

DATA SCIENCE and BUSINESS INTELLIGENT

Author: Egi Safitri

Meeting 6





- RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source).
- RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi.
- RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

Kelebihan Rapidminer

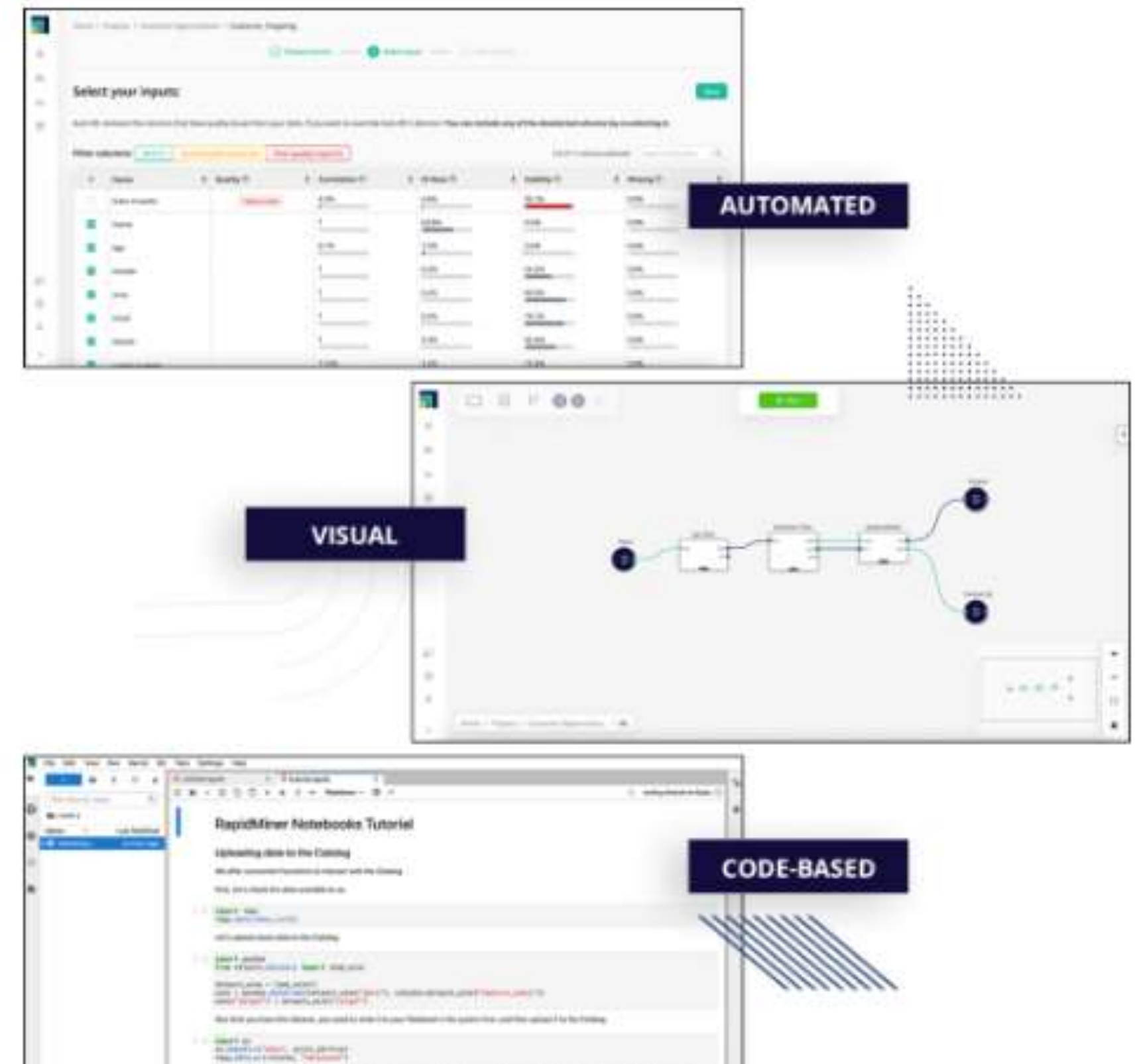
- Rapidminer cocok untuk pengguna yang tidak memiliki latar belakang non-programming tools.
- RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi.
- RapidMiner merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri.
- RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

Satu Platform Untuk Semua

Dengan menyediakan antarmuka yang intuitif untuk semua orang

Otomatisasi penuh untuk pemula, lingkungan JupyterLab terintegrasi, dan desain drag-and-drop visual.

- ✓ Data Science
- ✓ Domain Expert
- ✓ IT
- ✓ Leadership
- ✓ Non-programmer



Data Science Lifecycle

- RapidMiner membuat data science menarik dan mudah didekati dengan metodologi adopsi terstruktur untuk semua orang.
- Platform ini mempercepat seluruh life cycle - mulai dari data preparation hingga membuat model.
 - Terintegrasi dengan sistem, data, dan bahasa yang ada, serta merampingkan kolaborasi lintas tim, disiplin ilmu, dan geografi.



Sejarah Rapidminer

- RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund.
- RapidMiner didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3.

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



Rapidminer: Kakas Pemrograman Visual

- RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis.
- GUI ini akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data.
- File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis

Penggunaan Rapidminer

- Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
- Proses penemuan pengetahuan dimodelkan sebagai operator trees
- Representasi XML internal untuk memastikan format standar pertukaran data.
- Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
- Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain

Fitur-fitur Rapidminer

- Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
- Bentuk grafis yang canggih, seperti diagram histogram, tree chart dan 3D Scatter plots.
- Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- Menyediakan prosedur data mining dan machine learning termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi.
- Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.

Proses Instalasi

- Minimum spesifikasi:
 - RAM minimal 2 GB (disarankan 4 GB)
 - Laptop dengan 32/64-bit processor
 - Laptop dengan Operating System Windows 7, 8, 10, MacOS X atau Linux
 - Java Runtime Environment versi 6
- RapidMiner dapat diunduh pada situs resminya, yaitu www.rapid-i.com atau <https://rapidminer.com/> atau <https://rapidminer.com/platform/educational/> (versi education)
- Jalankan proses instalasi seperti biasa hingga berhasil.

Pengenalan Antarmuka

- RapidMiner menyediakan tampilan yang user friendly untuk memudahkan penggunaanya ketika menjalankan aplikasi.
- Tampilan pada RapidMiner dikenal dengan istilah Perspective.
- Pada RapidMiner terdapat 3 Perspective, yaitu; Welcome Perspective, Design Perspective dan Result Perspective.

Welcome/Home Perspective

- Ketika membuka aplikasi Anda akan disambut dengan tampilan yang disebut dengan Welcome Perspective.
- Pada bagian toolbar, terdapat toolbar Perspectives yang terdiri dari ikon-ikon untuk menampilkan perspective dari RapidMiner.
- Toolbar ini dapat dikonfigurasi sesuai dengan kebutuhan Anda. Sedangkan Views menunjukkan pandangan (view) yang sedang Anda tampilkan

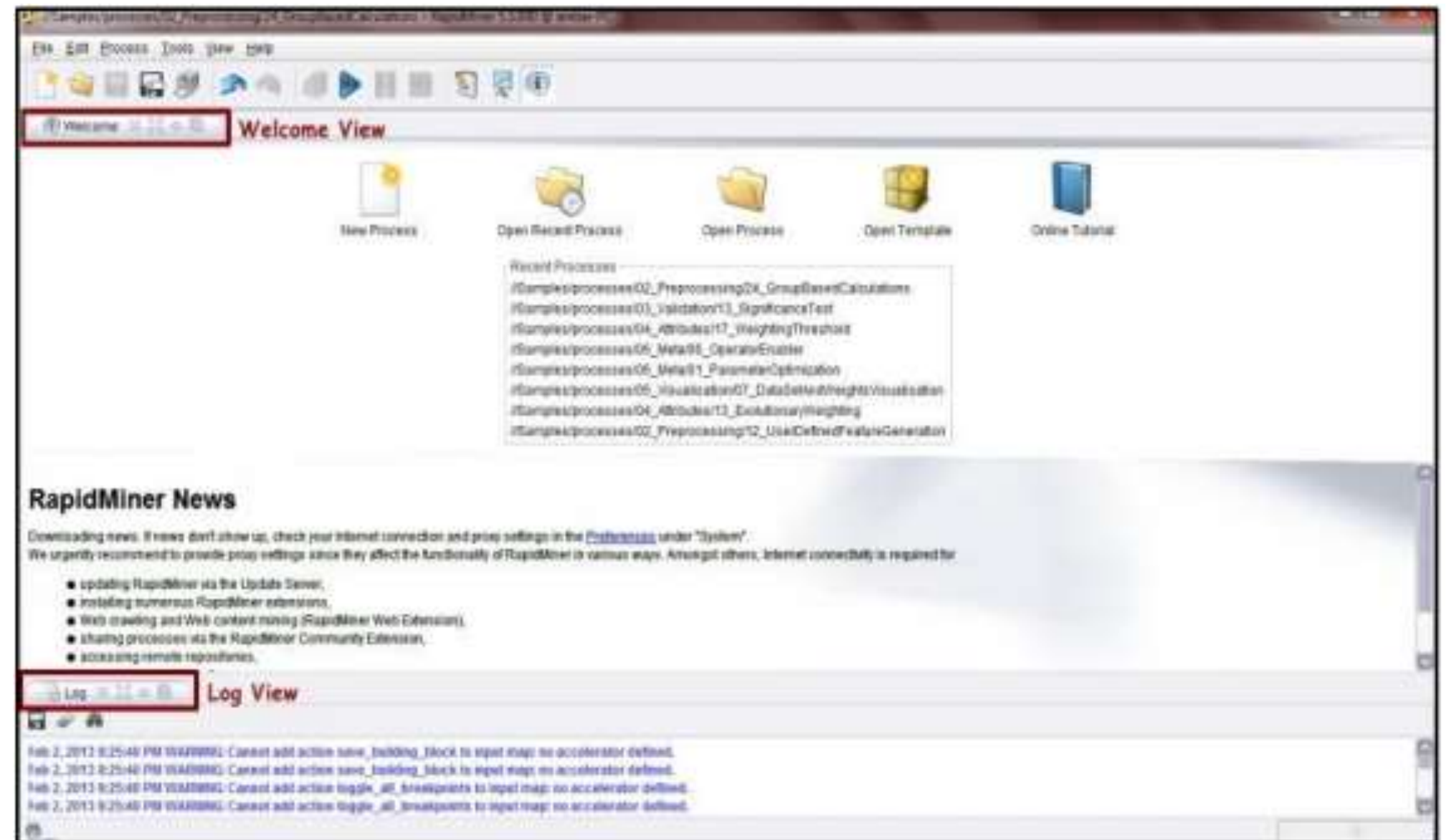


Welcome/Home Perspective

- Jika komputer Anda terhubung dengan internet, maka pada bagian bawah Welcome Perspective akan menampilkan berita terbaru mengenai RapidMiner. Bagian ini dinamakan News.
- Pada bagian tengah halaman terlihat daftar Last Processes (Recent Processes), bagian ini menampilkan daftar proses analisis yang baru saja dilakukan.
 - Hal ini akan memudahkan Anda jika ingin melanjutkan proses sebelumnya yang sudah ditutup, dengan mengklik dua kali salah satu proses yang ada pada daftar tersebut.
- Bagian Actions menunjukkan daftar aksi yang dapat Anda lakukan setelah membuka RapidMiner.

Welcome/Home Perspective

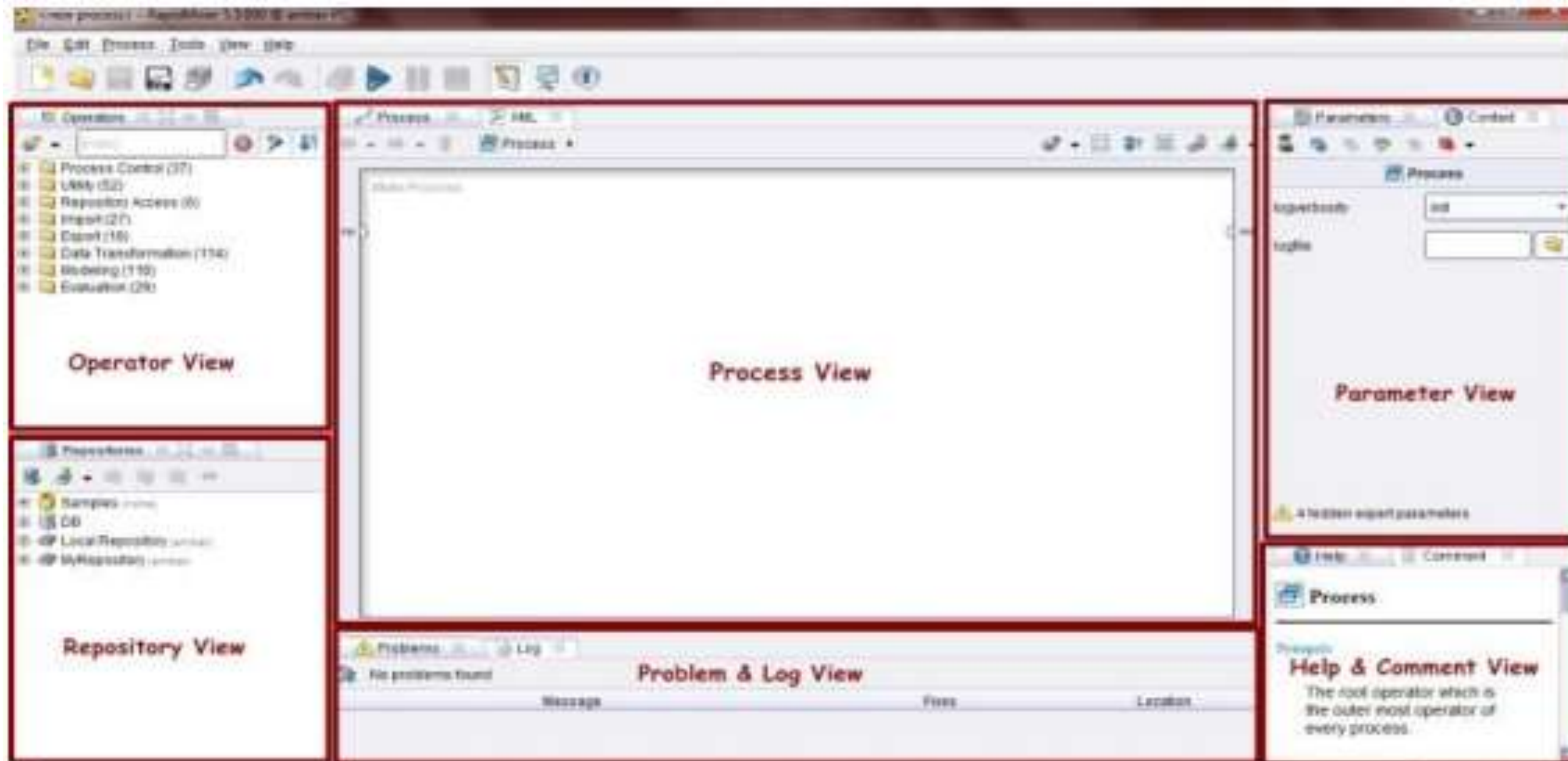
- RapidMiner dapat menampilkan beberapa view pada saat bersamaan.
- Pada tampilan Welcome Perspective terdapat Welcome view dan Log View.
 - Ukuran dari setiap view tersebut dapat diubah sesuai dengan kebutuhan Anda dengan Mengklik dan menarik garis batas diantara keduanya ke atas atau ke bawah.



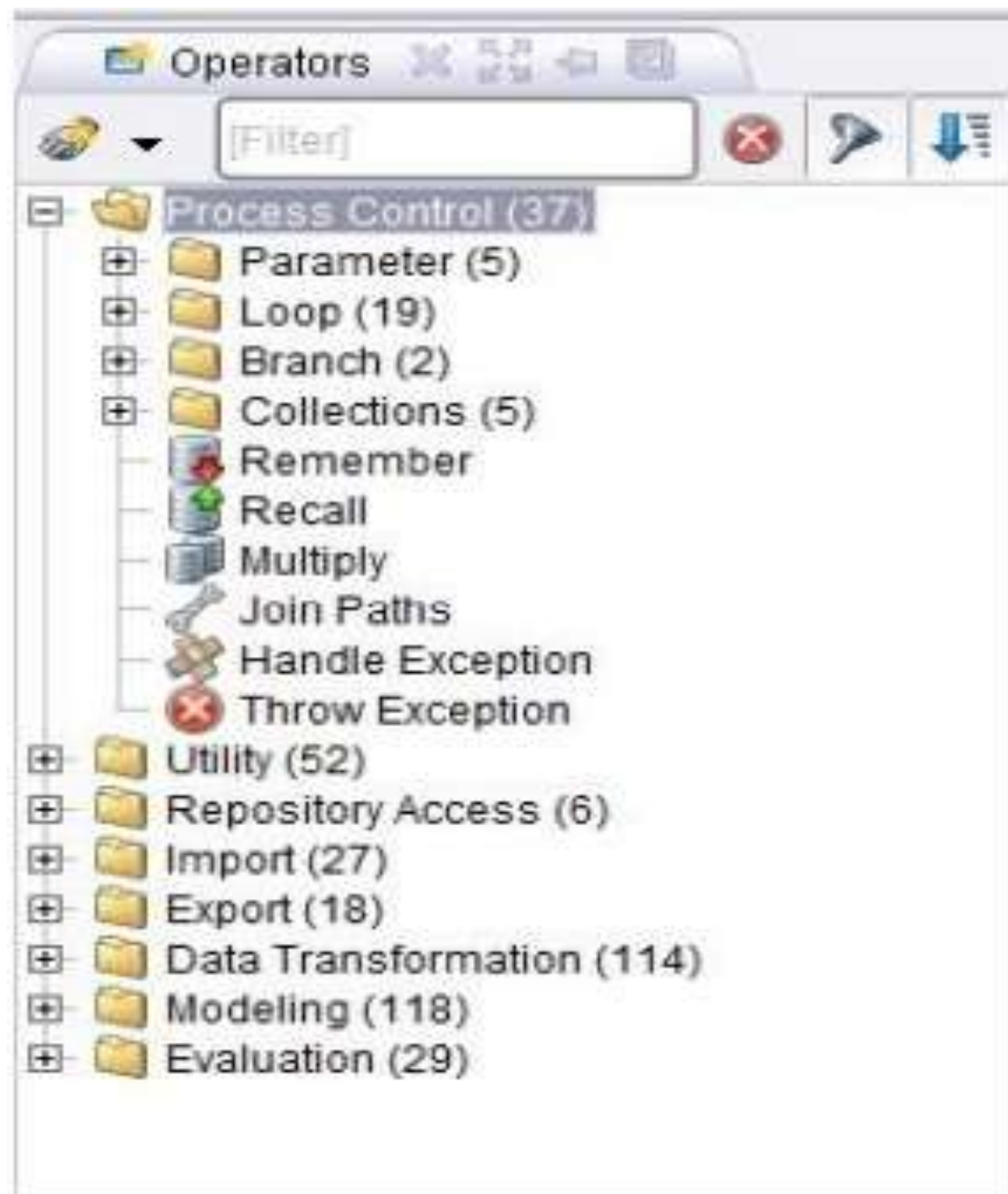
Design Perspective

- Design Perspective merupakan lingkungan kerja RapidMiner.
- Dimana Design Perspective ini merupakan perspective utama dari RapidMiner yang digunakan sebagai area kerja untuk membuat dan mengelola proses analisis.
- Perspective ini memiliki beberapa view dengan fungsinya masing-masing yang dapat mendukung Anda dalam melakukan proses analisis data mining.
 - Anda bisa mengganti perspective dengan mengklik salah satu ikon dari toolbar perspective.
 - Selain dengan cara tersebut, Anda juga bisa mengganti perspective dengan mengklik menu view, kemudian pilih perspective, lalu pilih perspective yang ingin Anda

Tampilan Design Perspective



Tampilan Operator View



- Operator View merupakan view yang paling penting pada perspective ini. Semua operator atau langkah kerja dari RapidMiner disajikan dalam bentuk kelompok hierarki di Operator View ini.
- Hal ini akan memudahkan Anda dalam mencari dan menggunakan operator dalam proses analisis yang sesuai dengan kebutuhan

Operator View (1)

- Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut
 - Process Control : Operator ini terdiri dari operator perulangan dan percabangan yang dapat mengatur aliran proses.
 - Utility : Operator bantuan, seperti operator macros, loggin, subproses, dan lain-lain.
 - Repository Access : Kelompok ini terdiri dari operator-operator yang dapat digunakan untuk membaca atau menulis akses pada repository.
 - Import : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk membaca data dan objek dari format tertentu seperti file, database, dan lain-lain.
 - Export : Kelompok ini terdiri dari banyak operator yang dapat digunakan untuk menulis data dan objek menjadi format tertentu.

Operator View (2)

- Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut.
 - Data Transformation : kelompok ini terdiri dari semua operator yang berguna untuk transformasi data dan metadata.
 - Modeling : kelompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru.
 - Evaluation : kelompok ini berisi operator yang dapat digunakan untuk menghitung kualitas pemodelan dan untuk data baru.

Repository View

- Repository View merupakan komponen utama dalam Design Perspective selain Operator View. View ini dapat Anda gunakan untuk mengelola dan menata proses Analisis Anda menjadi proyek.
- Pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan metadata.

Process View

- Process View menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut. Anda dapat menambahkan langkah baru dengan beberapa cara. hubungan diantara langkah-langkah ini dapat dibuat dan dilepas kembali.
- Pada dasarnya bekerja dengan RapidMiner ialah mendefinisikan proses analisis, yaitu dengan menunjukkan serangkaian langkah kerja tertentu.
- Dalam RapidMiner, komponen proses ini dinamakan sebagai operator.

Process View: Operator

- Deskripsi dari input yang diharapkan.
- Deskripsi dari output yang disediakan.
- Tindakan yang dilakukan oleh operator pada input, yang akhirnya mengarah dengan penyediaan output.
- Sejumlah parameter yang dapat mengontrol action performed.

Parameter View

- Beberapa operator dalam RapidMiner membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar.
- Namun terkadang parameter tidak mutlak dibutuhkan, meskipun eksekusi operator dapat dikendalikan dengan menunjukkan nilai parameter tertentu.
- Parameter view memiliki toolbar sendiri sama seperti view-view yang lain, dimana terdapat beberapa ikon dan nama-nama operator terkini yang diikuti dengan aktual parameter.

Parameter View (1)

- Huruf tebal berarti bahwa parameter mutlak harus didefinisikan oleh analis dan tidak memiliki nilai default.
- Sedangkan huruf miring berarti bahwa parameter diklasifikasikan sebagai parameter ahli dan seharusnya tidak harus diubah oleh pemula untuk analisis data.
- Poin pentingnya ialah beberapa parameter hanya ditunjukkan ketika parameter lain memiliki nilai tertentu.



Help & Comment View

- Setiap kali Anda memilih operator pada Operator View atau Process View, maka jendela bantuan dalam Help View akan menunjukkan penjelasan mengenai operator ini.
- Penjelasan yang ditampilkan dalam Help View meliputi:
 - Sebuah penjelasan singkat mengenai fungsi operator dalam satu atau beberapa kalimat.
 - Sebuah penjelasan rinci mengenai fungsi operator.
 - Daftar semua parameter termasuk deskripsi singkat dari parameter, nilai default (jika tersedia), petunjuk apakah parameter ini adalah parameter ahli serta indikasi parameter dependensi.

Help & Comment View

- Sedangkan Comment View merupakan area bagi Anda untuk menuliskan komentar pada langkah-langkah proses tertentu.
- Untuk membuat komentar, Anda hanya perlu memilih operator dan menulis teks di atasnya dalam bidang komentar.
- Kemudian komentar tersebut disimpan bersama-sama dengan definisi proses Anda. Komentar ini dapat berguna untuk melacak langkah-langkah tertentu dalam rancangan nantinya.

Problem View

- Problem View merupakan komponen yang sangat berharga dan merupakan sumber bantuan bagi Anda selama merancang proses analisis.
- Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam Problem View, seperti gambar di berikut ini.



Problem View(2)

- Pada kolom Message, Anda akan menemukan ringkasan pendek dari masalah.
- Kolom Location berisi tempat di mana masalah muncul dalam bentuk nama Operator dan nama port input yang bersangkutan.
- Kolom Fixes memberikan gambaran dari kemungkinan solusi tersebut, baik secara langsung sebagai teks (jika hanya ada satu kemungkinan Solusi) atau sebagai indikasi dari berapa banyak kemungkinan yang berbeda untuk memecahkan masalah.

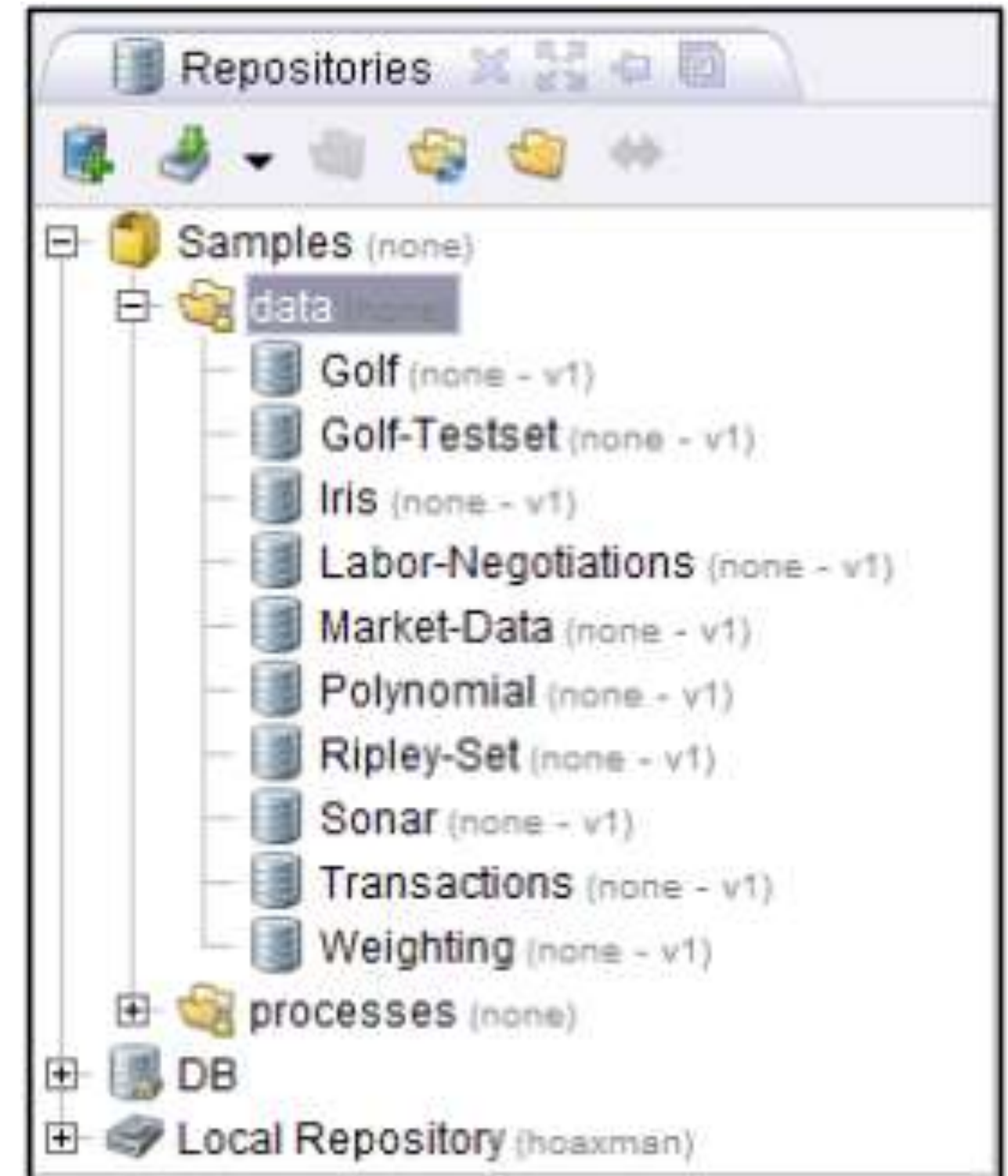


Cara Menggunakan Repository

- Repositori merupakan Tabel, database, koleksi teks, yang kita miliki untuk dapat digali datanya untuk mendapatkan informasi yang kita inginkan.
- Ini merupakan awal dari seluruh proses Data Mining. Maka dari itu adalah penting bagi kita untuk mengetahui cara menggunakan repository:
 - Sample data repository
 - Import repository

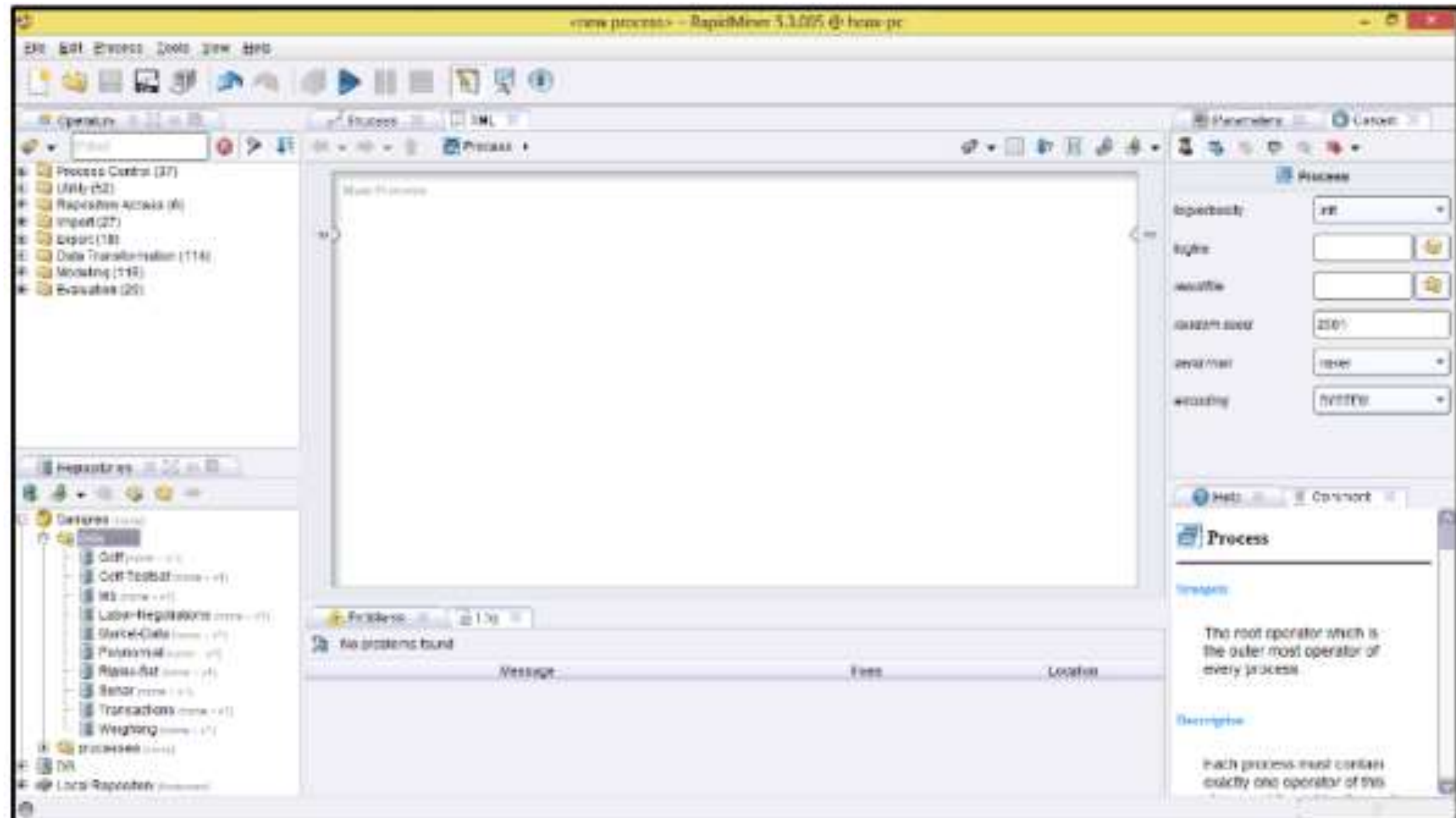
Sample Data Repository

- RapidMiner menyediakan contoh database yang dapat digunakan, berikut cara menggunakan Sample Data Repository.
- Pada bagian Repositori terdapat 3 buah lokasi repositori, yakni Samples, DB dan Local Repository.



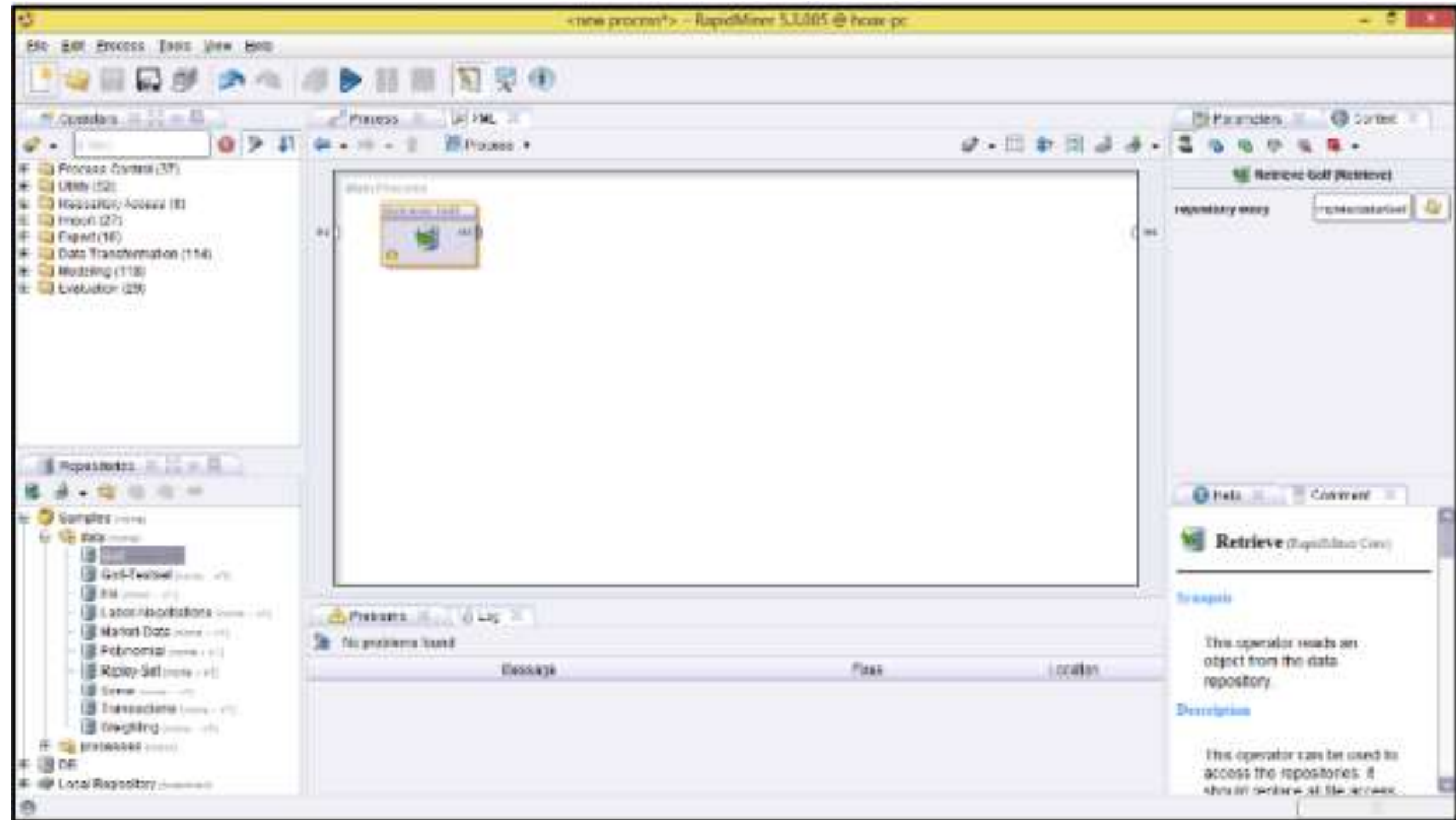
Sample Data Repository

- Untuk mengambil Sample Data Repository, buka hirarki Samples, masuk ke folder Data. Sehingga seperti gambar tampilan design perspective awal berikut:



Sample Data Repository

- Lakukan Drag dan Drop salah satu Example Repository. Kita ambil contoh Golf. Tarik dan lepaskan repository ke dalam Main Process, sehingga seperti gambar berikut.



Sample Data Repository

- Hubungkan output pada Database ke Result seperti Gambar diatas. Lalu klik ikon Play. Berikut adalah Sample data repository dari Golf.

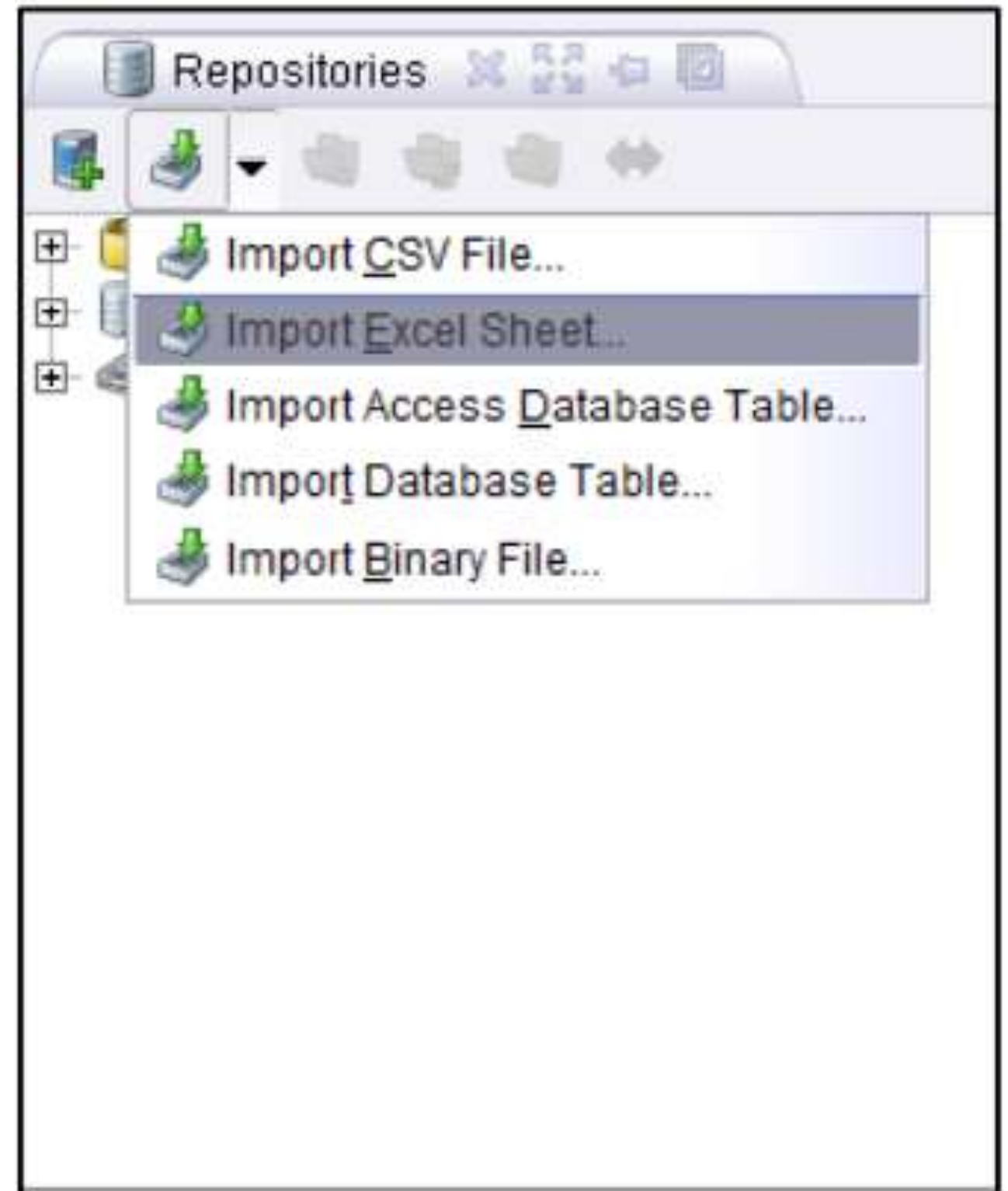


The screenshot shows a 'Result Overview' window for 'ExampleSet (Retrieve Golf)'. It includes a toolbar with 'Data View' (selected), 'Meta Data View', 'Plot View', 'Advanced Charts', and 'Annotations'. Below the toolbar, it states 'ExampleSet (14 examples, 1 special attribute, 4 regular attributes)'. The data is presented in a table with columns: Row No., Play, Outlook, Temperature, Humidity, and Wind.

Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

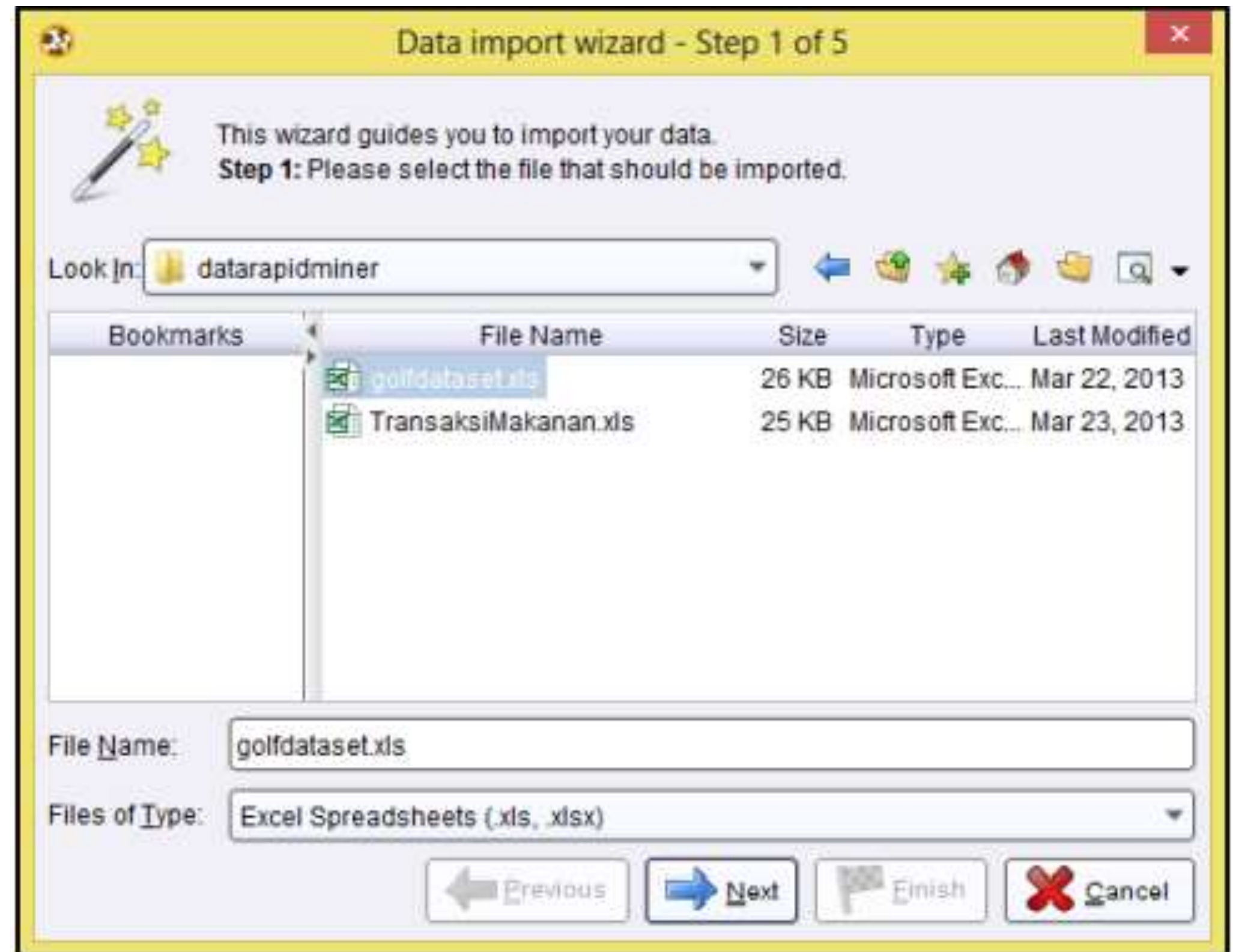
Import Repository

- RapidMiner menyediakan layanan agar pengguna dapat mengimport database miliknya dengan ekstensi .xls atau .xlsx, yakni file dari Microsoft Excel, Program yang relatif sering digunakan oleh pengguna.
- Berikut adalah cara untuk melakukan import file Microsoft Excel.
 - Lihat pada bagian Repository. Klik pada ikon import
 - Ada beberapa ekstensi file yang dapat kita masukkan kedalam repository kita. CSV File, Excel Sheen File, Access Database Table File, Database Table, Binary File.
 - Namun pada Dasarnya cara melakukan import pada semua file ini sama. Sebagai contoh, pilih Import Excel Sheet.



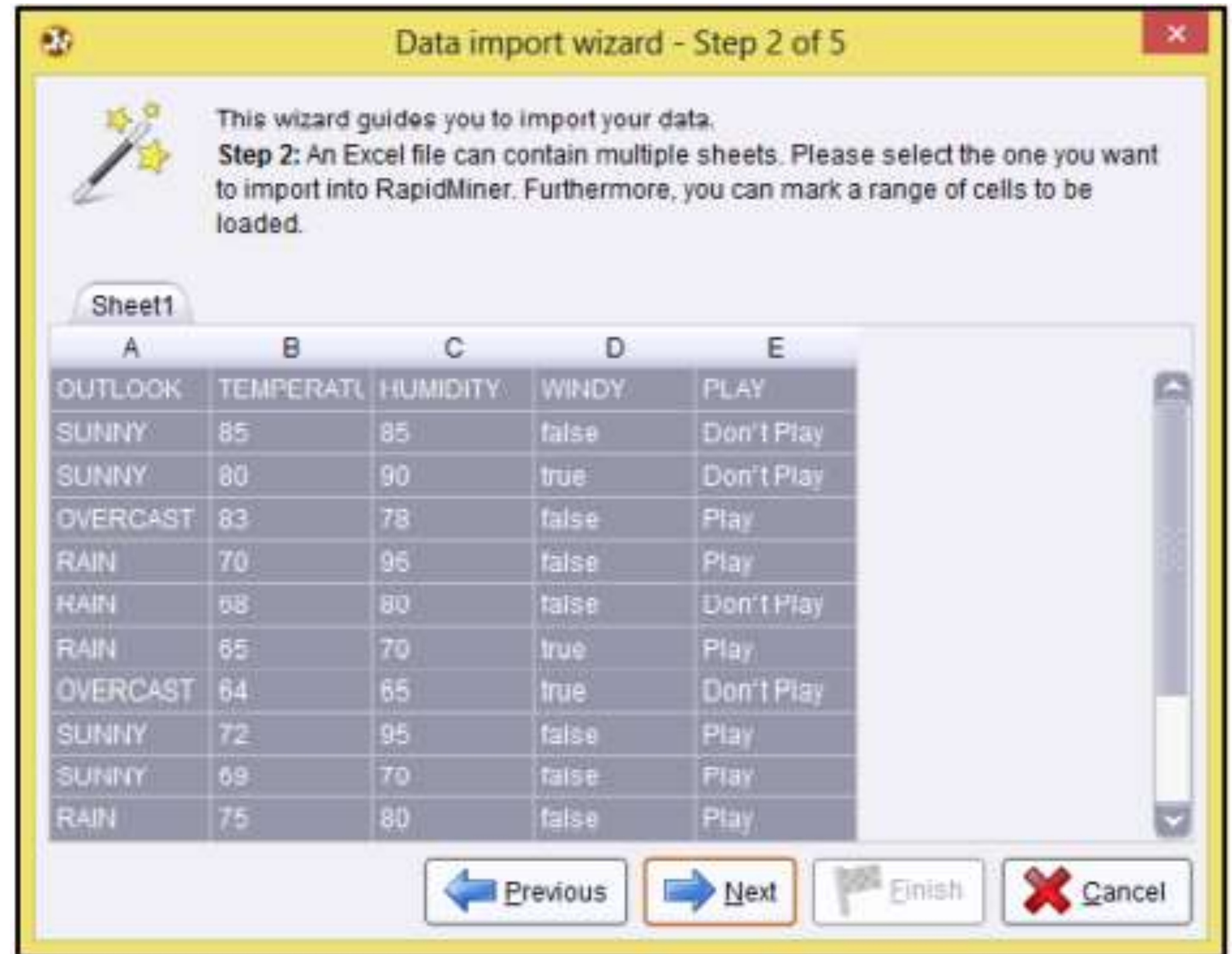
Import Repository

- Setelah itu, akan muncul window baru yakni Step 1 dari 5 Step Data import Wizard. Disini akan diarahkan oleh RapidMiner bagaimana langkah untuk melakukan import data.



Import Repository

- Kemudian pilih Sheet yang akan dimasukkan. Pada dasarnya, Repository RapidMiner hanya menyediakan 1 repositori untuk 1 buah table.



Import Repository

- Berikutnya ialah memberikan anotasi. Jika data kita tidak memiliki nama attribute, tidak usah melakukan apa-apa pada step 3 ini.

Data import wizard - Step 3 of 5

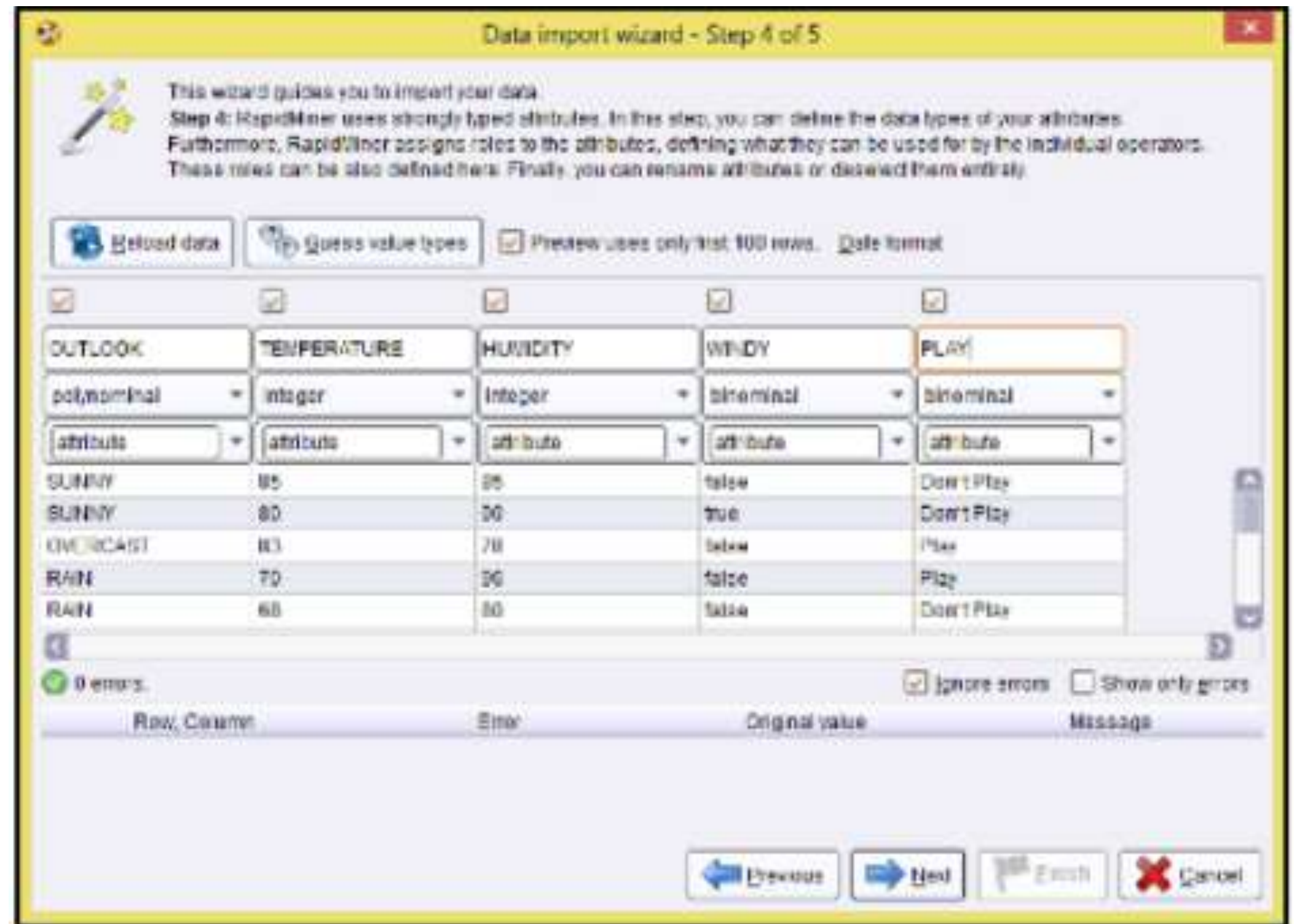
This wizard guides you to import your data.
Step 3: In RapidMiner, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotation	A	B	C	D	E
Name	OUTLOOK	TEMPERATU	HUMIDITY	WINDY	PLAY
-	SUNNY	85	85	false	Don't Play
-	SUNNY	80	90	true	Don't Play PLAY
-	OVERCAST	83	78	false	Play
-	RAIN	70	96	false	Play
-	RAIN	68	80	false	Don't Play
-	RAIN	65	70	true	Play
-	OVERCAST	64	65	true	Don't Play
-	SUNNY	72	95	false	Play
-	SUNNY	69	70	false	Play
-	RAIN	75	80	false	Play

Navigation buttons: **Previous**, **Next**, **Finish**, **Cancel**

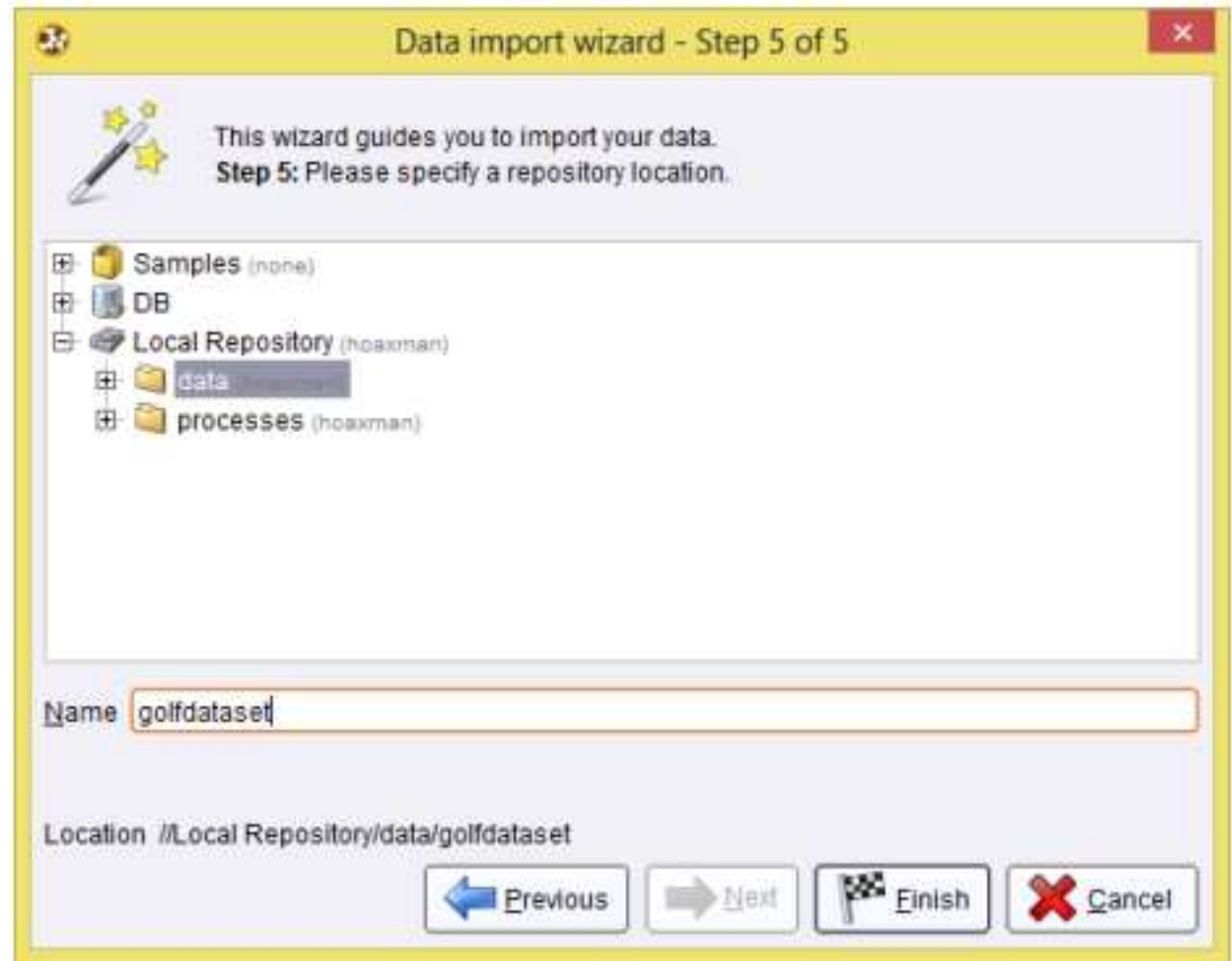
Import Repository

- Step ke 4 adalah memberikan tipe data pada tabel kita. Sebenarnya RapidMiner akan memberikan tipe data yang tepat secara otomatis.
- Namun, jika kita merasa tipe data yang diberikan RapidMiner tidak cocok, kita bisa mengubahnya



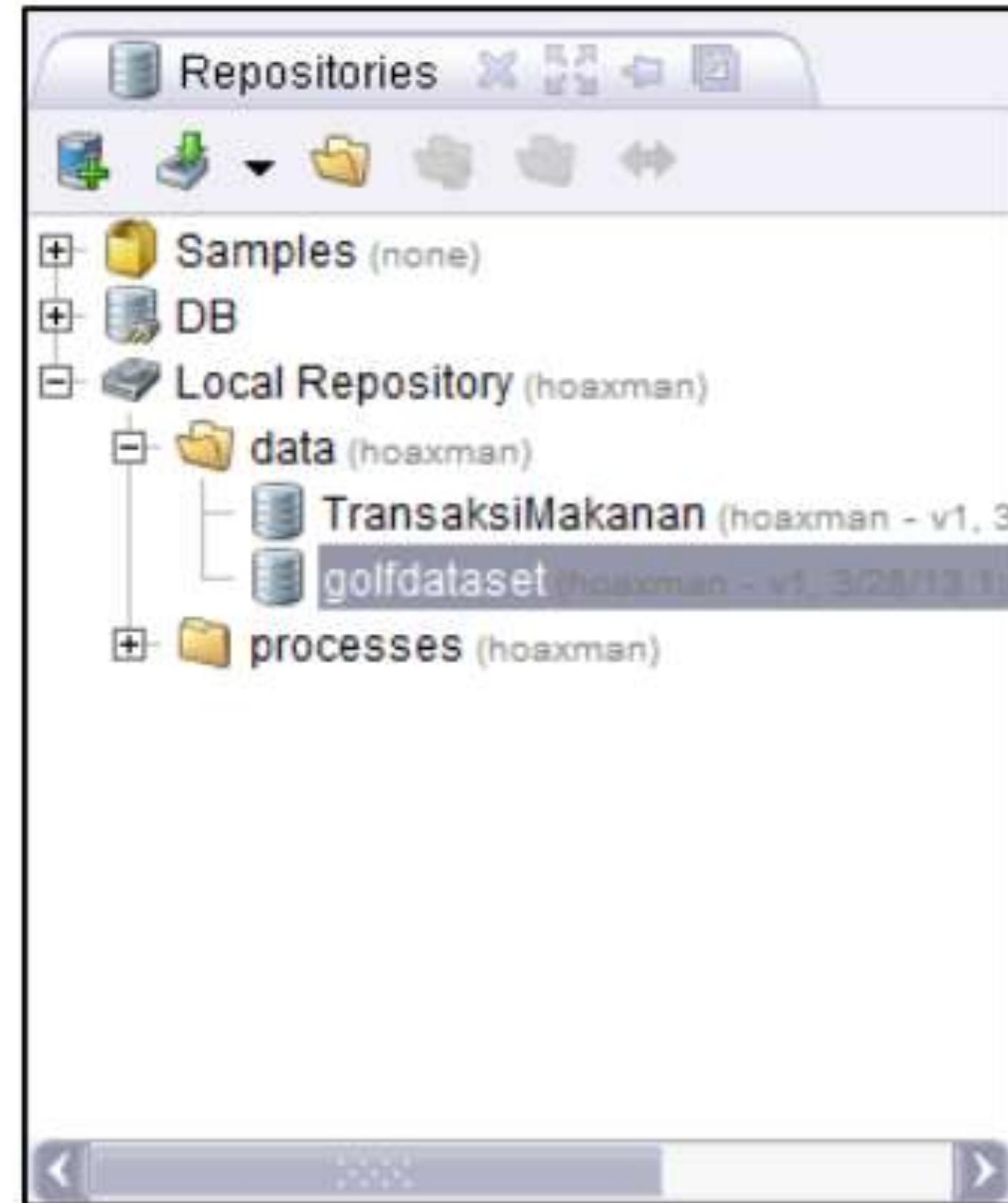
Import Repository

- Step ke 5 adalah memasukkan database kita ke dalam repository.
- Disarankan untuk memasukkannya kedalam Local Repository untuk memudahkan kita mencarinya.
- Jangan lupa untuk memberikan nama repository kita.
- Klik tombol **Finish**.



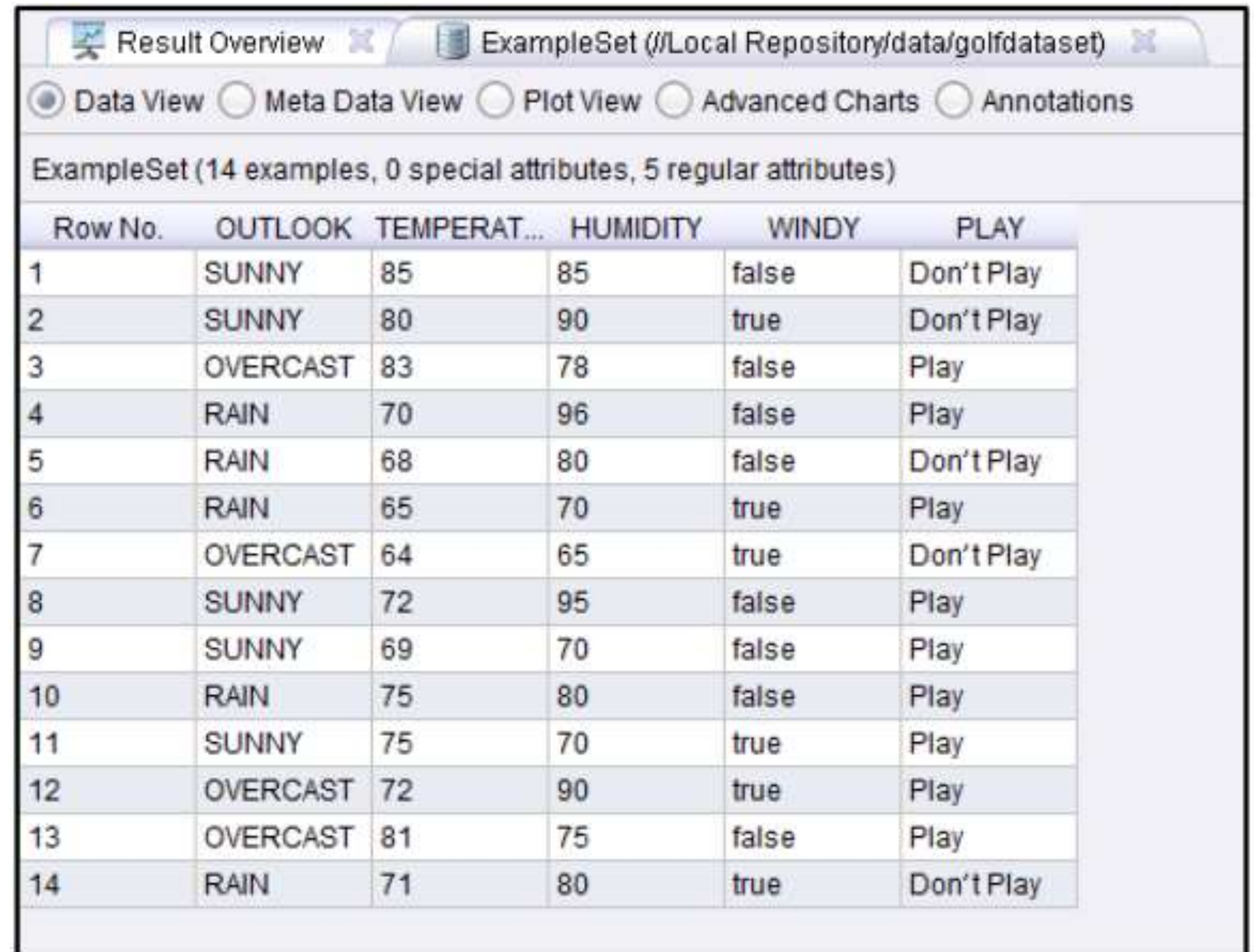
Import Repository

- Hasil Import Repository akan terlihat pada bagian Repository
- Untuk melihat isi dari repository kita, hubungkan output pada repository ke arah result.



Import Repository

- Setelah klik ikon Play, berikutnya akan muncul isi dari tabel yang kalian miliki.



The screenshot shows a software window titled 'Result Overview' with a sub-tab 'ExampleSet (//Local Repository/data/golfdataset)'. Below the title bar, there are radio buttons for 'Data View' (selected), 'Meta Data View', 'Plot View', 'Advanced Charts', and 'Annotations'. The main content area displays 'ExampleSet (14 examples, 0 special attributes, 5 regular attributes)' and a table with 6 columns: 'Row No.', 'OUTLOOK', 'TEMPERAT...', 'HUMIDITY', 'WINDY', and 'PLAY'. The table contains 14 rows of data.

Row No.	OUTLOOK	TEMPERAT...	HUMIDITY	WINDY	PLAY
1	SUNNY	85	85	false	Don't Play
2	SUNNY	80	90	true	Don't Play
3	OVERCAST	83	78	false	Play
4	RAIN	70	96	false	Play
5	RAIN	68	80	false	Don't Play
6	RAIN	65	70	true	Play
7	OVERCAST	64	65	true	Don't Play
8	SUNNY	72	95	false	Play
9	SUNNY	69	70	false	Play
10	RAIN	75	80	false	Play
11	SUNNY	75	70	true	Play
12	OVERCAST	72	90	true	Play
13	OVERCAST	81	75	false	Play
14	RAIN	71	80	true	Don't Play

