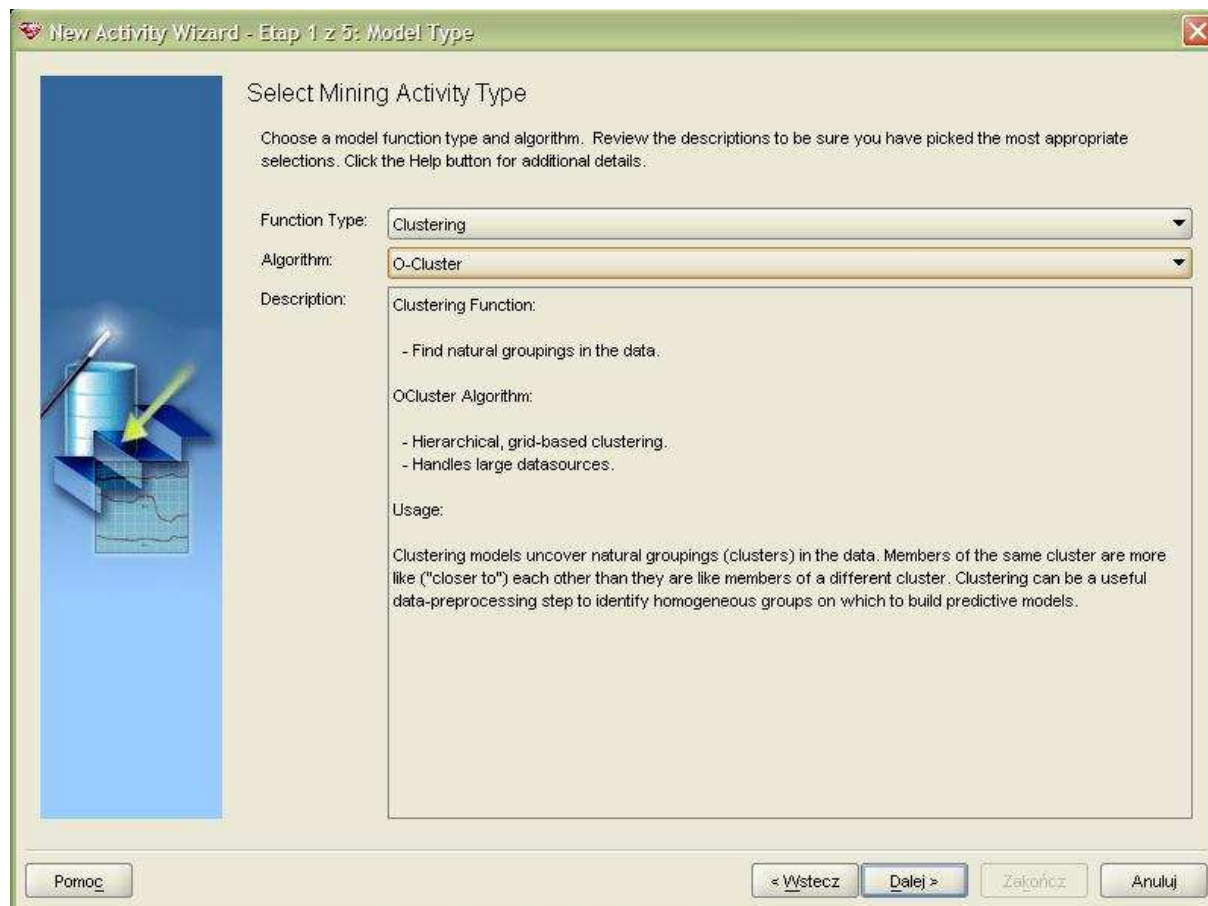


## Laboratorium 9

### Grupowanie z wykorzystaniem algorytmu O-Cluster.

1. Uruchom narzędzie Oracle Data Miner i połącz się z serwerem bazy danych.
2. Z menu głównego wybierz Activity→Build. Na ekranie powitalnym kliknij przycisk Dalej>.
3. Z listy Function Type wybierz Clustering. Rozwiń listę Algorithm i wybierz z niej algorytm O-Cluster. Kliknij przycisk Dalej>.



4. Wskaż schemat STUDENT i tabelę MINING\_DATA\_BUILD\_V jako źródło danych do eksploracji. Jako klucz podstawowy wskaż atrybut CUST\_ID. Kliknij przycisk Dalej>.

New Activity Wizard - Etap 2 z 4: Data

Select the Case Table

Select the table containing the "cases" (individual records/rows) that will be input to your mining activity. You can unselect any table columns that you know should not be considered as mining attributes. You can also join additional data in with the case table by selecting the checkbox below.

Schema:

Table/View:

Join additional data with case table

Unique Identifier:  Single Key:   
 Compound, or None

NOTE: Compound (multi-column), or absence of unique identifiers requires creation of a supporting table. This can take a significant amount of time and disk space.

Select Columns:

Select	Name	Data Type
<input checked="" type="checkbox"/>	AFFINITY_CARD	NUMBER
<input checked="" type="checkbox"/>	AGE	NUMBER
<input checked="" type="checkbox"/>	BOOKKEEPING_APPLICATION	NUMBER
<input checked="" type="checkbox"/>	BULK_PACK_DISKETTES	NUMBER
<input checked="" type="checkbox"/>	COUNTRY_NAME	VARCHAR2
<input checked="" type="checkbox"/>	CUST_GENDER	CHAR
<input checked="" type="checkbox"/>	CUST_ID	NUMBER
<input checked="" type="checkbox"/>	CUST_INCOME_LEVEL	VARCHAR2
<input checked="" type="checkbox"/>	CUST_MARITAL_STATUS	VARCHAR2
<input checked="" type="checkbox"/>	EDUCATION	VARCHAR2
<input checked="" type="checkbox"/>	PLAT_PAPER_MONITOR	NUMBER

[Sampling Settings...](#)

Pomoc

5. Podaj nazwę i krótki opis procesu eksploracji. Kliknij przycisk **Dalej>**.

New Activity Wizard - Etap 4 z 4: Activity Name

Activity Name

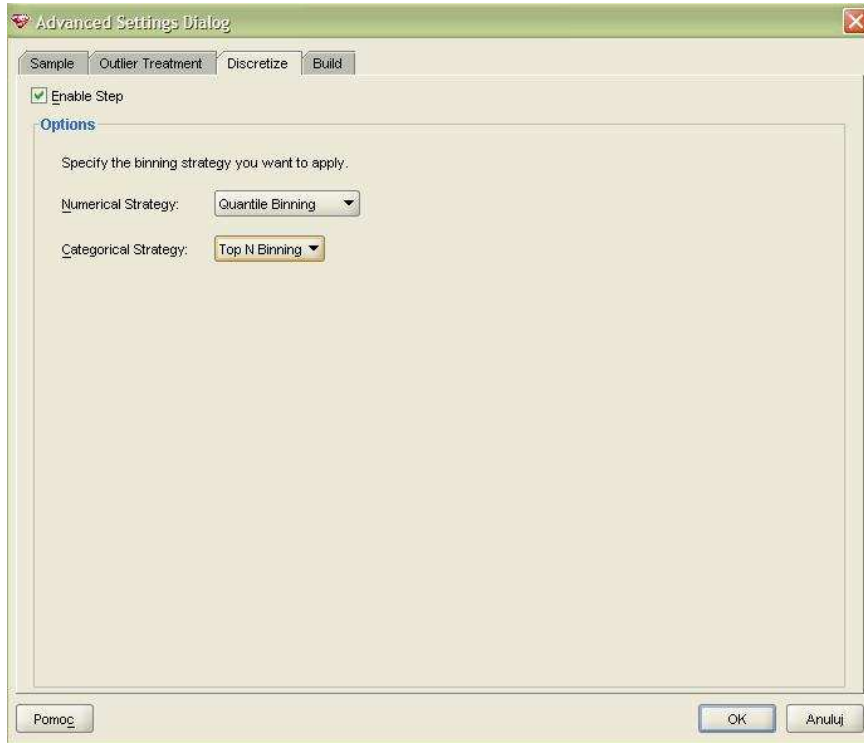
Enter the name for the new Mining Activity.

Name:

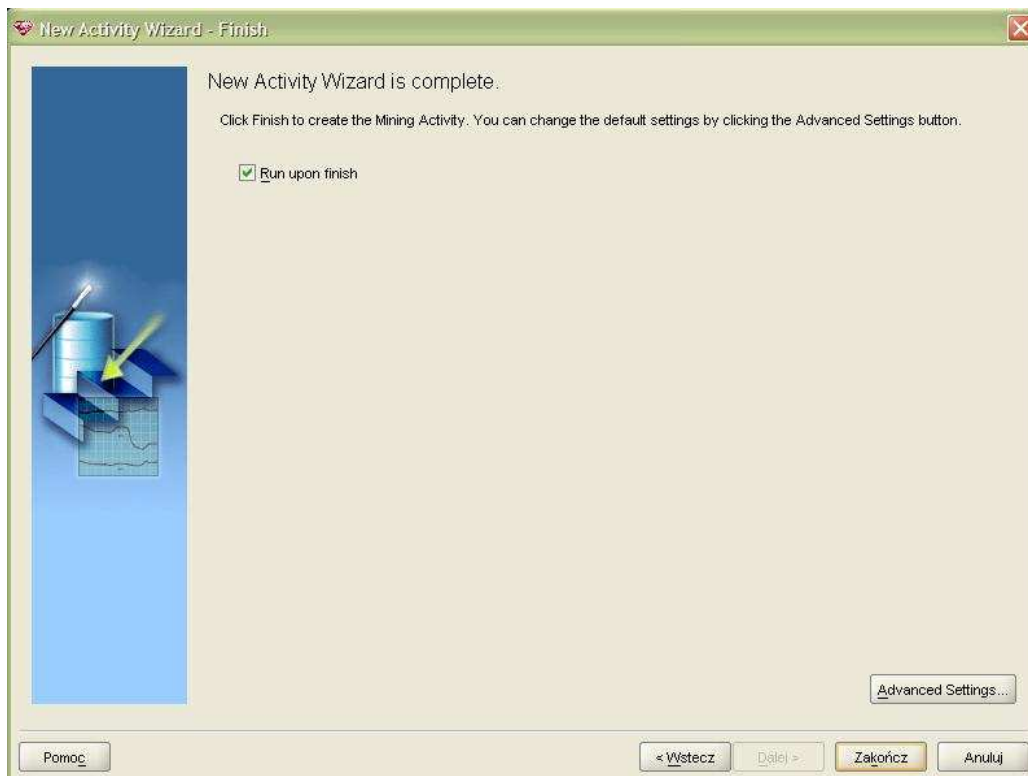
Comment:

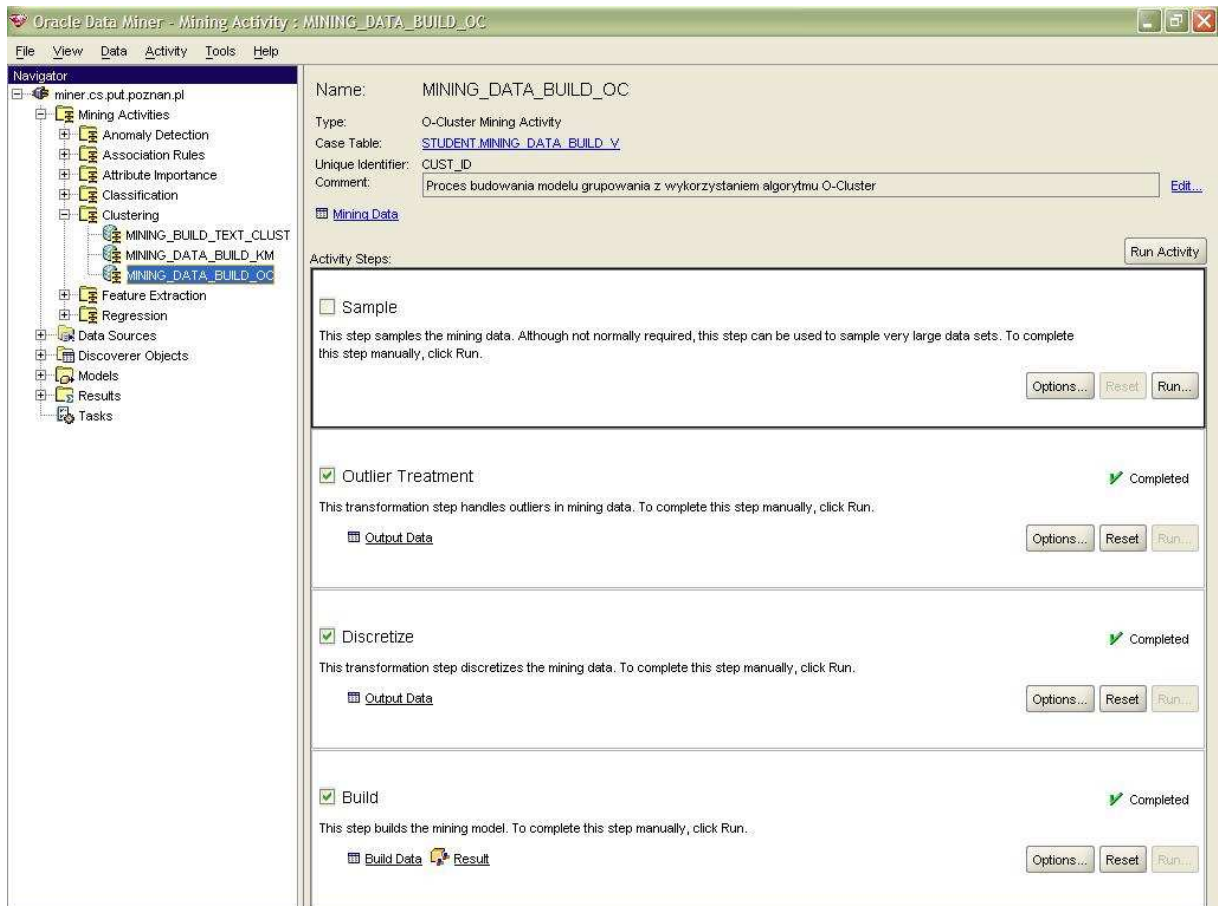
Pomoc

6. Kliknij przycisk **Advanced Settings**. Upewnij się, że na zakładce **Sample** opcja próbkowania jest wyłączona (pole wyboru **Enable Step** jest odznaczone). Algorytm O-Cluster wymaga dyskretyzacji, dlatego przejdź na zakładkę **Discretize**. Dla atrybutów numerycznych wybierz strategię **Quantile Binning**, natomiast dla atrybutów kategorycznych strategię **Top N Binning**. Kliknij przycisk **OK**.

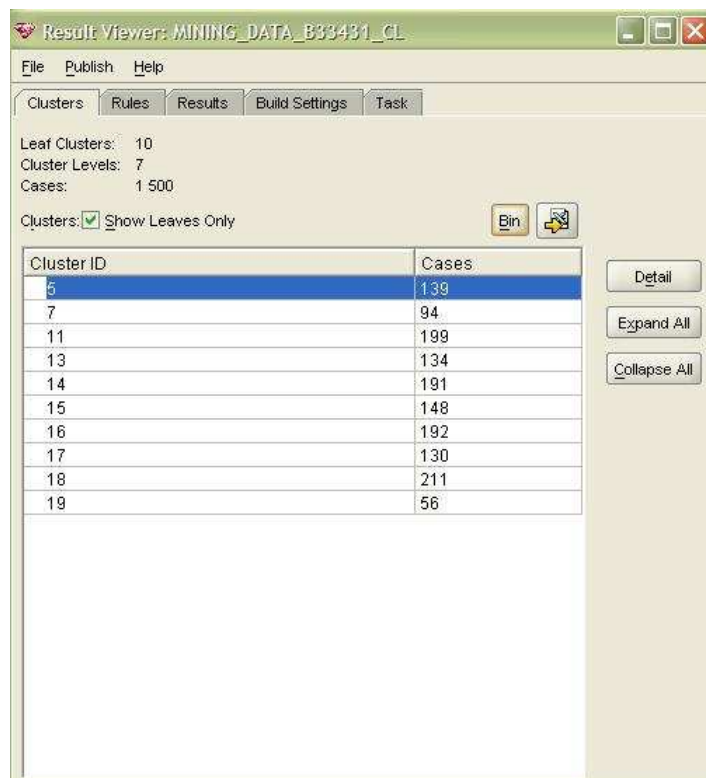


7. Upewnij się, że opcja **Run upon finish** jest włączona. Kliknij przycisk **Zakończ**.





8. Kliknij na odnośnik Result w bloku Build. Zaznacz opcję Show Leaves Only.



9. Przejdź na zakładkę Rules. Zaznacz opcję Only Show Rules for Leaf Clusters. Wybierz dowolny klastery i przeanalizuj atrybuty, które trafiają do wybranego klastra.

The screenshot shows the Oracle Data Miner Result Viewer interface. The 'Rules' tab is active, and the 'Only Show Rules for Leaf Clusters' checkbox is checked. A table lists clusters with their IDs, confidence percentages, and support counts. Cluster 16 is highlighted in blue. Below the table, the 'Rule Detail' section shows the logical rule for cluster 16, including the 'IF' condition, the 'THEN' result, and the confidence and support values.

Cluster ID	Confidence (%)	Support Count
5	81,2949640288	113
7	82,9787234043	78
11	75,3768844221	150
13	75,3731343284	101
14	81,6753926702	156
15	81,0810810811	120
16	85,9375	165
17	78,4615384615	102
18	84,8341232227	179
19	78,5714285714	44

**Rule Detail**

IF  
 AFFINITY\_CARD in (0,0,1,0) and AGE <= 79.0 and AGE >= 39.0 and BOOKKEEPING\_APPLICATION = 1.0 and BULK\_PACK\_DISKETTES in (0,0,1,0) and COUNTRY\_NAME equal (United States of America) and CUST\_GENDER in (F, M) and CUST\_INCOME\_LEVEL in (C: 50,000 - 69,999, D: 70,000 - 89,999, E: 90,000 - 109,999, F: 110,000 - 129,999, G: 130,000 - 149,999, H: 150,000 - 169,999, I: 170,000 - 189,999, J: 190,000 - 249,999, K: 250,000 - 299,999, L: 300,000 and above) and CUST\_MARITAL\_STATUS in (Divorc., Married, NeverM) and EDUCATION in (< Bach., 10th, 7th-8th, Assoc-A, Assoc-V, Bach., HS-grad, Masters, PhD, Profsc) and FLAT\_PANEL\_MONITOR in (0,0,1,0) and HOME\_THEATER\_PACKAGE = 1.0 and HOUSEHOLD\_SIZE in (2,0,3,0,9+) and OCCUPATION in (Machine, Other, Prof.) and OS\_DOC\_SET\_KANJI = 0.0 and YRS\_RESIDENCE <= 9.0 and YRS\_RESIDENCE >= 2.0 and Y\_BOX\_GAMES = 0.0

THEN  
 cluster equal 16

Confidence (%)=85.9375  
 Support =165

## Ćwiczenie samodzielne

Na podstawie tabeli PRACOWNICY zbuduj następującą perspektywę:

```
SQL> desc pracownicy_v
```

Name	Null?	Type
NAZWISKO		VARCHAR2 ( 15 )
ETAT		VARCHAR2 ( 10 )
ROK		NUMBER
PLACA_POD		NUMBER ( 6 , 2 )
PLACA_DOD		NUMBER ( 6 , 2 )
ID_ZESP		NUMBER ( 2 )
ID_SZEFA		NUMBER ( 4 )
ID_PRAC	NOT NULL	NUMBER ( 4 )

Utworzoną przez siebie perspektywę wykorzystaj do zbudowania modelu znajdującego  $n$  grup najbardziej podobnych do siebie pracowników (wartość  $n$  wyznacz eksperymentalnie). Do znalezienia grup zastosuj algorytm O-Cluster. Wyświetl informacje o uzyskanym modelu.

Aby zobaczyć działanie algorytmu, zastosuj zbudowany przez siebie model do zawartości perspektywy PRACOWNICY\_V. Zwróć uwagę, jak algorytm przydziela profesorów do grup.

Ostatecznym celem ćwiczenia jest napisanie zapytania, które wyświetli przypisanie pracowników do poszczególnych grup (wraz z prawdopodobieństwem przypisania pracownika do danej grupy). Postaraj się uzyskać podobny efekt:

NAZWISKO	ETAT	CLUSTER_ID	PROBABILITY
BLAZEWICZ	PROFESOR	4	.9972
SLOWINSKI	PROFESOR	4	.904
WEGLARZ	DYREKTOR	6	.9999
BRZEZINSKI	PROFESOR	7	.9343
MORZY	PROFESOR	7	.9254
KROLIKOWSKI	ADIUNKT	8	.6857
BIALY	STAZYSTA	8	.9286
ZAKRZEWICZ	STAZYSTA	8	.9791
KOSZLAJDA	ADIUNKT	8	.8453
JEZIERSKI	ASYSTENT	8	.7038
MATYSIAK	ASYSTENT	8	.904
MAREK	SEKRETARKA	9	.9575
KONOPKA	ASYSTENT	9	.9036
HAPKE	ASYSTENT	9	.598

Uwaga:

Pamiętaj, aby dane przetwarzane przez algorytm O-Cluster poddać dyskretyzacji!