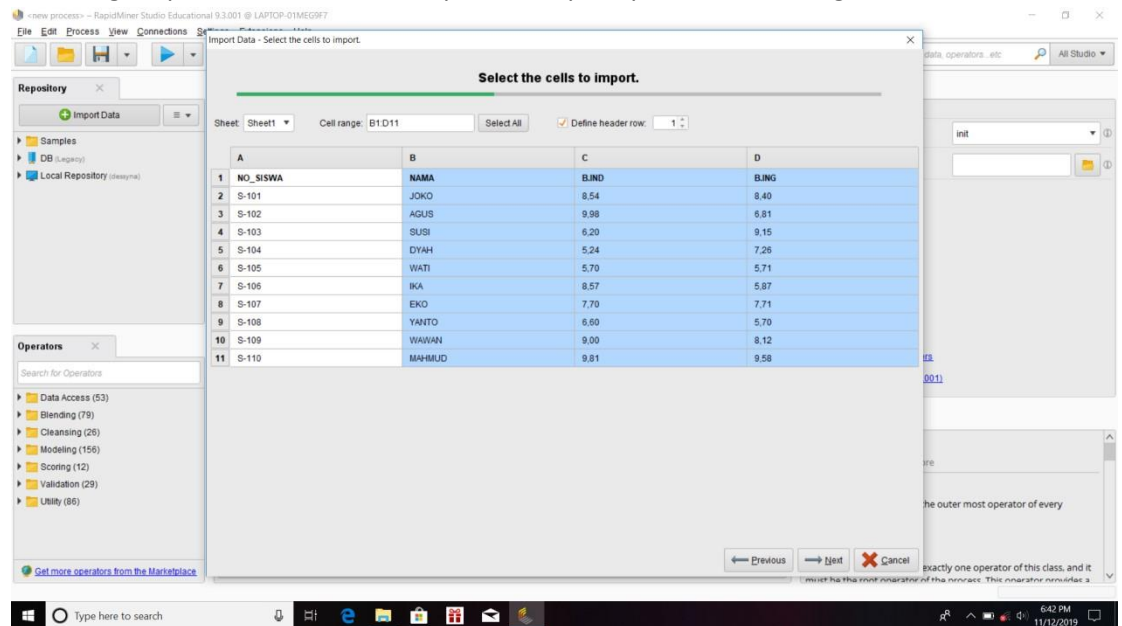


MODUL 10 CLUSTERING : K-MEANS

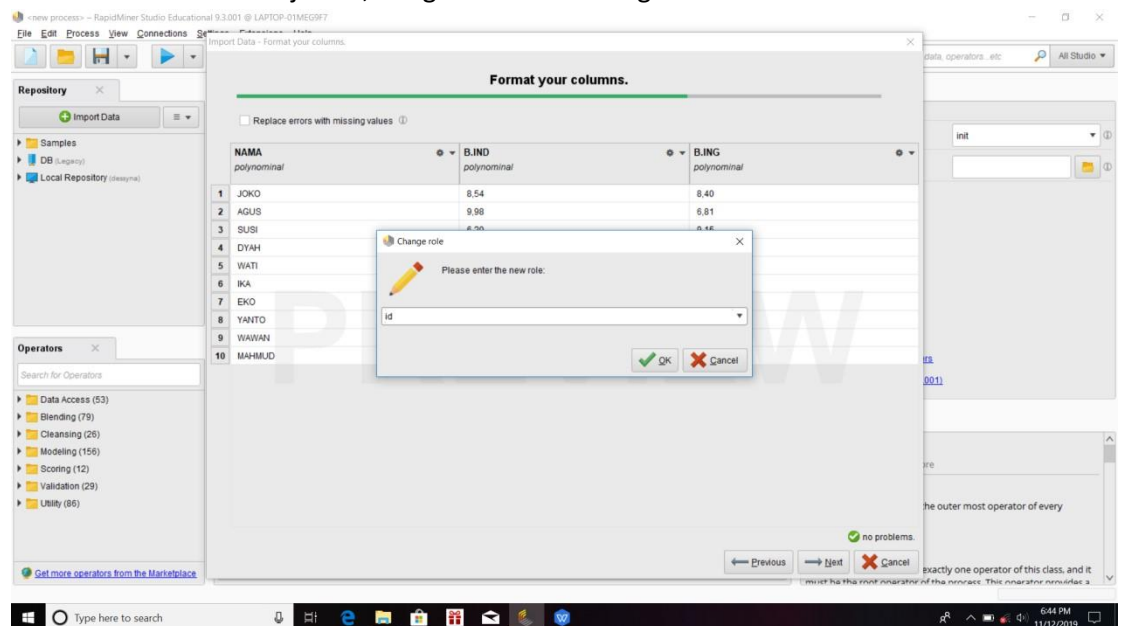
Langkah - langkah Praktikum

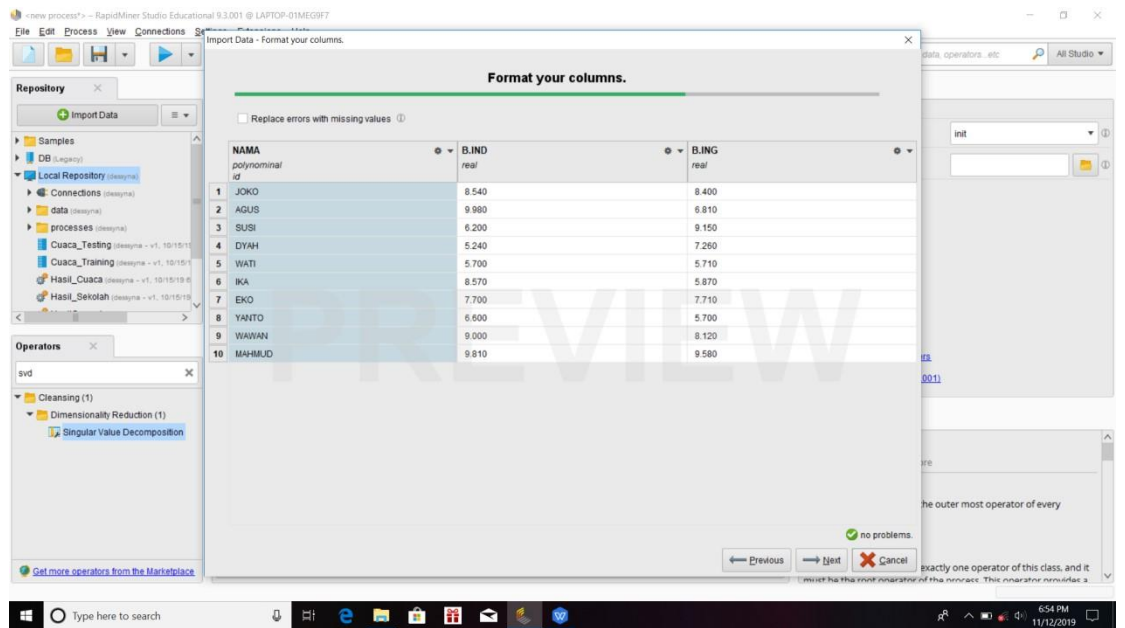
a. Algoritma K-Means Menggunakan RapidMiner

1. Buka Ms. Excel, dan buatlah tabel data nilai ujian siswa tersebut, Simpan dengan nama Tabel_NilaiUjian.xls
2. Jalankan aplikasi RapidMiner
3. Gunakan file Tabel_NilaiUjian.xls sebagai data yang akan digunakan dalam proses Clustering. Import file ini ke dalam repositori seperti pada Modul 8 Kegiatan 8.4.2

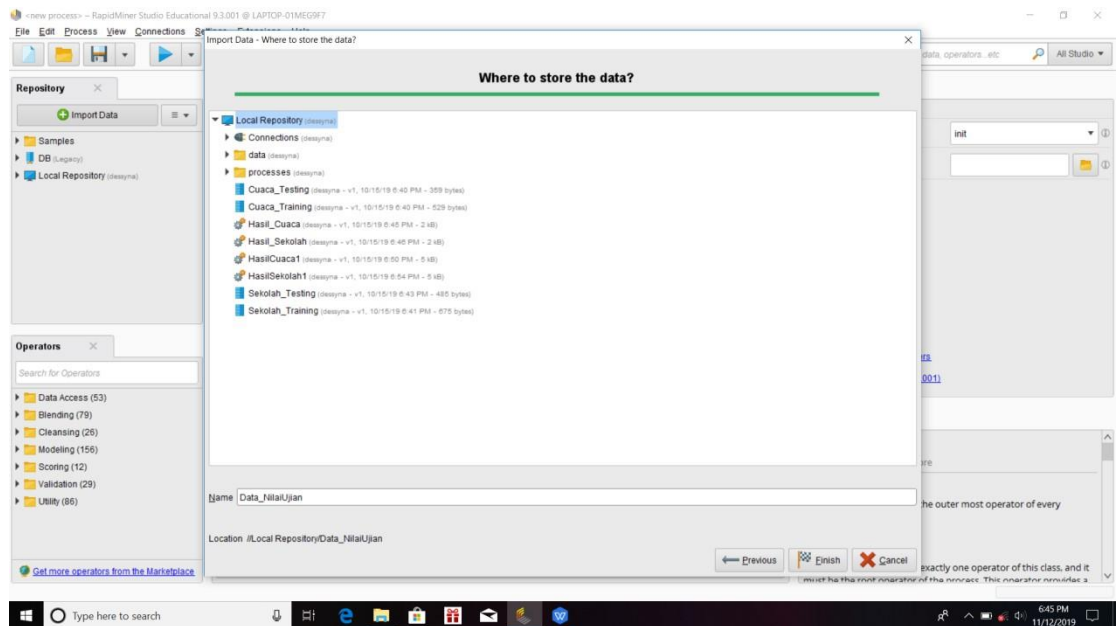


4. Ubah kolom NAMA menjadi id, dengan cara klik Change role >> id >> OK

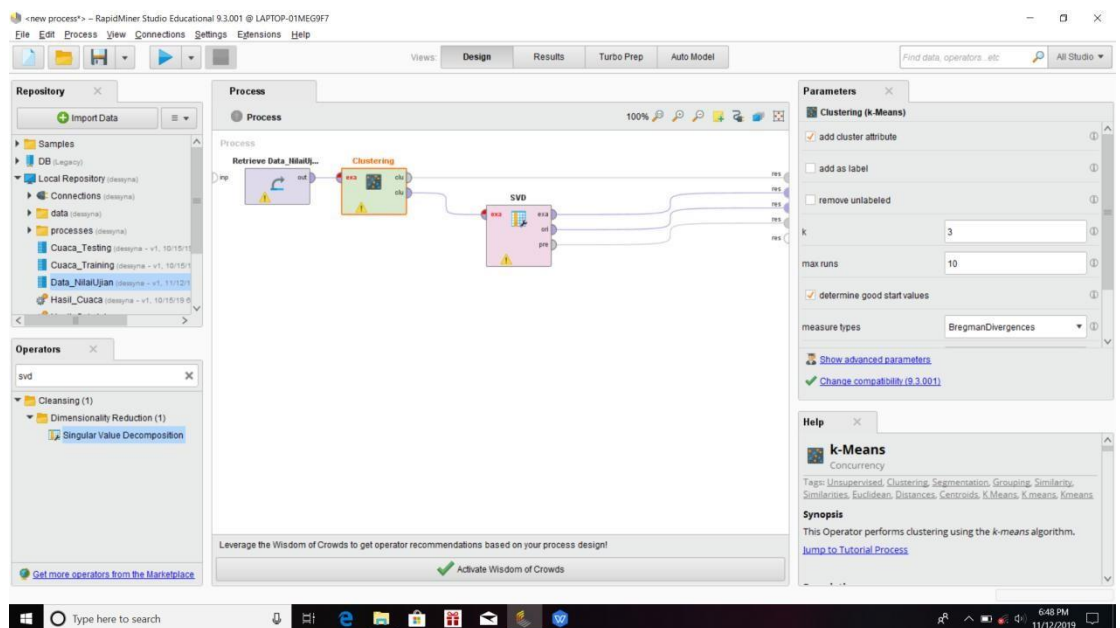




5. Beri nama Data_NilaiUiandan masukkan pada repositories. Kemudian klik finish



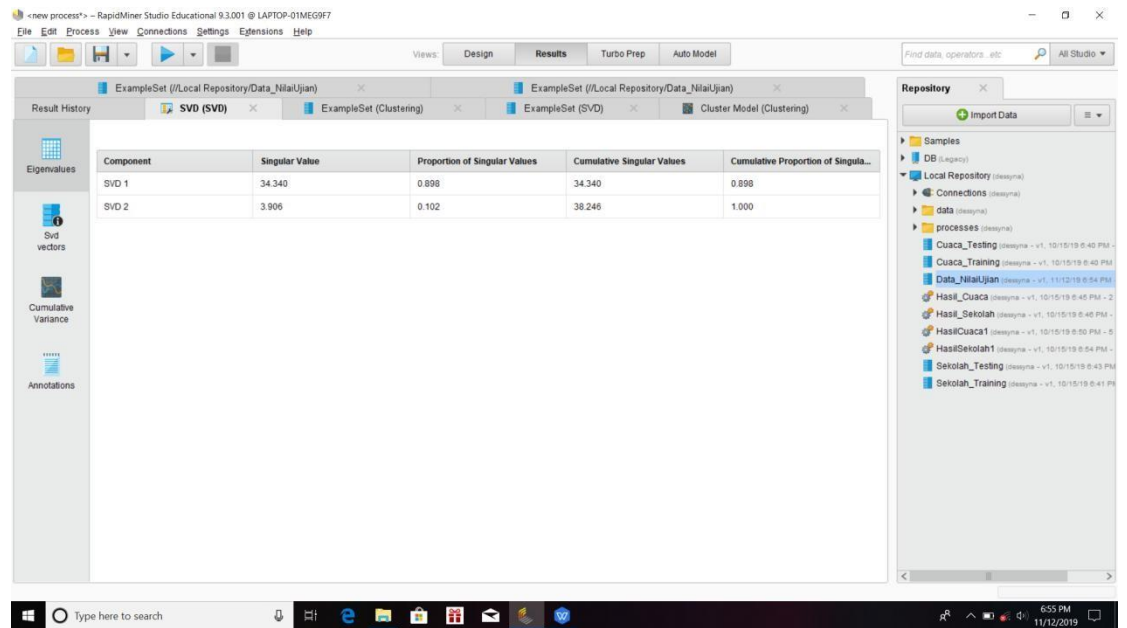
6. Gunakan Data_NilaiUjian ini dan masukkan dalam area process.
7. Tambahkan operator k-Means. Ubah nama operator ini menjadi k-Means. Hubungkan utput operator retrieve ke entry exa operator ini dan output clu (cluster model) dihubungkan ke connector res panel. Ubah nilai parameter k = 3 pada operator ini. Angka ini digunakan untuk menentukan jumlah kelompok siswa.



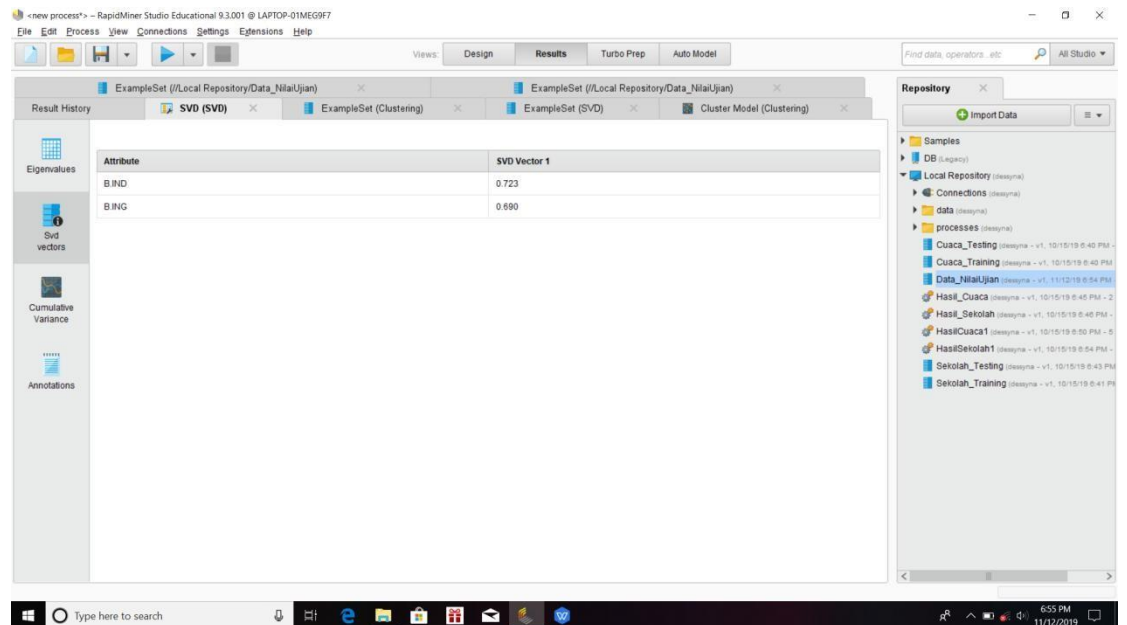
8. Tambahkan pula operator SVD (Singular Value Decompositon). hubungkan output clu (clustered set) ke-2 operaor clustering(k-Means) ke dalam entry exa operator SVD dan 3 port output exa, ori dan pre terhadap connector panel res.
9. Jalankan process dengan menekan tombol Run.
10. Berikut hasil process Clusteringdengan algoritma K-Means.

a) SVD

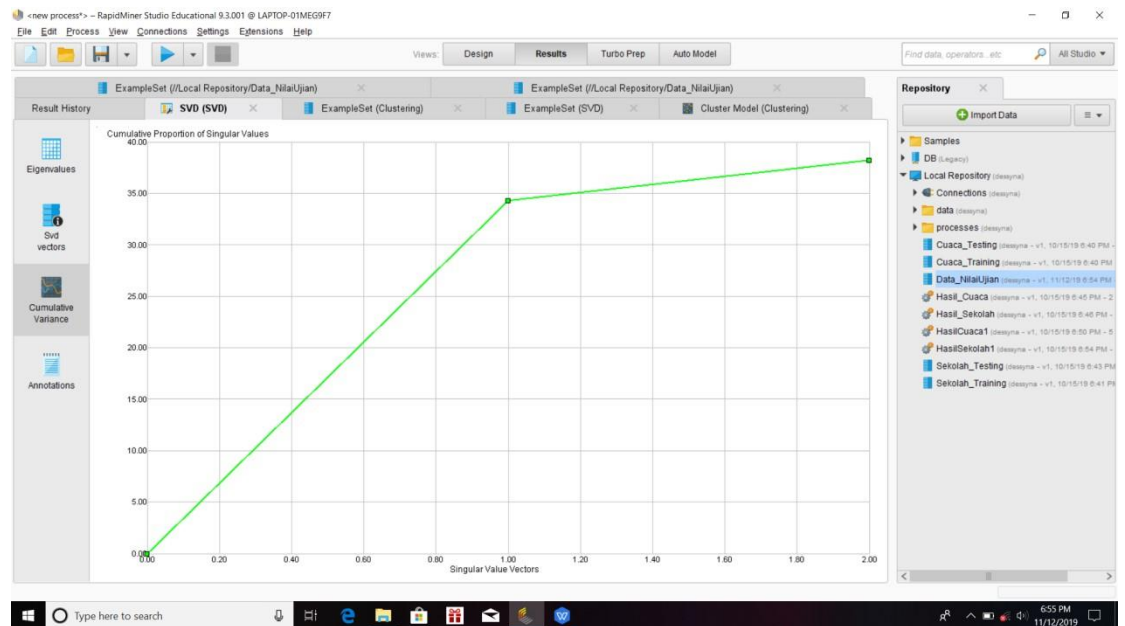
i. Nilai Eigenvalue



ii. Nilai Svd vectors

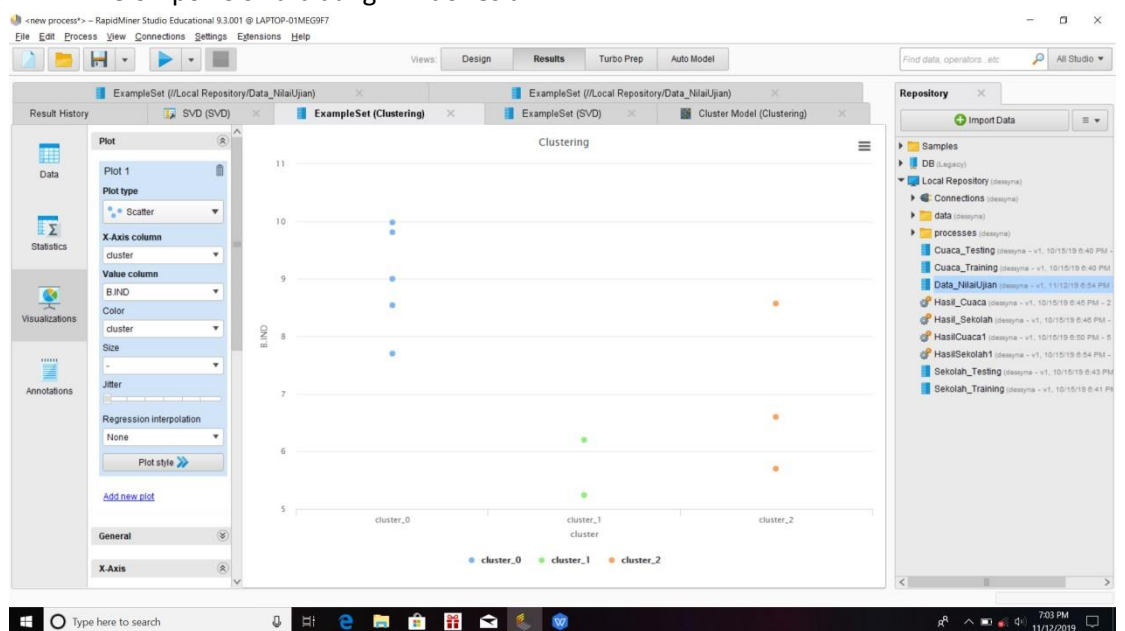


iii. Nilai cumulative variance

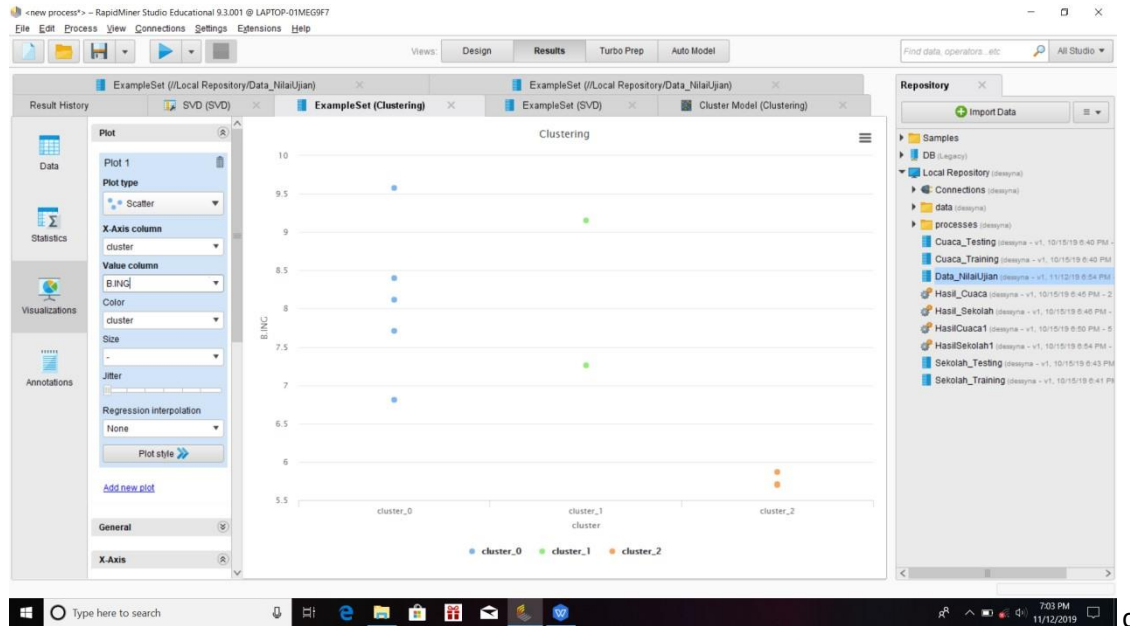


b) ExampleSet (k-Means)

i. Kelompok Siswa bidang B.Indonesia



ii. Kelompok Siswa bidang B.Inggris



c)

ExampleSet (SVD)

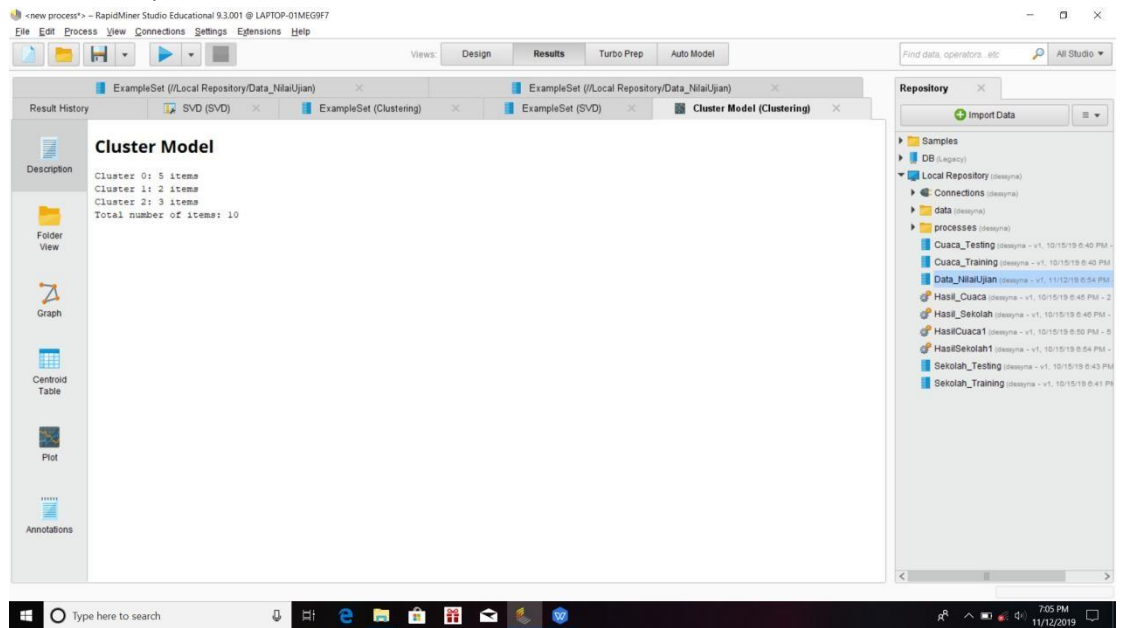
The screenshot shows the RapidMiner Studio interface. The main window displays the 'ExampleSet (SVD)' table. The table has 10 rows and 4 columns: 'Row No.', 'NAMA', 'cluster', and 'svd_1'. The data is as follows:

Row No.	NAMA	cluster	svd_1
1	JOKO	cluster_0	0.349
2	AGUS	cluster_0	0.347
3	SUSI	cluster_1	0.315
4	DYAH	cluster_1	0.256
5	WATI	cluster_2	0.235
6	IKA	cluster_2	0.299
7	EKO	cluster_0	0.317
8	YANTO	cluster_2	0.254
9	WAWAN	cluster_0	0.353
10	MAHMUD	cluster_0	0.399

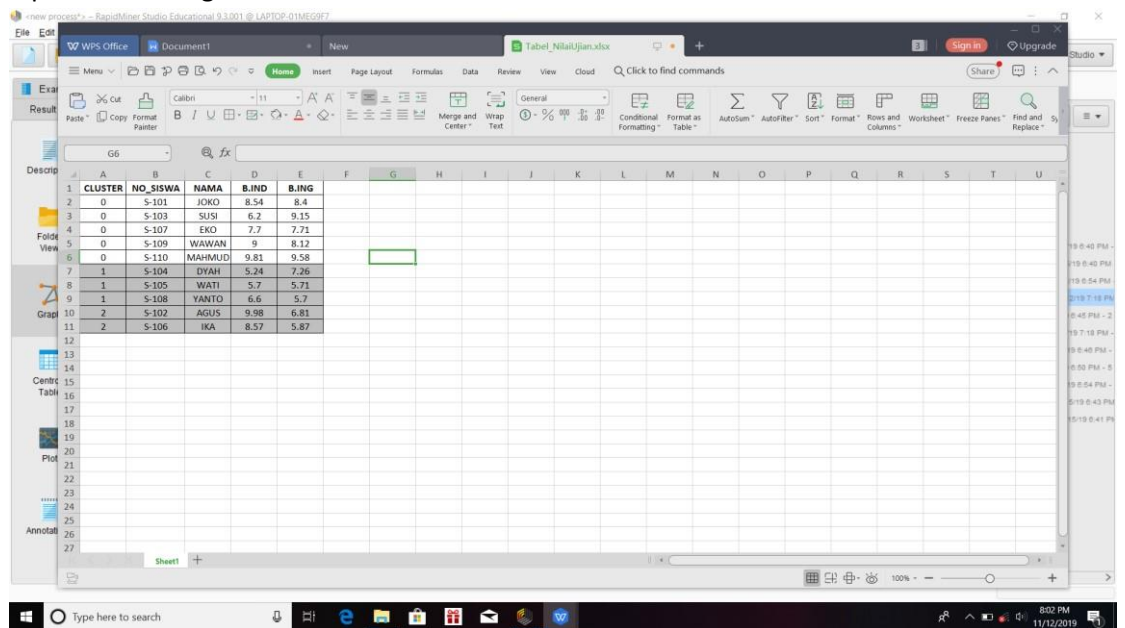
The table is titled 'ExampleSet (10 examples, 2 special attributes, 1 regular attribute)'. The 'Repository' panel on the right shows a list of data sources, including 'Cuaca_Testing', 'Cuaca_Training', 'Data_NilaiUjian', 'Hasil_Cuaca', 'HasilSekolah', 'Sekolah_Testing', and 'Sekolah_Training'.

d) Cluster Model (Clustering)

i. Description



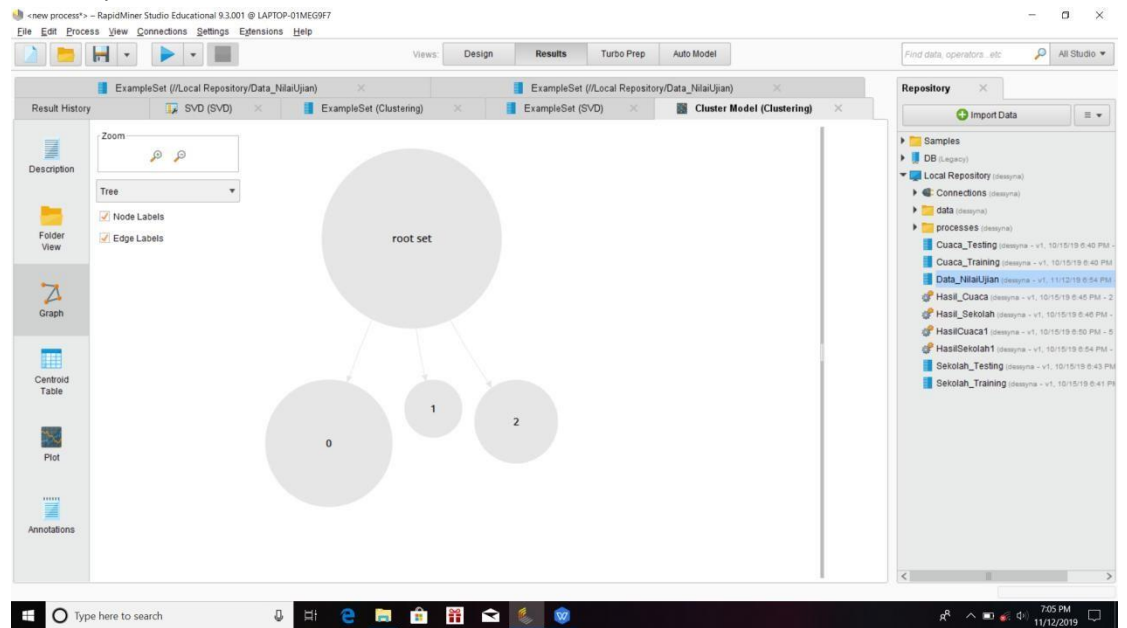
b. Interpretasi Hasil Algoritma K-Means



Pembagian kelompok yang diajukan untuk lomba olimpiade :

1. Cluster_2 yang diajukan untuk lomba olimpiade bidang B.Indonesia
2. Cluster_0 yang diajukan untuk lomba olimpiade bidang B.Ingggris

ii. Graph



Tugas

1. Buatlah tabe berikut dengan menggunakan Microsoft Excel !

The following table represents the data shown in the Excel spreadsheet across two screenshots.

No. Siswa	Nama	B.IND	B.ING	MTK	IPA
S-101	JOKO	6.946091051	9.76655197	6.400701886	5.561797903
S-102	AGUS	8.927866861	9.552811877	7.834146055	9.407904028
S-103	SUSI	8.783662979	6.069341438	5.001231269	7.734338663
S-104	DYAH	9.142285245	7.702192587	8.041770661	9.03306207
S-105	WATI	8.065555689	7.923331391	5.338498288	6.72221931
S-106	IKA	7.354345484	9.118292395	9.668485457	9.223160414
S-107	EKO	7.37341481	5.782211515	7.846805065	8.986749852
S-108	YANTO	6.187876862	6.138585665	9.380143341	5.640649491
S-109	WAWAN	6.352438361	7.667098448	9.027599108	8.30302766
S-110	MAHMUD	9.557750774	7.208304925	8.676005152	7.828432092
S-111	BUDI	5.000221478	7.982335875	9.244884025	5.583383635
S-112	SANTI	7.832484303	7.905048622	7.586385154	5.714966367
S-113	DIAN	6.36937844	8.454482483	9.701079636	6.878686741
S-114	DANI	9.205411027	8.834176759	5.213821488	5.075841129
S-115	AHMAD	6.672282117	5.622583461	7.497860401	5.956249118
S-116	BAYU	5.128274954	5.062253146	5.497970115	8.562173563
S-117	RISA	5.914562545	9.257454574	6.03388747	5.181892126
S-118	RANI	9.642489592	7.121403324	8.023001825	6.752380073
S-119	YANI	7.50765808	5.367688022	9.5561954	8.865848409
S-120	RATIH	6.524777301	8.502775869	6.016231659	9.135420483
S-121	INDAH	6.485487012	6.574451248	5.150098078	5.319564425
S-122	JONO	9.785377166	9.835131244	7.027548972	5.730500811
S-123	SARAH	8.208108527	6.924354059	6.814151404	8.481927472
S-124	NANA	7.590406452	7.360962342	5.828736084	7.597229715
S-125	BAMBANG	8.038076793	5.107019747	5.73208641	9.34551479
S-126	HADI	7.006021651	7.475932852	6.264661311	6.584696993
S-127	NANA	5.075351866	7.682685913	6.130761757	6.264368747
S-128	FEBRI	9.597445846	8.514587056	6.785882682	5.172053349
S-129	DENI	7.557827573	8.063801813	6.086578441	8.757632073
S-130	TONI	6.340335171	6.170818596	8.295743535	6.658043851

2. Lakukan kembali kegiatan 10.4.1 dan 10.4.2 pada modul 10 tabel Tabel_NilaiUjian 30 siswa tersebut dengan ketentuan jumlah Cluster = 4. catat dan tulis semua hasilnya pada lembar jawaban anda

- a) Mengimport data nilai 30 siswa ke dalam rapidMiner

Import Data - Select the cells to import.

Select the cells to import.

Sheet: Sheet1 Cell range: B1:F31 Selected All Define header row: 1

	A	B	C	D	E	F
11	S-110	MAHMUD	6.959	9.358	8.502	7.832
12	S-111	BUDI	7.309	5.966	7.853	7.104
13	S-112	SANTI	9.603	9.650	8.657	8.708
14	S-113	DIAN	9.718	6.269	6.783	9.131
15	S-114	DANI	8.769	8.318	7.891	8.824
16	S-115	AHMAD	8.133	5.294	9.953	9.381
17	S-116	BAYU	6.811	8.177	9.560	6.083
18	S-117	RISA	5.286	5.307	9.870	8.942
19	S-118	RANI	8.148	8.629	8.585	8.704
20	S-119	YANI	6.653	7.034	5.025	7.687
21	S-120	RATIH	8.230	5.251	5.866	9.685
22	S-121	INDAH	6.730	7.101	8.758	6.544
23	S-122	JONO	9.067	6.117	7.414	9.644
24	S-123	SARAH	9.490	8.822	6.435	8.797
25	S-124	RIMA	5.491	9.094	6.624	7.843
26	S-125	BAMBANG	7.226	9.144	8.444	8.862
27	S-126	HADI	9.206	5.546	5.639	9.591
28	S-127	NANA	9.841	5.325	6.623	6.677
29	S-128	FEBRI	7.874	9.052	7.166	7.076
30	S-129	DENI	7.630	7.199	6.770	9.081
31	S-130	TONI	8.844	8.158	5.125	8.338

no problems.

iterating using the k-means algorithm.

b) Merubah nama menjadi id

Import Data - Format your columns.

Format your columns.

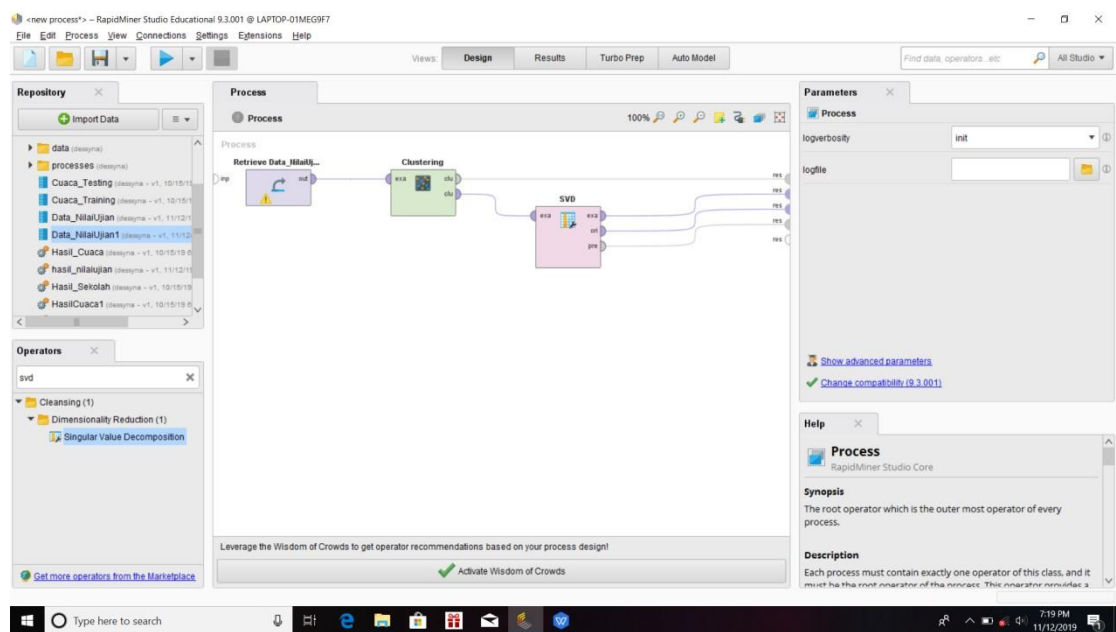
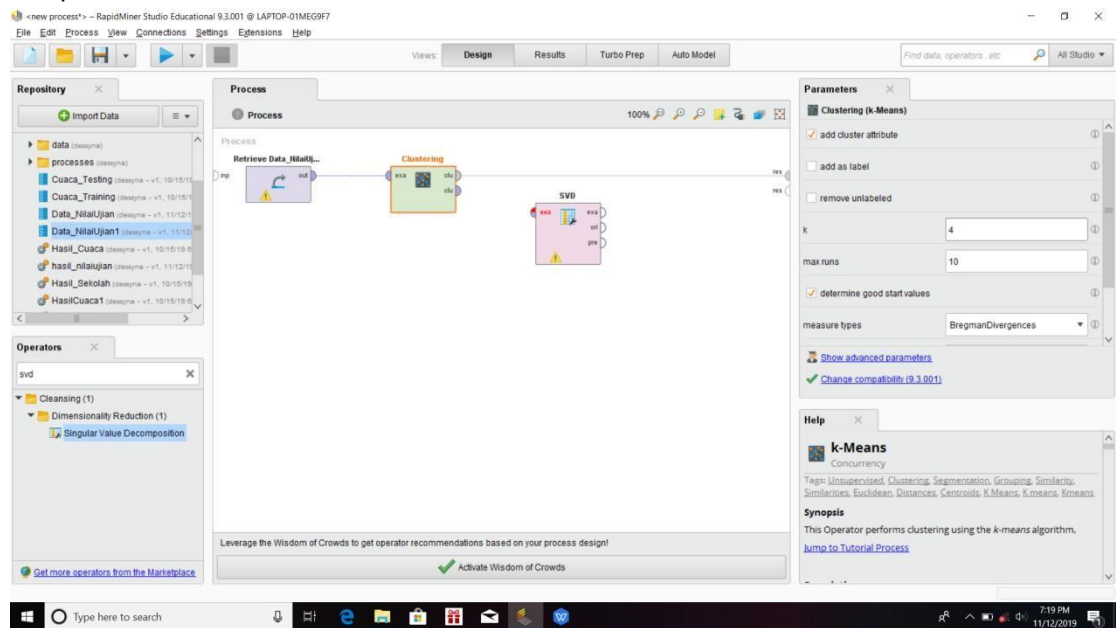
☐ Replace errors with missing values

	NAMA polynomial id	B.IND real	B.ING real	MTK real	IPA real
1	JOKO	5.671	8.252	7.110	7.535
2	AGUS	6.878	7.069	6.072	5.310
3	SUSI	8.750	5.994	9.847	5.603
4	DIYAH	5.092	7.966	5.672	9.594
5	WATI	9.254	7.735	8.849	5.909
6	IKA	6.805	6.294	9.209	9.586
7	EKO	7.156	9.033	5.983	7.439
8	YANTO	9.807	9.605	7.593	5.922
9	WAWAN	6.220	9.844	8.225	8.910
10	MAHMUD	6.959	9.358	8.502	7.832
11	BUDI	7.309	5.966	7.853	7.104
12	SANTI	9.603	9.650	8.657	8.708
13	DIAN	9.718	6.269	6.783	9.131
14	DANI	8.769	8.318	7.891	8.824
15	AHMAD	8.133	5.294	9.953	9.381
16	BAYU	6.811	8.177	9.560	6.083
17	RISA	5.286	5.307	9.870	8.942
18	RANI	8.148	8.629	8.585	8.704
19	YANI	6.653	7.034	5.025	7.687

no problems.

iterating using the k-means algorithm.

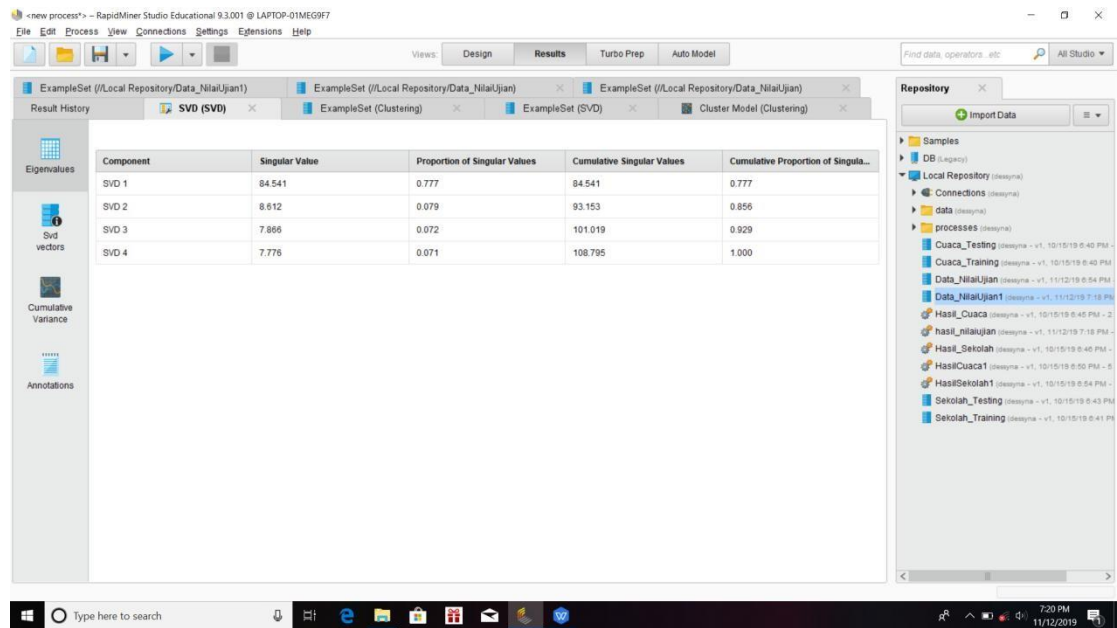
- c) Drag data nilai ke dalam lebar kerja lalu ditambahkan operator k-Means dan SVD pada K-Means dirubah k = 4



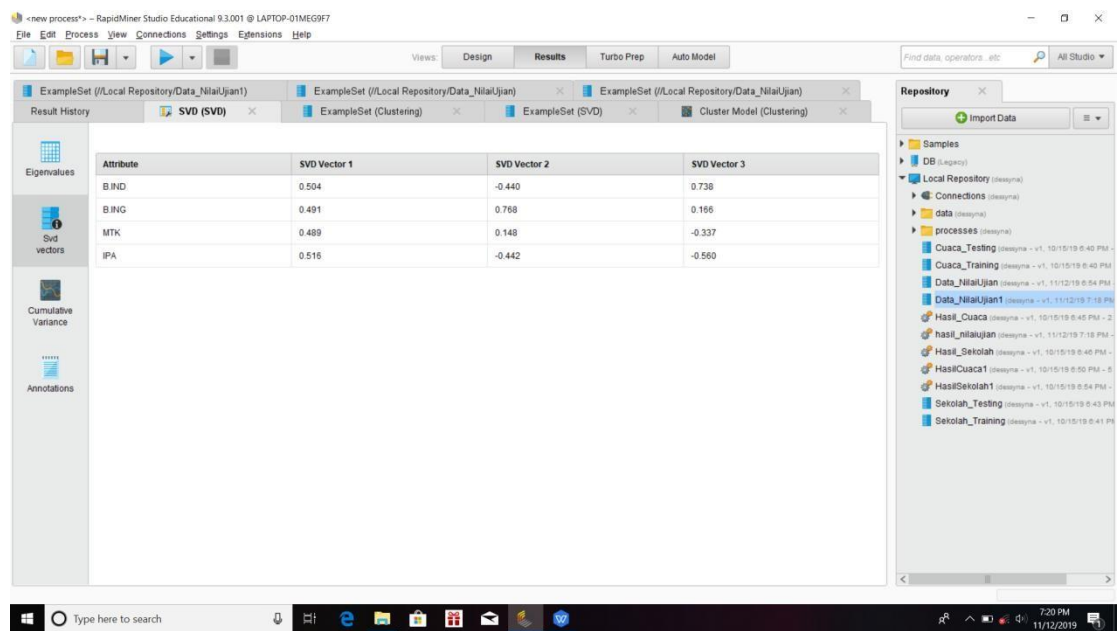
- d) Menjalankan program dengan cara di Run atau klik F11

a. SVD

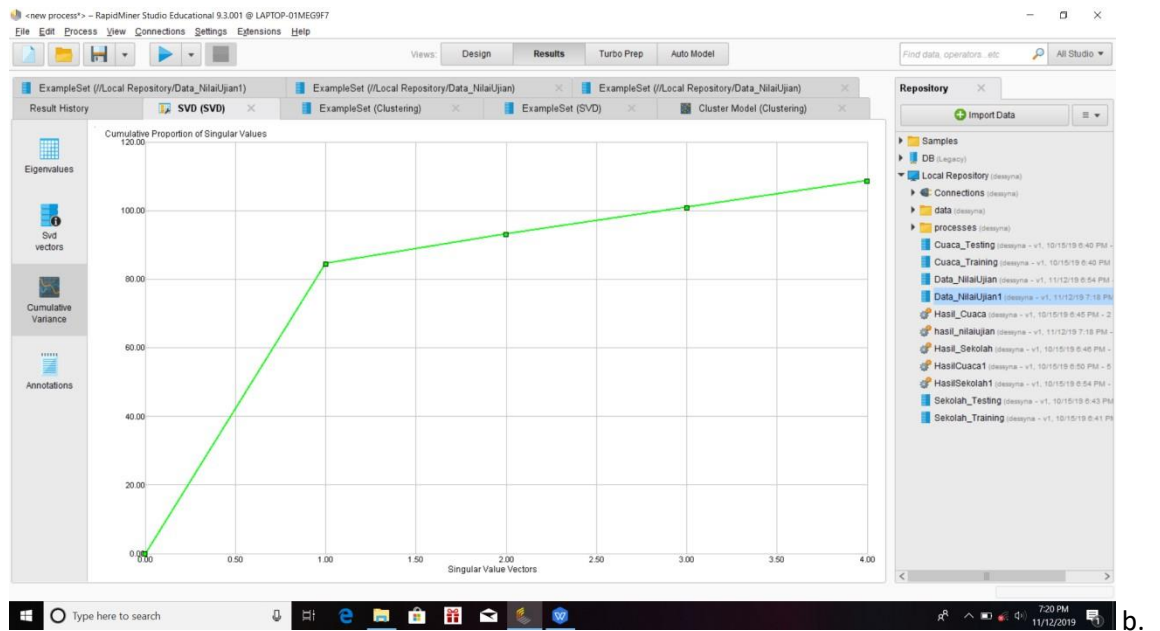
i. Nilai Eigenvalue



ii. Nilai Svd vectors



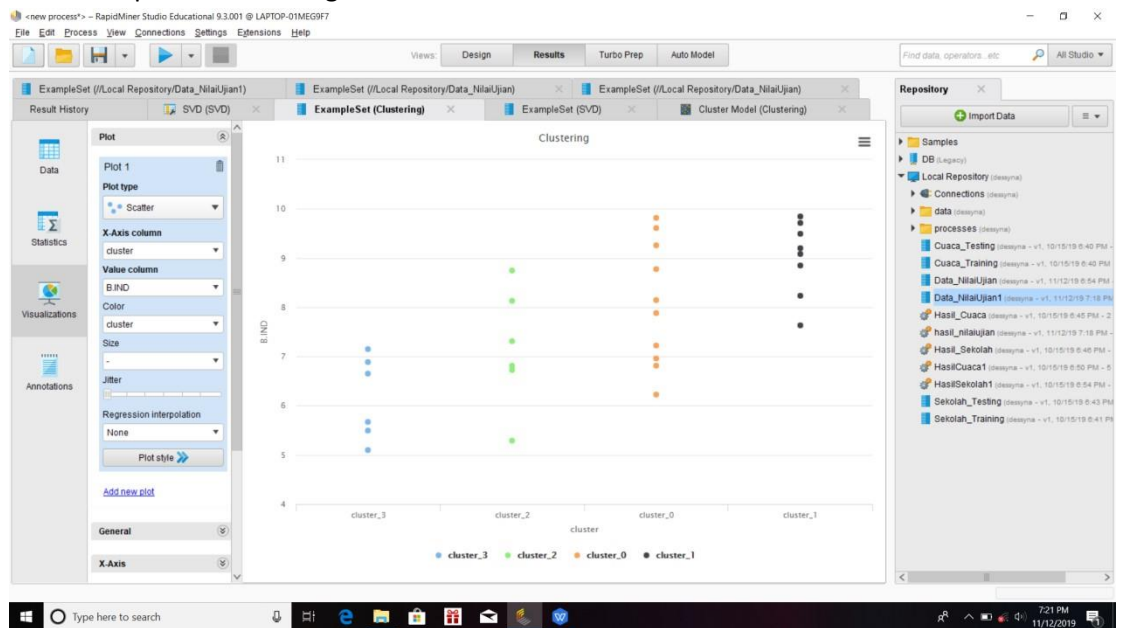
iii. Nilai Cumulative variance



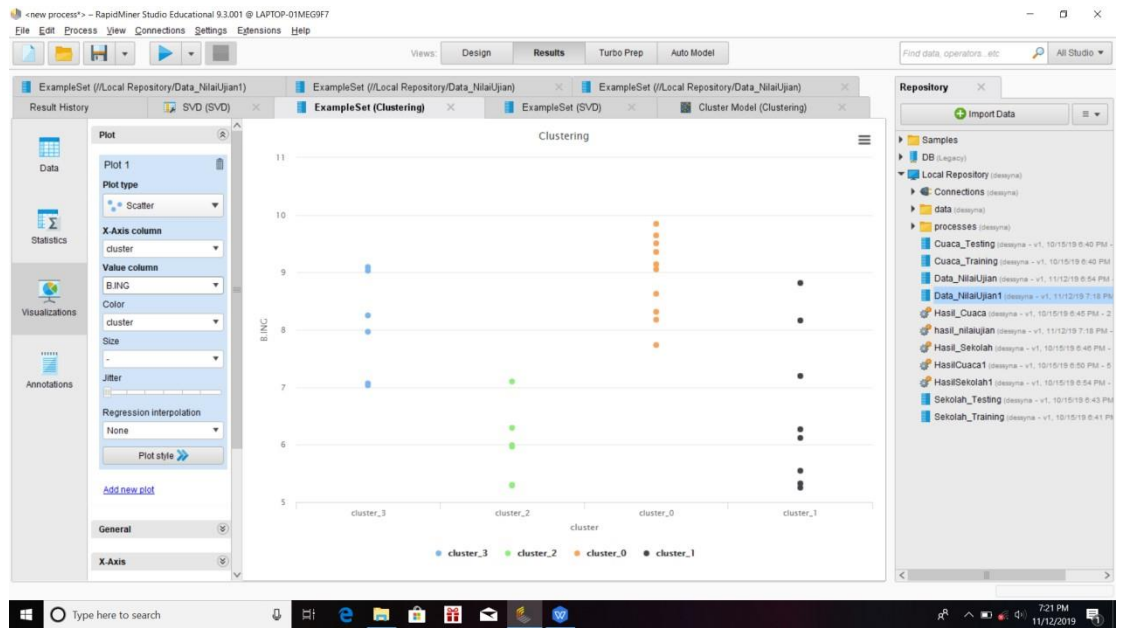
b.

ExampleSet (k-Means)

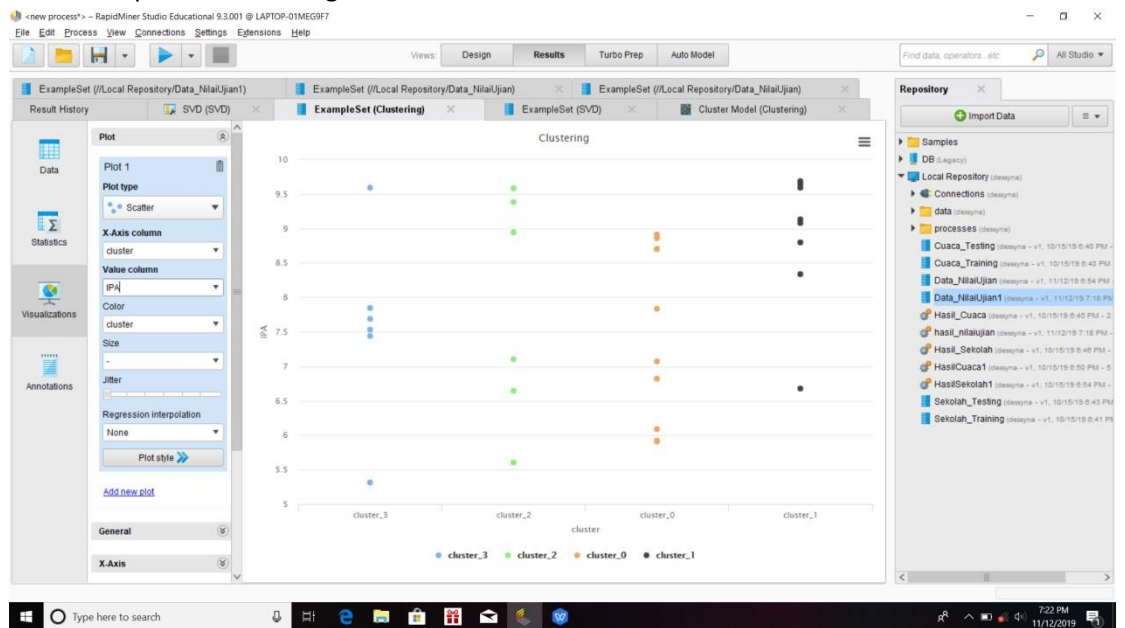
i. Kelompok Siswa bidang B.Indonesia



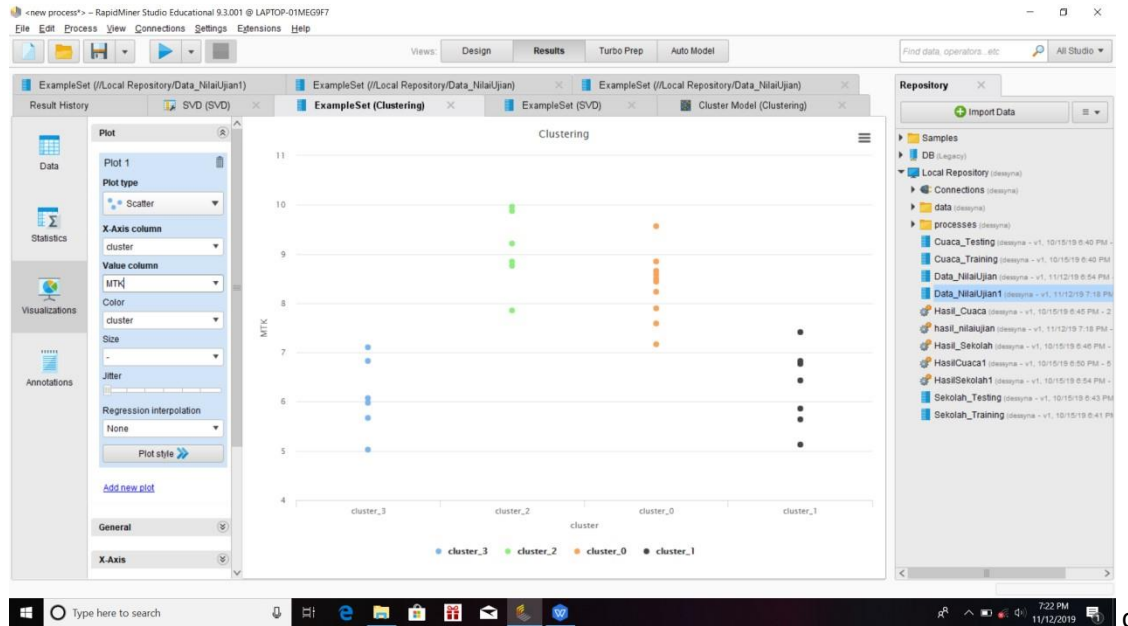
ii. Kelompok Siswa bidang B.Inggris



iii. Kelompok Siswa bidang IPA



iv. Kelompok Siswa bidang Matematika



c.

ExampleSet (SVD)

The screenshot shows the RapidMiner Studio interface with the 'ExampleSet (SVD)' table open. The table has the following columns: 'Row No.', 'NAMA', 'cluster', and 'svd_1'. The data is as follows:

Row No.	NAMA	cluster	svd_1
1	JOKO	cluster_3	0.169
2	AGUS	cluster_3	0.150
3	SUSI	cluster_2	0.172
4	DYAH	cluster_3	0.168
5	WATI	cluster_0	0.187
6	IKA	cluster_2	0.189
7	EKO	cluster_3	0.175
8	YANTO	cluster_0	0.194
9	WAWAN	cluster_0	0.196
10	MAHMUD	cluster_0	0.193
11	BUDI	cluster_2	0.167
12	SANTI	cluster_0	0.216
13	DIAN	cluster_1	0.189
14	DIAN	cluster_0	0.188
15	AHMAD	cluster_2	0.194
16	BAYU	cluster_0	0.180
17	RISA	cluster_2	0.174

Below the table, it says 'ExampleSet (30 examples, 2 special attributes, 1 regular attribute)'.

new process* - RapidMiner Studio Educational 9.3.001 @ LAPTOP-01MEG9F7

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

ExampleSet (/Local Repository/Data_NilaiUjian) ExampleSet (/Local Repository/Data_NilaiUjian) ExampleSet (/Local Repository/Data_NilaiUjian)

Result History SVD (SVD) ExampleSet (Clustering) ExampleSet (SVD) Cluster Model (Clustering)

Open in Turbo Prep Auto Model

Filter (30 / 30 examples): all

Row No.	NAMA	cluster	svd_1
14	DANI	cluster_0	0.188
15	AHMAD	cluster_2	0.194
16	BAYU	cluster_0	0.180
17	RISA	cluster_2	0.174
18	RANI	cluster_0	0.201
19	YANI	cluster_3	0.156
20	RATH	cluster_1	0.173
21	INDAH	cluster_2	0.173
22	JONO	cluster_1	0.191
23	SARAH	cluster_1	0.199
24	RAMA	cluster_3	0.173
25	BAMBANG	cluster_0	0.199
26	HADI	cluster_1	0.178
27	NANA	cluster_1	0.170
28	FEBRI	cluster_0	0.184
29	DENI	cluster_1	0.182
30	TONI	cluster_1	0.181

ExampleSet (30 examples, 2 special attributes, 1 regular attribute)

Repository

- Import Data
- Samples
- DB (Legacy)
- Local Repository (default)
 - Connections (default)
 - data (default)
 - processes (default)
 - Cuaca_Testing (default - v1, 10/15/19 9:40 PM)
 - Cuaca_Training (default - v1, 10/15/19 9:40 PM)
 - Data_NilaiUjian (default - v1, 11/12/19 7:18 PM)
 - Hasil_Cuaca (default - v1, 10/15/19 9:45 PM - 2)
 - Hasil_NilaiUjian (default - v1, 11/12/19 7:18 PM - 2)
 - Hasil_Sekolah (default - v1, 10/15/19 9:50 PM - 5)
 - HasilCuaca1 (default - v1, 10/15/19 9:50 PM - 5)
 - Sekolah_Testing (default - v1, 10/15/19 9:43 PM)
 - Sekolah_Training (default - v1, 10/15/19 9:41 PM)

Type here to search

7:23 PM 11/12/2019

d.

Cluster Mode (Clustering)

i. Description

new process* - RapidMiner Studio Educational 9.3.001 @ LAPTOP-01MEG9F7

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

ExampleSet (/Local Repository/Data_NilaiUjian) ExampleSet (/Local Repository/Data_NilaiUjian) ExampleSet (/Local Repository/Data_NilaiUjian)

Result History SVD (SVD) ExampleSet (Clustering) ExampleSet (SVD) Cluster Model (Clustering)

Cluster Model

Description

Cluster 0: 10 items
Cluster 1: 8 items
Cluster 2: 6 items
Cluster 3: 6 items
Total number of items: 30

Folder View

Graph

Centroid Table

Plot

Annotations

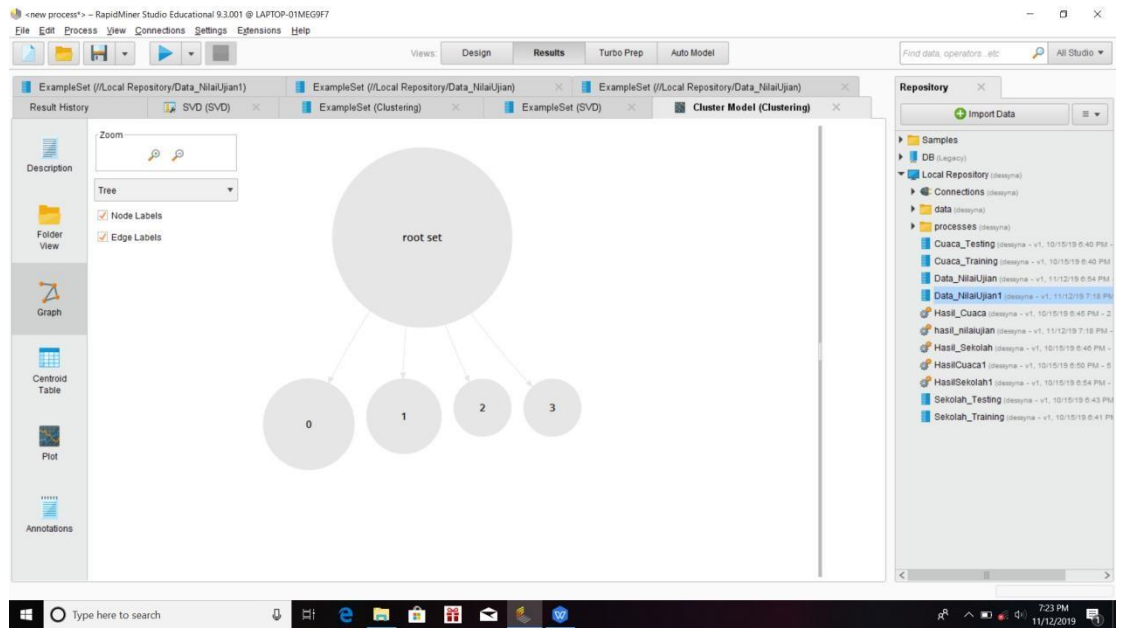
Repository

- Import Data
- Samples
- DB (Legacy)
- Local Repository (default)
 - Connections (default)
 - data (default)
 - processes (default)
 - Cuaca_Testing (default - v1, 10/15/19 9:40 PM)
 - Cuaca_Training (default - v1, 10/15/19 9:40 PM)
 - Data_NilaiUjian (default - v1, 11/12/19 7:18 PM)
 - Hasil_Cuaca (default - v1, 10/15/19 9:45 PM - 2)
 - Hasil_NilaiUjian (default - v1, 11/12/19 7:18 PM - 2)
 - Hasil_Sekolah (default - v1, 10/15/19 9:40 PM - 5)
 - HasilCuaca1 (default - v1, 10/15/19 9:50 PM - 5)
 - HasilSekolah1 (default - v1, 10/15/19 9:54 PM - 5)
 - Sekolah_Testing (default - v1, 10/15/19 9:43 PM)
 - Sekolah_Training (default - v1, 10/15/19 9:41 PM)

Type here to search

7:23 PM 11/12/2019

ii. Graph



3. Tulislah masing-masing nama siswa yang terdapat dalam kelompok Cluster 0, Cluster 1, Cluster 2, dan Cluster 3.

CLUSTER	NAMA
0	WATI
0	YANTO
0	WAWAN
0	MAHMUD
0	SANTI
0	DANI
0	BAYU
0	RANI
0	BAMBIANG
0	FEBRI
1	DIAN
1	RATIH
1	JONO
1	SARAH
1	HADI
1	NANA
1	DENI
1	TONI
2	SUSI
2	IKA
2	BUDI
2	AHMAD
2	RISA
2	INDAH
3	JOKO
3	AGUS

