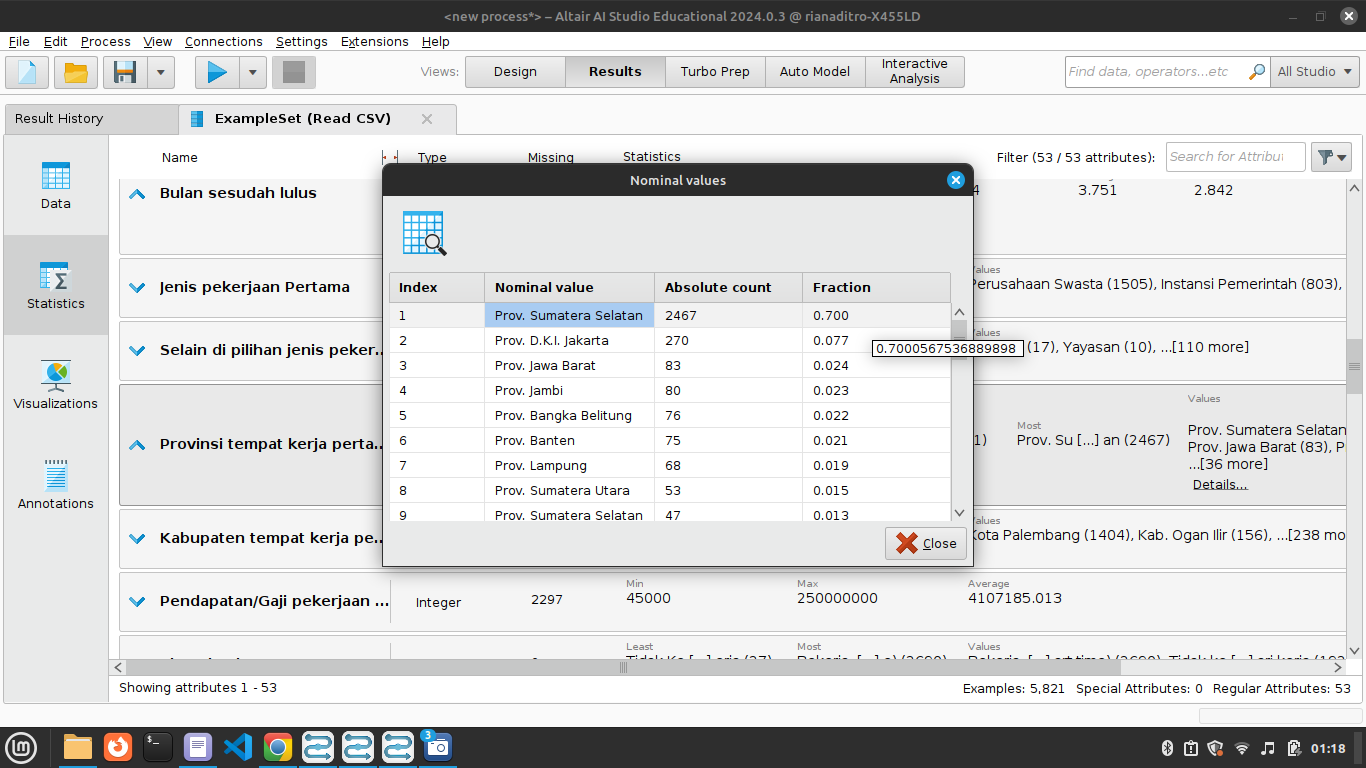
1. IPK bertipe nominal, harus diubah.



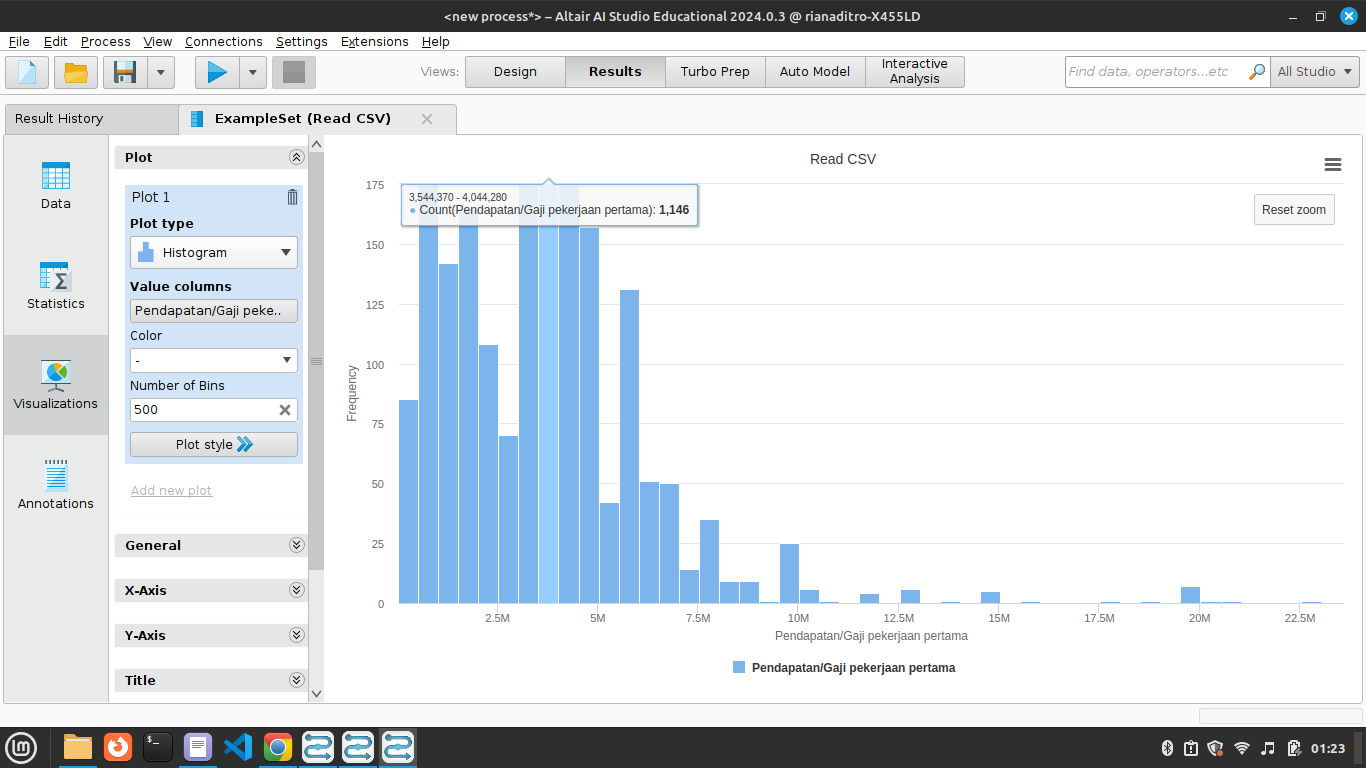
1. Ditemukan data yang tidak bisa dipercaya pada Bulan sebelum lulus, maka dari itu, penelitian ini akan mengambil data dari Bulan setelah lulus.



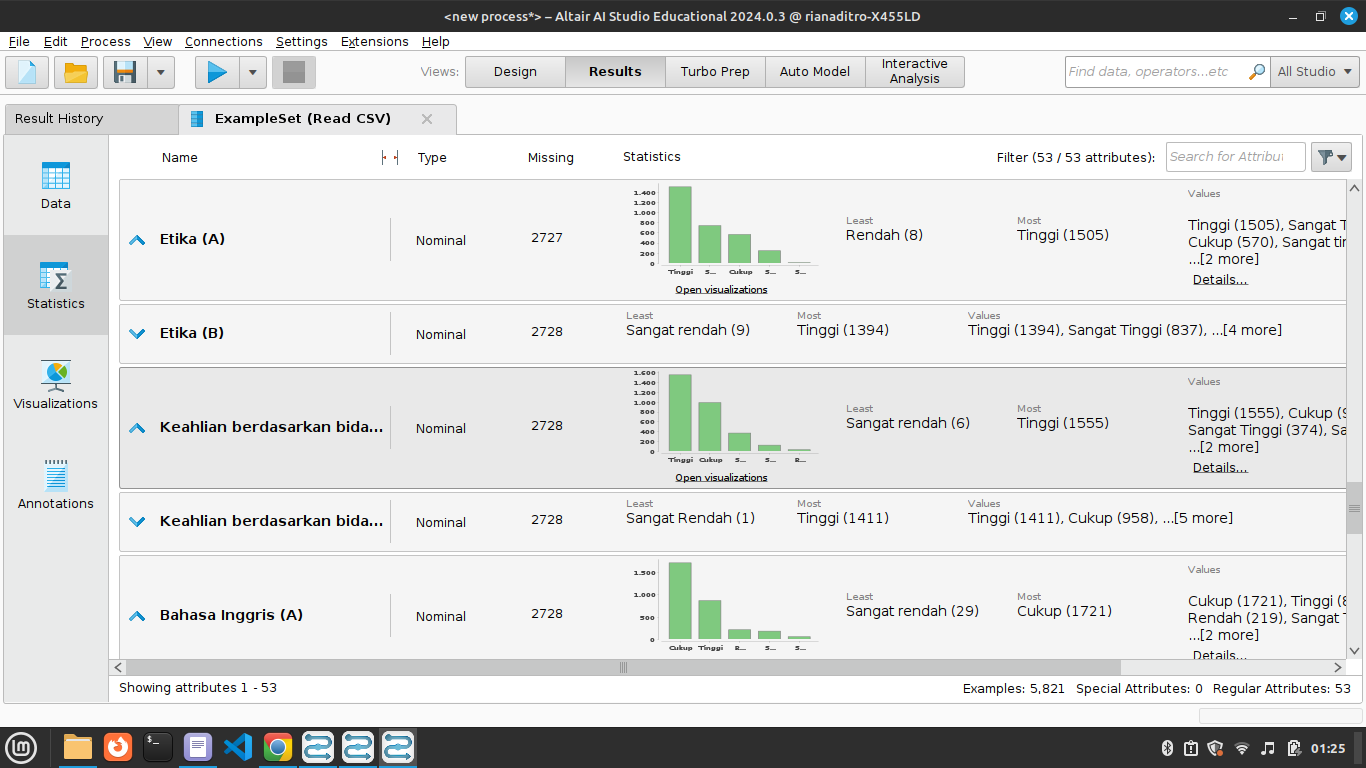
1. Bulan setelah lulus memiliki missing value, perlu dilakukan pemilahan.



1. Provinsi tempat kerja pertama akan dibuat sebagai atribut baru yaitu LOKASI KERJA dengan nilai ‘dalam daerah’ untuk provinsi Sumatera Selatan dan ‘luar daerah’ untuk lainnya.



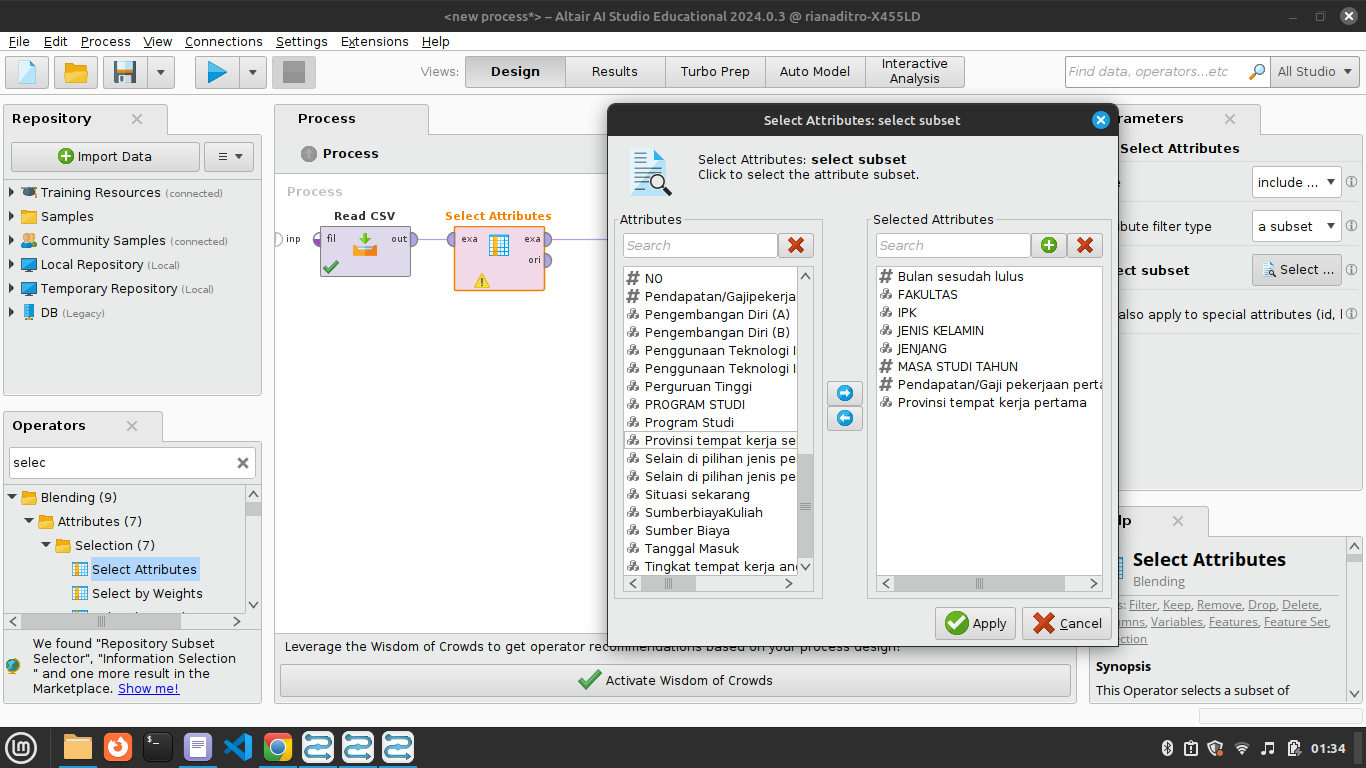
1. Pendapatan/Gaji pekerjaan pertama akan dibuat sebagai atribut baru yaitu EKSPEKTASI GAJI dengan nilai kategori ‘dibawah 4 juta’ dan ‘diatas 4 juta’



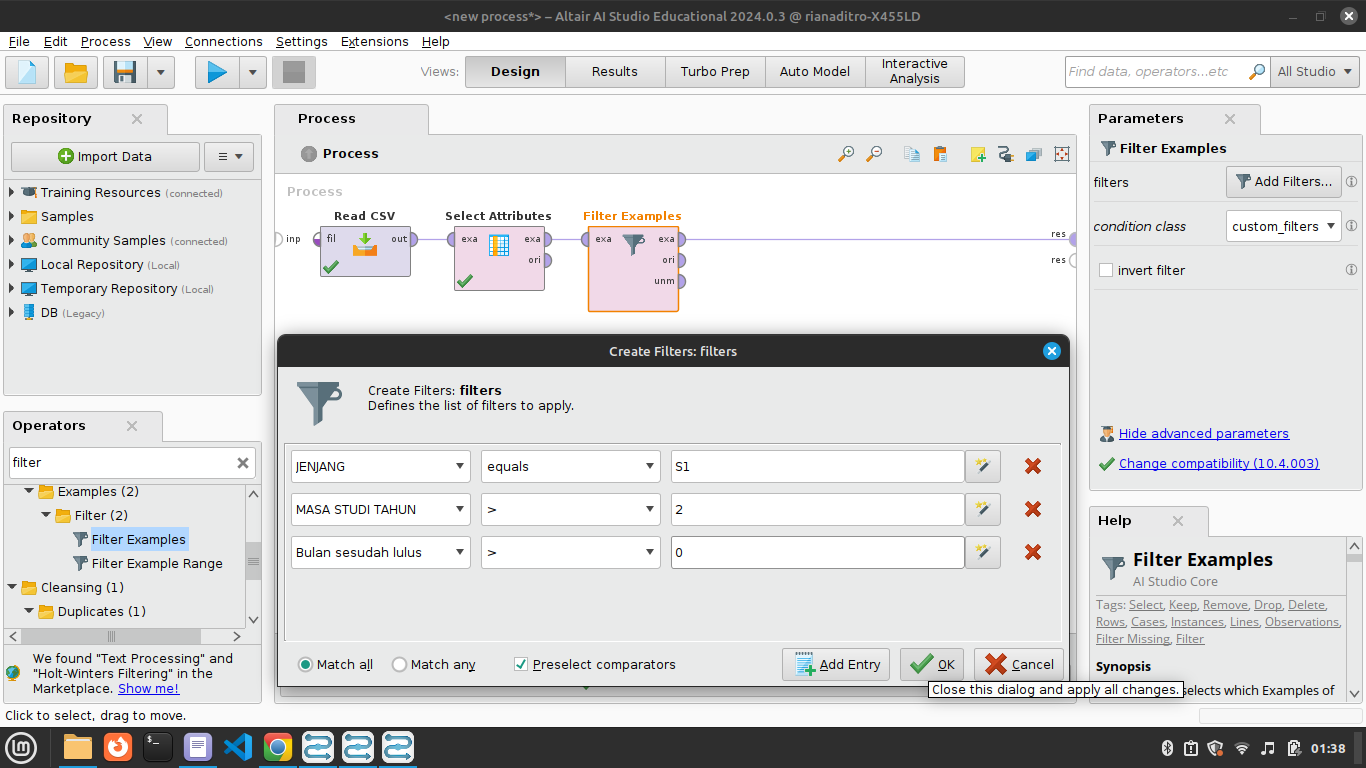
1. Atribut Etika, Bahasa Inggris, dan lain-lain memiliki distribusi data yang sama. Dinilai tidak memberikan bobot tambahan dalam pelatihan model, maka tidak akan diambil.

# 

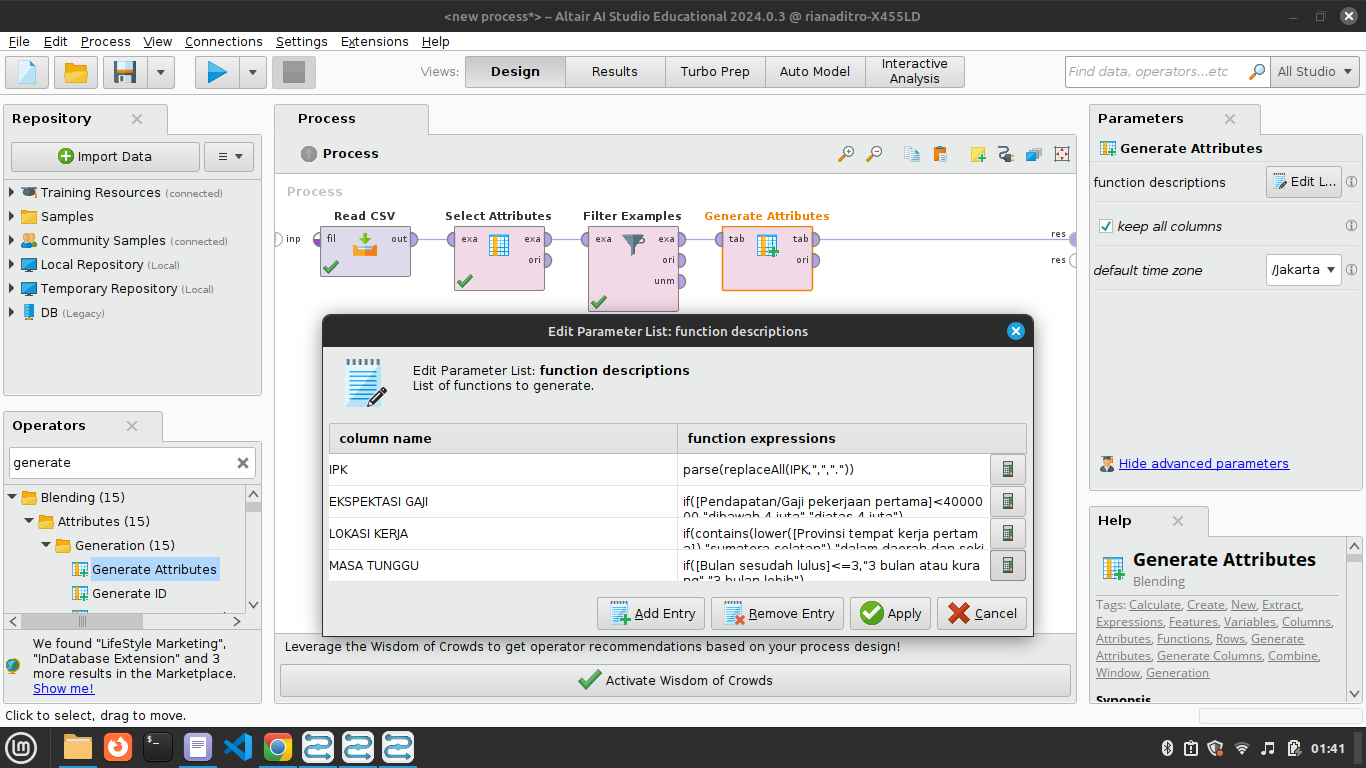
# Data Processing



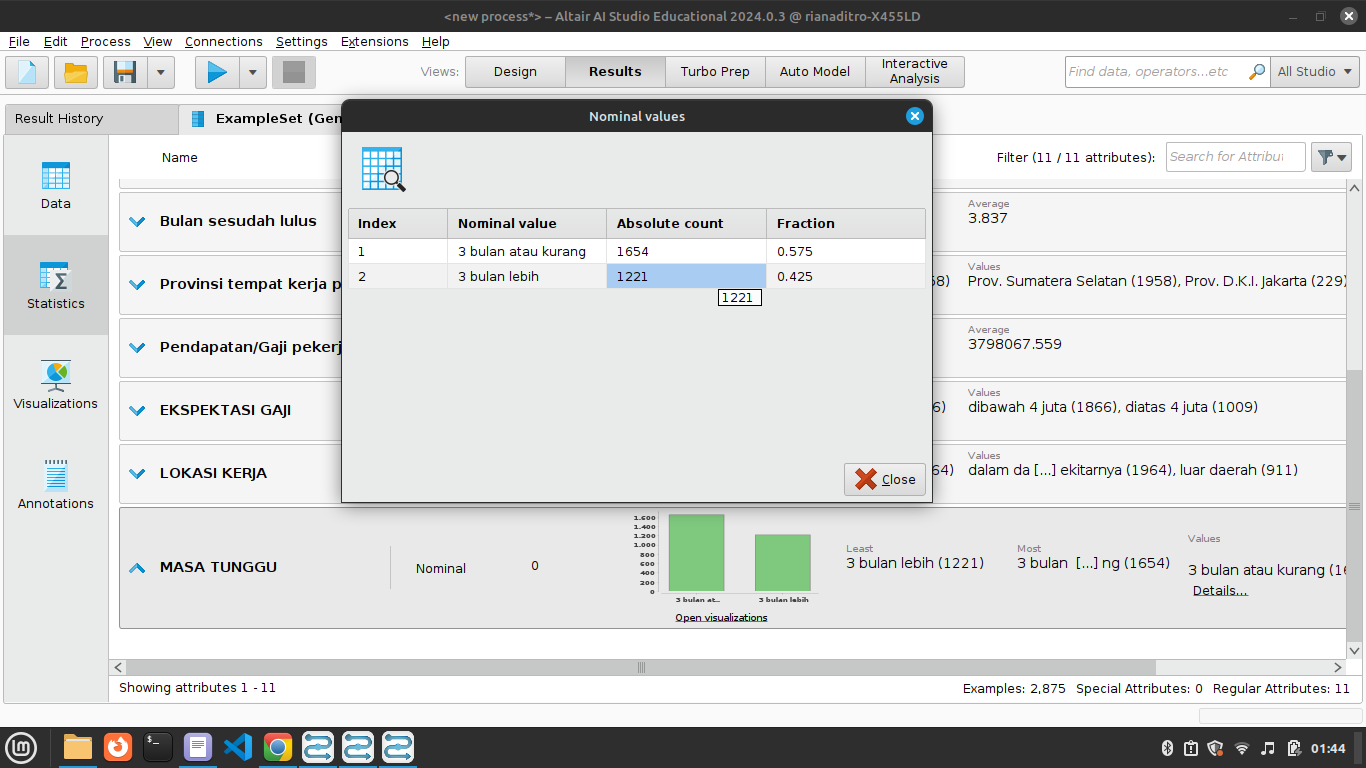
1. Select Atribute: untuk pemilahan data yang digunakan



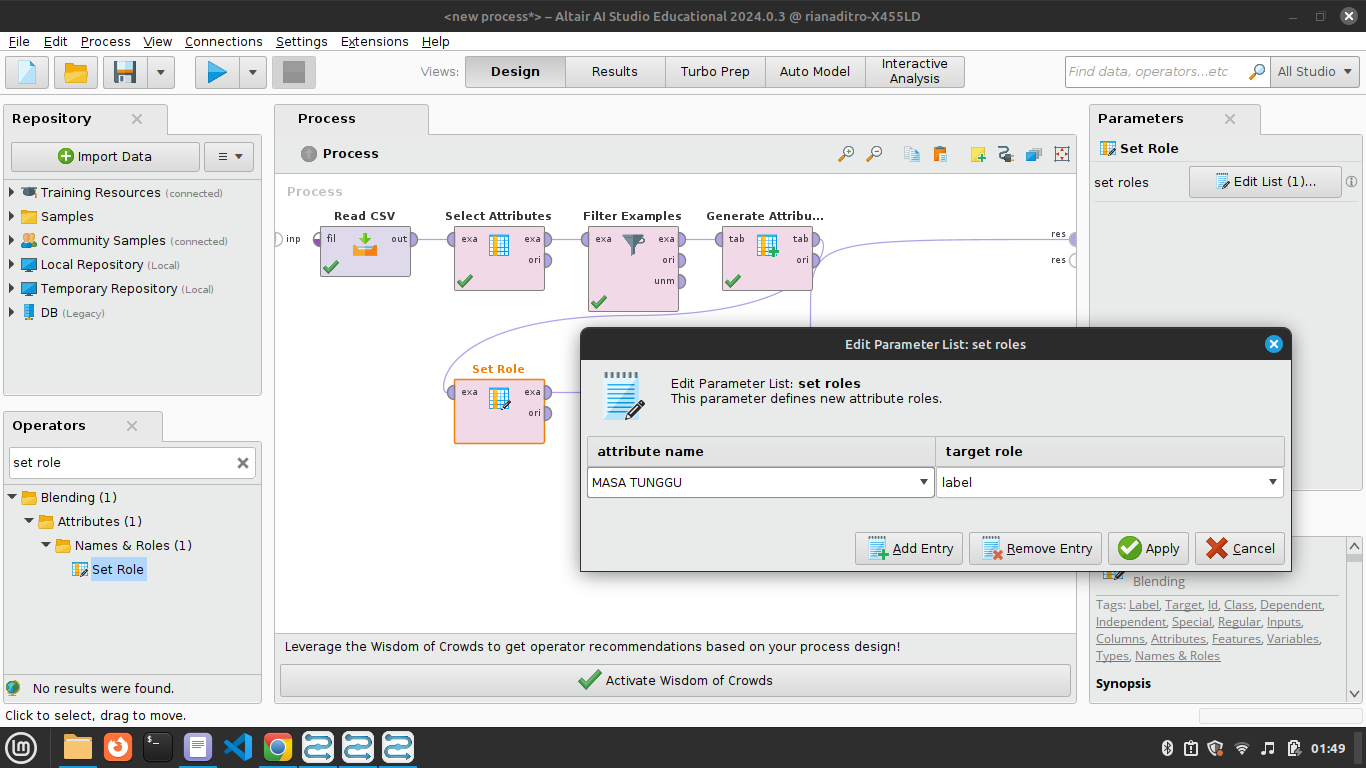
1. Filter Example: menghilangkan missing value dan data tidak valid.



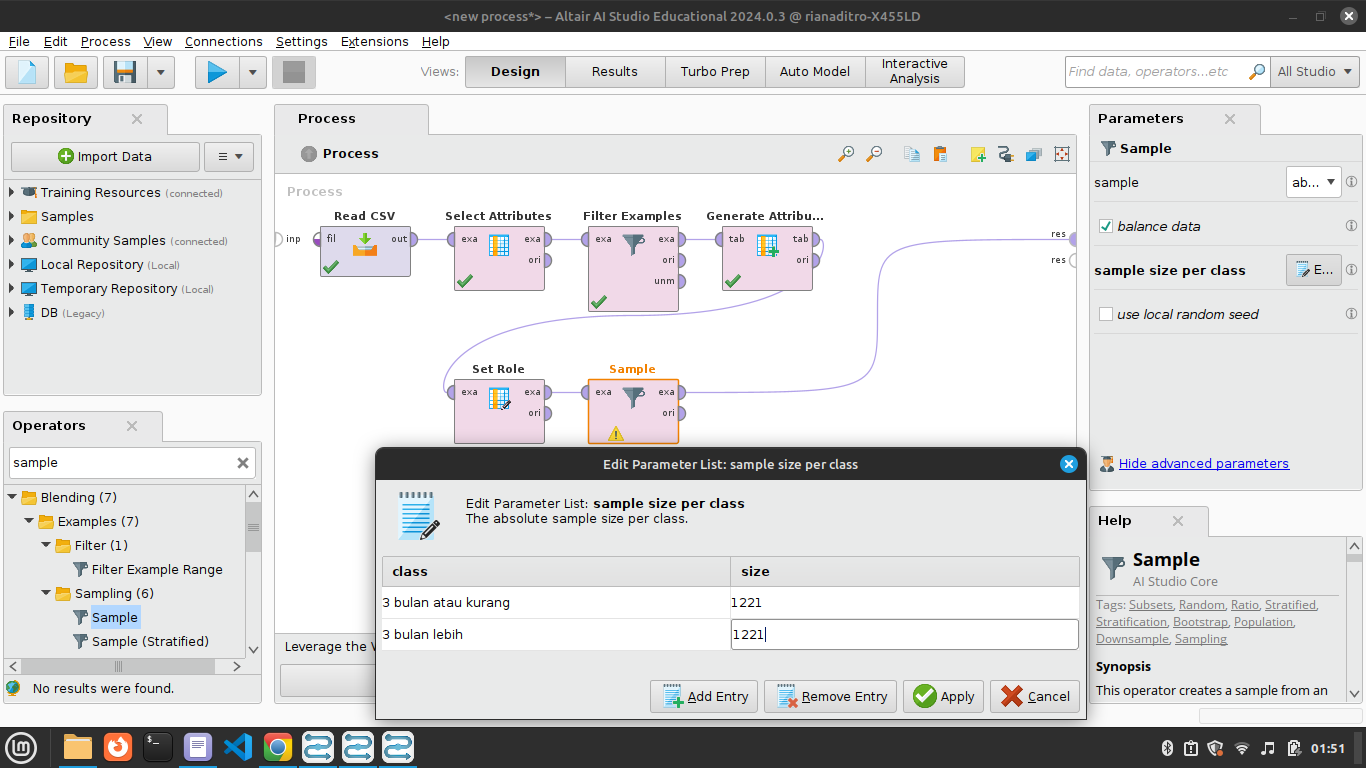
1. Generate Atribute: mengubah format IPK, dan membuat atribut baru.



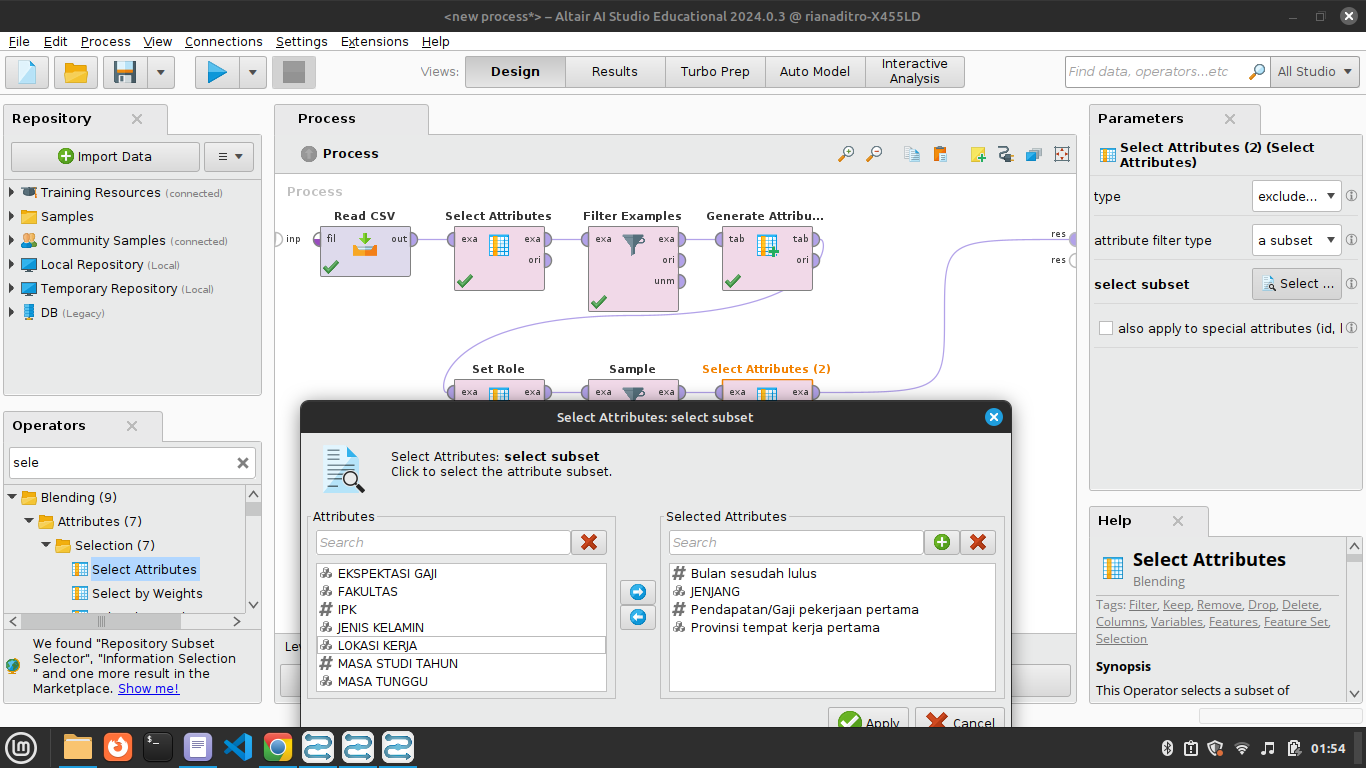
1. Cek atribut MASA TUNGGU. Atribut ini akan kita gunakan sebagai label. Untuk hasil yang optimal kita imbangkan proporsi datanya dengan menyamakan jumlah data berdasarkan label.



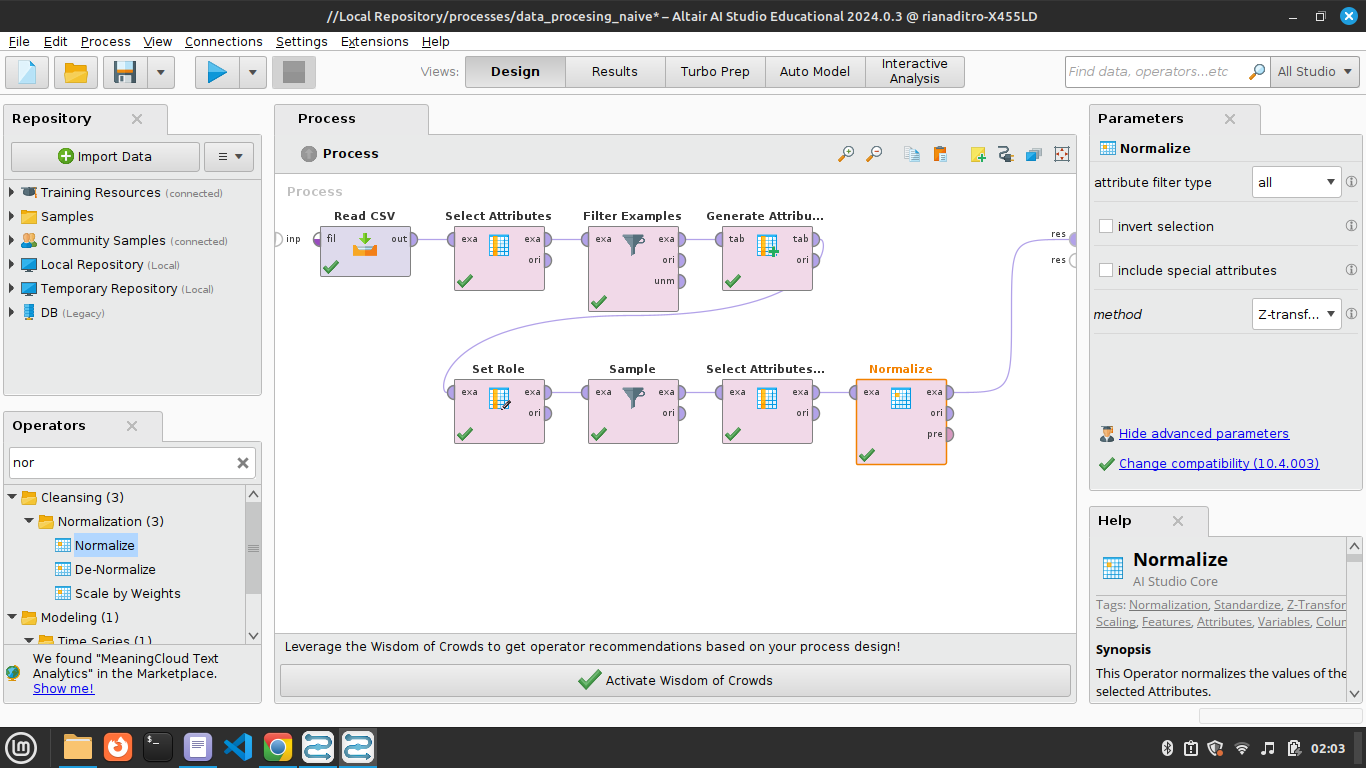
1. Set Role: mengubah MASA TUNGGU menjadi label.



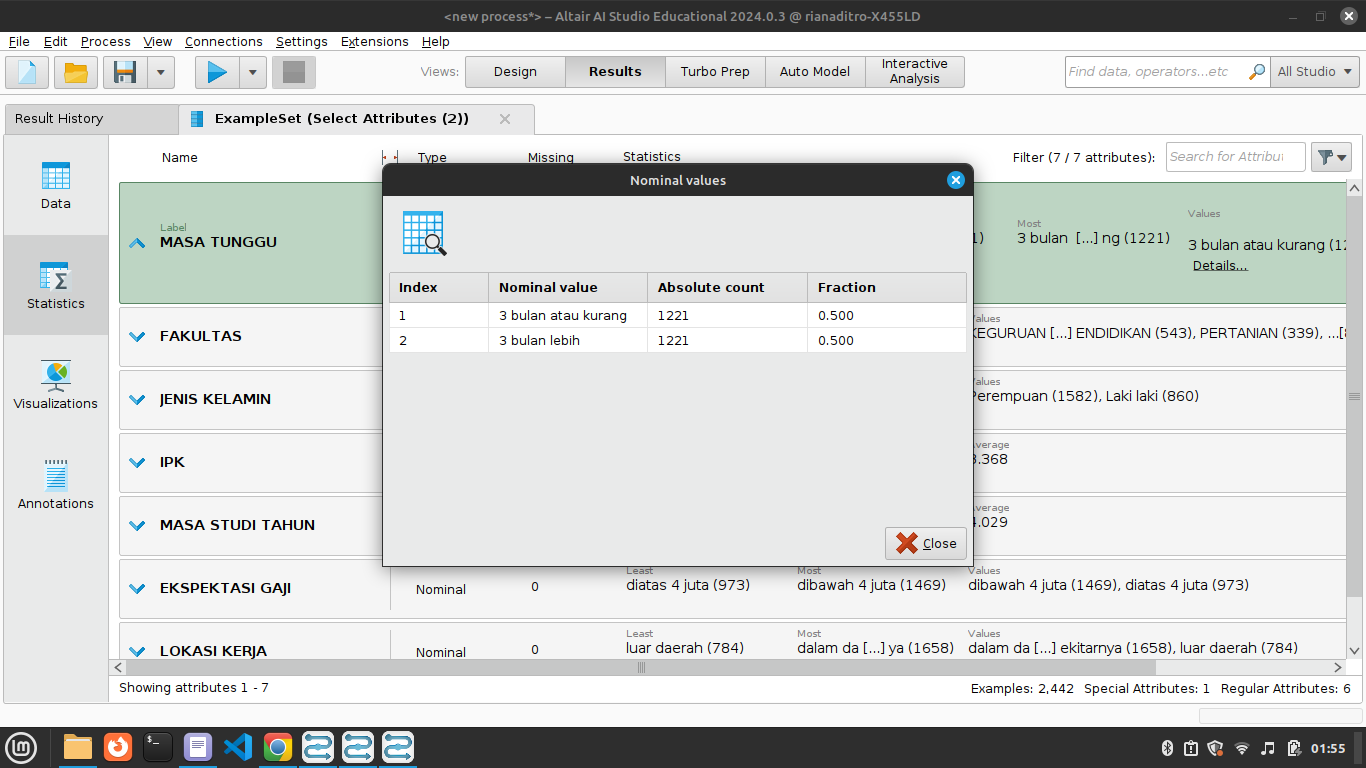
1. Sample: membagi data berdasarkan class pada label.



1. Select Atribute: hilangkan atribut yang sudah tidak diperlukan.



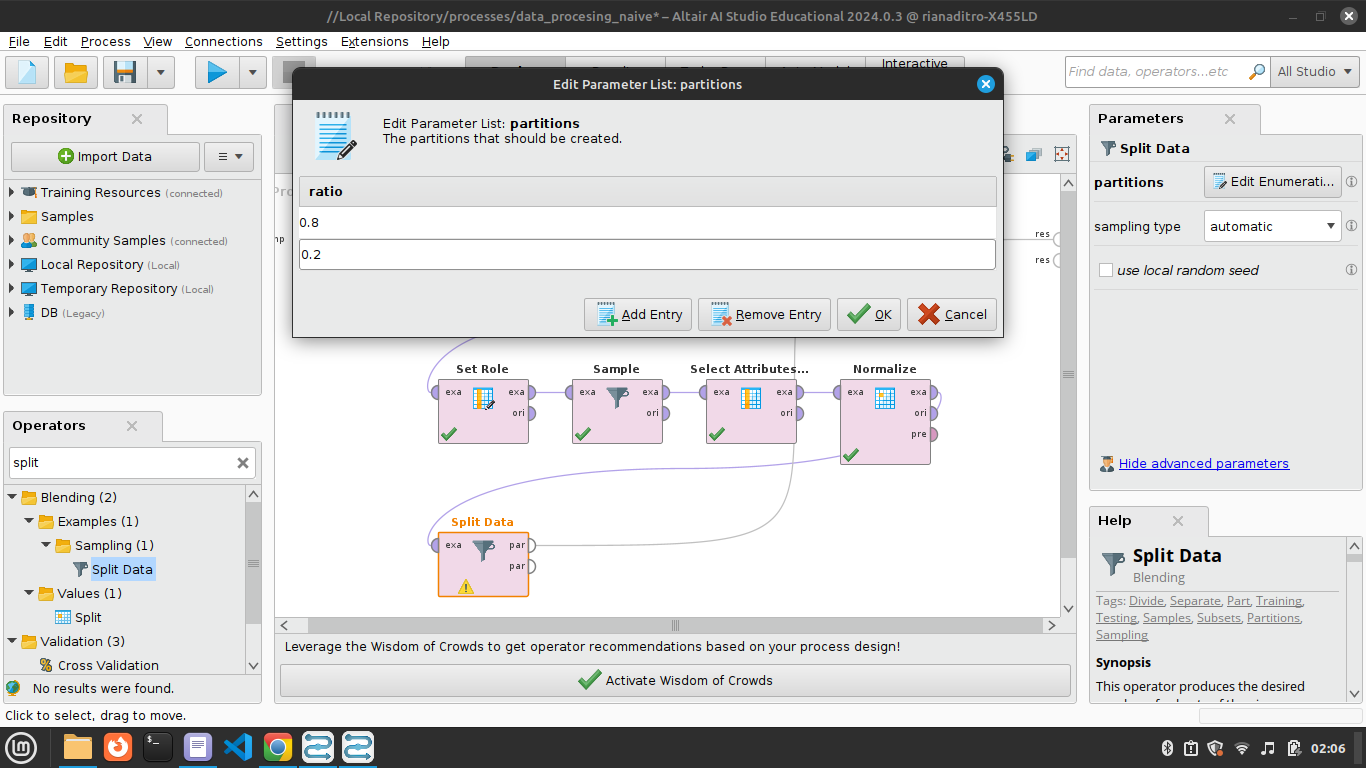
1. Normalization: normalisasi nilai di rentang 0 dan 1.



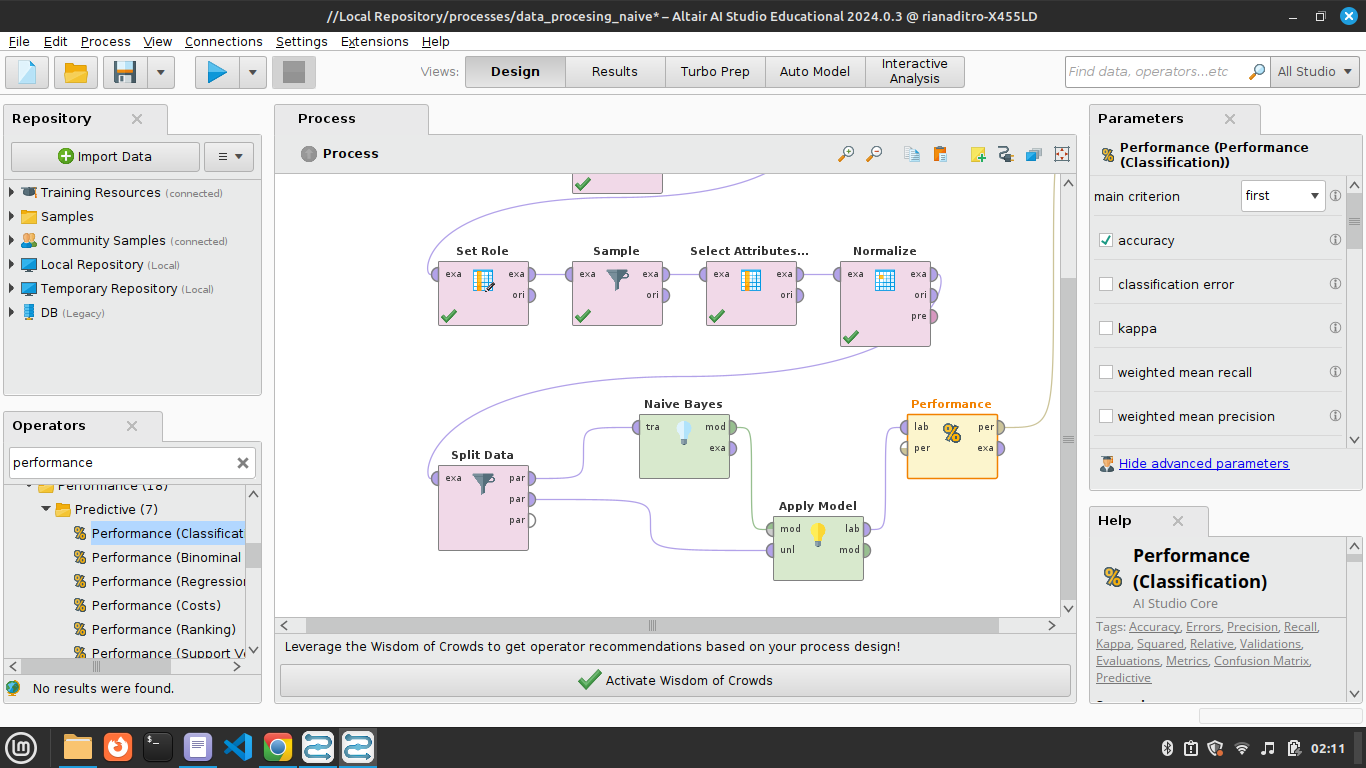
1. Total 2442 data siap digunakan. Tambahkan operator ‘Write CSV’ untuk mengekspor data ini ke file CSV jika anda ingin memisahkan proces antara pengolahan data dan pelatihan model.

# 

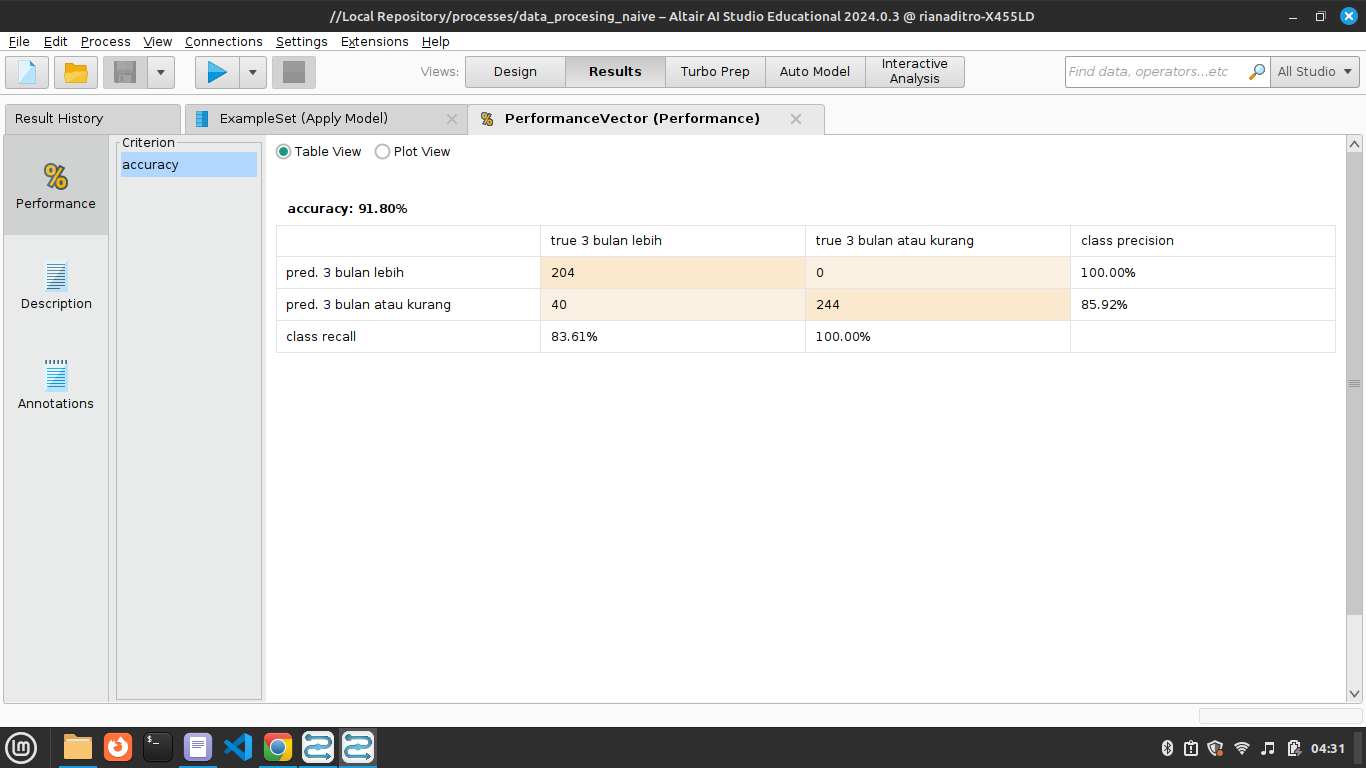
# Data Training & Testing



1. Split Data: 80:20 untuk data latih:data uji, atur sampling type menjadi ‘stratified’ agar pembagian merata antara label (opsional, karena data sudah merata).



1. Tambahkan operator Naive Bayes, Apply Model dan Performance(classification)
2. Hubungkan sesuai gambar.



1. Didapatkan akurasi sebesar 91.80%