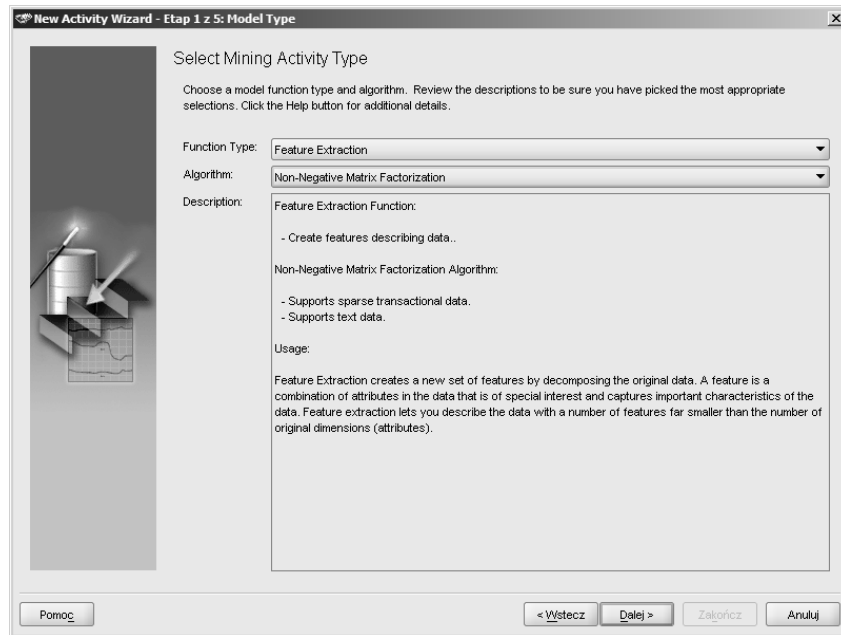


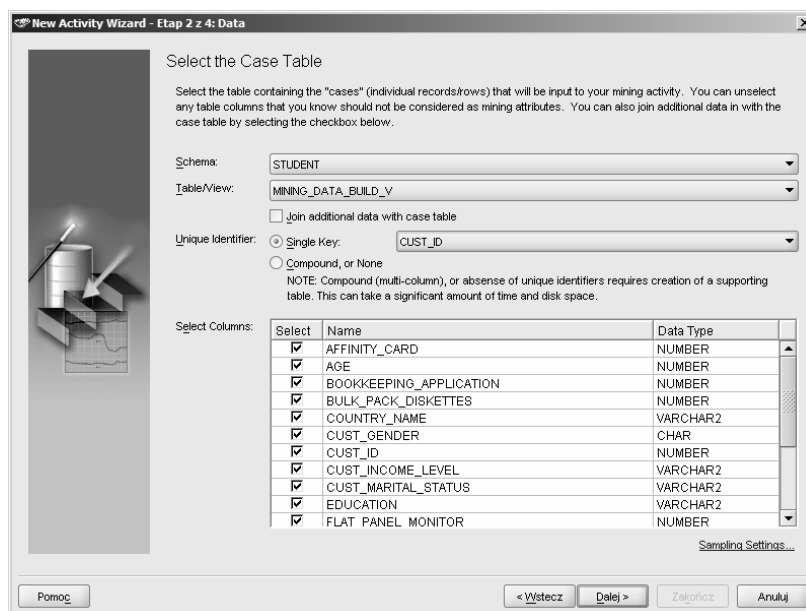
Laboratorium 10

Odkrywanie cech i algorytm Non-Negative Matrix Factorization.

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 Feature Extraction. Rozwiń listę Algorithm i wybierz z niej algorytm Non-Negative Matrix Factorization. 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>.



5. W kolejnym kroku pozostaw domyślne ustawienia atrybutów. Kliknij przycisk **Dalej**>.

New Activity Wizard - Etap 3 z 4: Data Usage

Review Data Usage Settings

Review the column settings. You can change the column settings to better match your understanding of the data. The default settings have been determined for each column based on the activity type and the characteristics of the data.

[Data Summary](#)

Name	Alias	Input	Data Type	Mining Type	Spar...
<input type="checkbox"/> STUDENT_MINING_DATA_B...					
AFFINITY_CARD	AFFINITY_CARD	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
AGE	AGE	<input checked="" type="checkbox"/>	NUMBER	numerical	<input type="checkbox"/>
BOOKKEEPING_APPLIC...	BOOKKEEPING_APPLIC...	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
BULK_PACK_DISKETTES	BULK_PACK_DISKETTES	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
COUNTRY_NAME	COUNTRY_NAME	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
CUST_GENDER	CUST_GENDER	<input checked="" type="checkbox"/>	CHAR	categorical	<input type="checkbox"/>
CUST_ID	CUST_ID	<input type="checkbox"/>	NUMBER	numerical	<input type="checkbox"/>
CUST_INCOME_LEVEL	CUST_INCOME_LEVEL	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
CUST_MARITAL_STATUS	CUST_MARITAL_STATUS	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
EDUCATION	EDUCATION	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
FLAT_PANEL_MONITOR	FLAT_PANEL_MONITOR	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
HOME_THEATER_PACK...	HOME_THEATER_PACK...	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
HOUSEHOLD_SIZE	HOUSEHOLD_SIZE	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
OCCUPATION	OCCUPATION	<input checked="" type="checkbox"/>	VARCHAR2	categorical	<input type="checkbox"/>
OS_DOC_SET_KANJI	OS_DOC_SET_KANJI	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
PRINTER_SUPPLIES	PRINTER_SUPPLIES	<input type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
Y_BOX_GAMES	Y_BOX_GAMES	<input checked="" type="checkbox"/>	NUMBER	categorical	<input type="checkbox"/>
YRS_RESIDENCE	YRS_RESIDENCE	<input checked="" type="checkbox"/>	NUMBER	numerical	<input type="checkbox"/>

Pomoc < Wstecz Dalej > Zakończ Anuluj

6. W kolejnym kroku podaj nazwę dla procesu eksploracji oraz krótki opis procesu eksploracji. Kliknij przycisk **Dalej**>.

New Activity Wizard - Etap 4 z 4: Activity Name

Activity Name

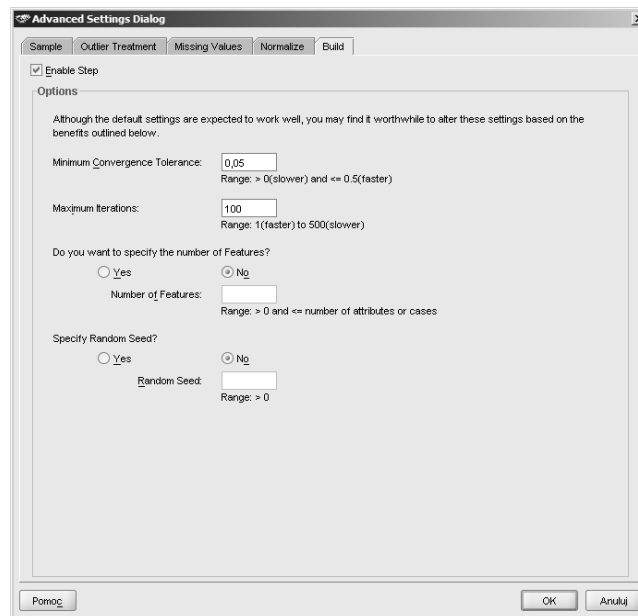
Enter the name for the new Mining Activity.

Name:

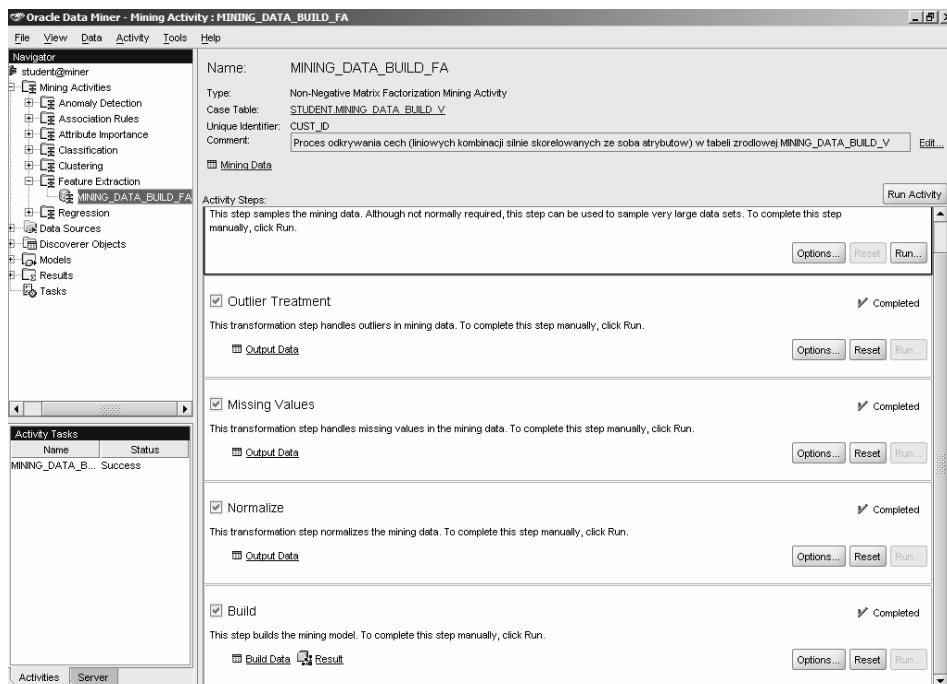
Comment:

Pomoc < Wstecz Dalej > Zakończ Anuluj

- Kliknij przycisk **Advanced Settings**. Upewnij się, że na zakładce **Sample** opcja próbkowania jest wyłączona (pole wyboru **Enable Step** jest odznaczone). Przejdź na zakładkę **Outlier Treatment**. Upewnij się, czy znajdowanie osobliwości jest włączone (możesz pozostawić domyślne kryterium odległości trzech odchyłeń standardowych od średniej jako wyznacznika osobliwości). Przejdź na zakładkę **Missing Values** i upewnij się, że brakujące wartości będą automatycznie zamieniane na średnią (dla atrybutów numerycznych) lub wartość modalną (dla atrybutów kategoriycznych). Przejdź na zakładkę **Normalize** i upewnij się, że atrybuty numeryczne będą automatycznie normalizowane do przedziału 0-1. Przejdź na zakładkę **Build** i zmień maksymalną liczbę iteracji przy wyliczaniu macierzy na 100.



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



9. Kliknij na odnośnik **Result** w bloku **Build**. Z listy rozwijanej wybierz cechę numer 6. Na każdą cechę składa się liniowa kombinacja wartości oryginalnych atrybutów. Im większa wartość współczynnika przy danej wartości atrybutu, tym bardziej dana wartość atrybutu wpływa na cechę. Przykładowo, cecha numer 6 reprezentuje przede wszystkim mężczyzn mieszkających w Stanach Zjednoczonych posiadających monitor z płaskim ekranem i nie posiadających karty lojalnościowej. Osoby pasujące do tej charakterystyki będą posiadać wysoką wartość liczbową cechy numer 6.

Attribute Name	Value	Coefficient
COUNTRY_NAME	United States of America	0,3163391661
CUST_GENDER	M	0,2999607549
BOOKKEEPING_APPLICATION	1	0,2857431082
AFFINITY_CARD	0	0,2787278199
FLAT_PANEL_MONITOR	1	0,1665125107
OS_DOC_SET_KANJI	0	0,1617860266
HOME_THEATER_PACKAGE	0	0,1526723875
CUST_GENDER	F	0,1261172239
AGE		0,1197151097
EDUCATION	HS-grad	0,0885863326
Y_BOX_GAMES	0	0,0805394104
YRS_RESIDENCE		0,0803186071
HOUSEHOLD_SIZE	3	0,0754248904
HOME_THEATER_PACKAGE	1	0,0735655537
CUST_MARITAL_STATUS	Married	0,0733516295
HOUSEHOLD_SIZE	2	0,0600809052
BULK_PACK_DISKETTES	0	0,0598793864
EDUCATION	Bach.	0,0469408646
AFFINITY_CARD	1	0,0465607340
OCCUPATION	Exec.	0,0464067811
OCCUPATION	Cleric.	0,0463271834
CUST_MARITAL_STATUS	Divor	0,0422104566

10. Kliknij przycisk **Filter**. Oznacz pole wyboru **Check All**. Wskaż jako interesujący Cię atrybut **CUST_MARITAL_STATUS**. Kliknij przycisk **OK**.

Attributes:

Check All

Include	Attribute Name
<input type="checkbox"/>	AFFINITY_CARD
<input type="checkbox"/>	AGE
<input type="checkbox"/>	BOOKKEEPING_APPLICATION
<input type="checkbox"/>	BULK_PACK_DISKETTES
<input type="checkbox"/>	COUNTRY_NAME
<input type="checkbox"/>	CUST_GENDER
<input type="checkbox"/>	CUST_INCOME_LEVEL
<input checked="" type="checkbox"/>	CUST_MARITAL_STATUS
<input type="checkbox"/>	EDUCATION
<input type="checkbox"/>	FLAT_PANEL_MONITOR
<input type="checkbox"/>	HOME_THEATER_PACKAGE
<input type="checkbox"/>	HOUSEHOLD_SIZE
<input type="checkbox"/>	OCCUPATION
<input type="checkbox"/>	OS_DOC_SET_KANJI
<input type="checkbox"/>	Y_BOX_GAMES
<input type="checkbox"/>	YRS_RESIDENCE

Filter by Coefficient where the value is

between (min. value):

and (max. value):

Filter by absolute values

Sort By:

Selected Attributes

Include in the list

Exclude from the list

Pomoc OK Anuluj

11. Znajdź cechę opisującą osoby, które nigdy nie wyszły za mąż/ożeniły się.

Result Viewer: MINING_DATA_B26645_NF

Features: 4

Fetch Size: 100

Attribute Name	Value	Coefficient
CUST_MARITAL_STATUS	NeverM	0,1221569950
CUST_MARITAL_STATUS	Married	0,0724813938
CUST_MARITAL_STATUS	Divorc.	0,0406133203
CUST_MARITAL_STATUS	Separ.	0,0081798385
CUST_MARITAL_STATUS	Mabsent	0,0028687549
CUST_MARITAL_STATUS	Widowed	0,0018179329
CUST_MARITAL_STATUS	Mar-AF	0,0003130304

12. W nawigatorze obiektów rozwiń gałąź Models, następnie gałąź Feature Extraction i sprawdź, jak nazywa się zbudowany przez Ciebie model.

Oracle Data Miner : Feature Extraction

File View Data Activity Tools Help

Navigator

Name	Type	Activity Name	Function	Algorithm	Creation Date	Size(MB)	Build Duration	Server Task...
MINING_DATA_B26645_NF	Mining Model	MINING_DAT...FEATURE_E...	NONNEGATI...		2006-08-18	0,052	2	DM4.\$MININ...

Activity Tasks

Name	Status
MINING_DATA_BUILD_FA	Success

Activities Server

13. Uruchom narzędzie iSQLPlus i zaloguj się do bazy danych. Odczytaj podstawowe informacje o modelu z repozytorium ODM.

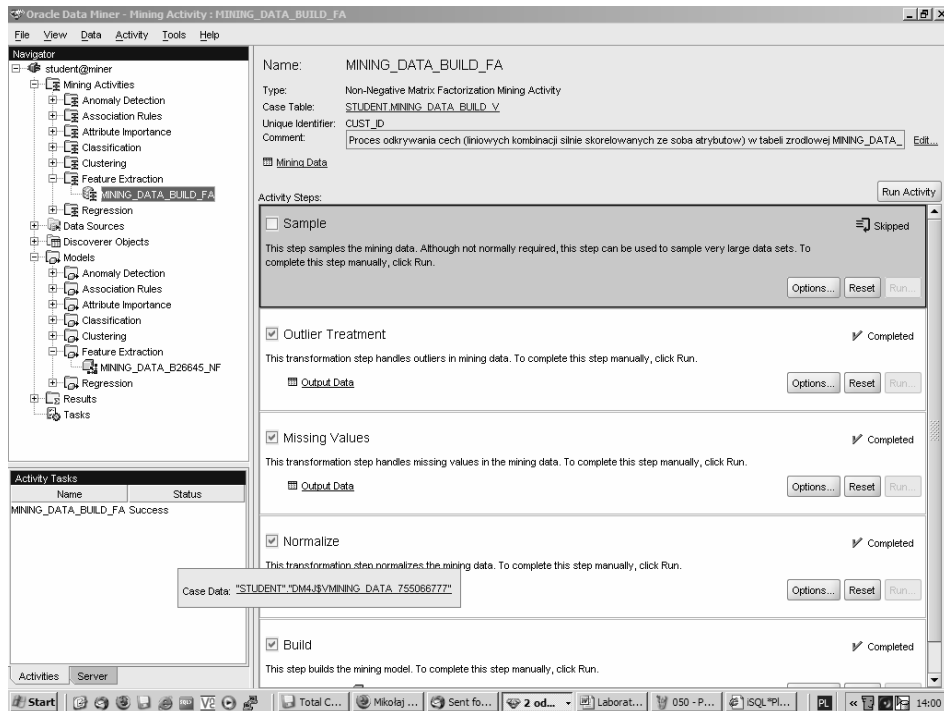
```
SELECT setting_name, setting_value
FROM TABLE(DBMS_DATA_MINING.GET_MODEL_SETTINGS('MINING_DATA_B26645_NF'))
ORDER BY setting_name;
```

SETTING_NAME	SETTING_VALUE
ALGO_NAME	ALGO_NONNEGATIVE_MATRIX_FACTOR
FEAT_NUM_FEATURES	9
JDMS_FUNCTION_TYPE	FEATURE_EXTRACTION
NMFS_CONV_TOLERANCE	0.05
NMFS_NUM_ITERATIONS	100
NMFS_RANDOM_SEED	-1

14. Do wyświetlenia pełnego modelu (wszystkich cech i wszystkich składowych cech) posłuż się poniższym zapytaniem

```
SELECT F.feature_id, A.attribute_name, A.attribute_value, A.coefficient
FROM
TABLE(DBMS_DATA_MINING.GET_MODEL_DETAILS_NMF('MINING_DATA_B26645_NF')) F,
TABLE(F.attribute_set) A
ORDER BY feature_id, attribute_name, attribute_value;
```

15. Wróć do narzędzia Oracle Data Mining. W nawigatory obiektów rozwiń gałąź Mining Activities, następnie gałąź Feature Extraction. Kliknij odnośnik Output Data w bloku Normalize i sprawdź, jak nazywa się automatycznie skonstruowana perspektywa zawierająca przetransformowane dane.

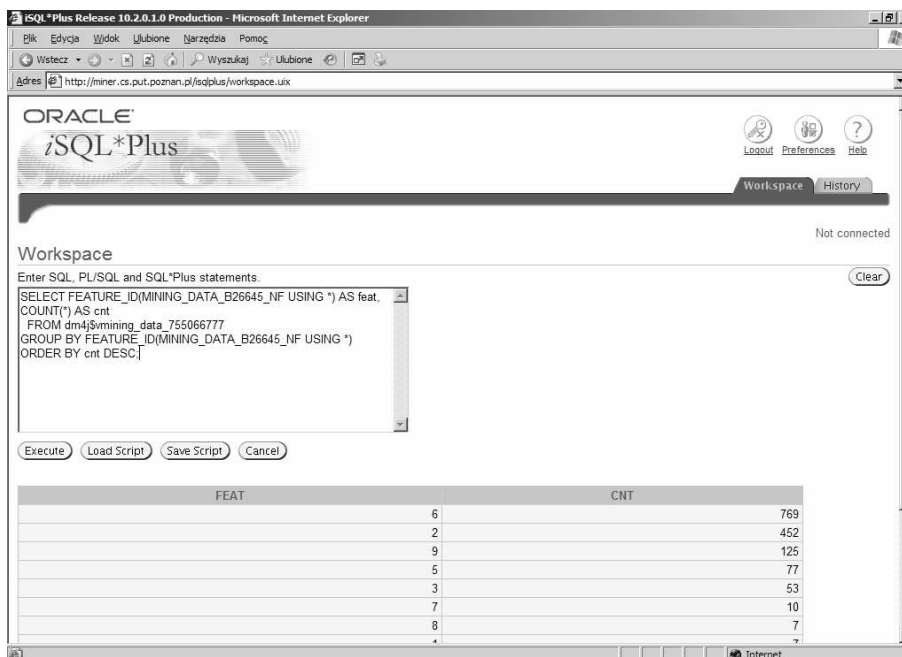


16. Wróć do narzędzia iSQLPlus. Wyświetl liczbę klientów opisanych poszczególnymi cechami. Wykorzystaj poniższe zapytanie SQL. FEATURE_ID() to funkcja zwracająca identyfikator cechy zgodnie z podanym modelem. Klauzula USING umożliwia ograniczenie modelu tylko do podanej listy atrybutów.

```

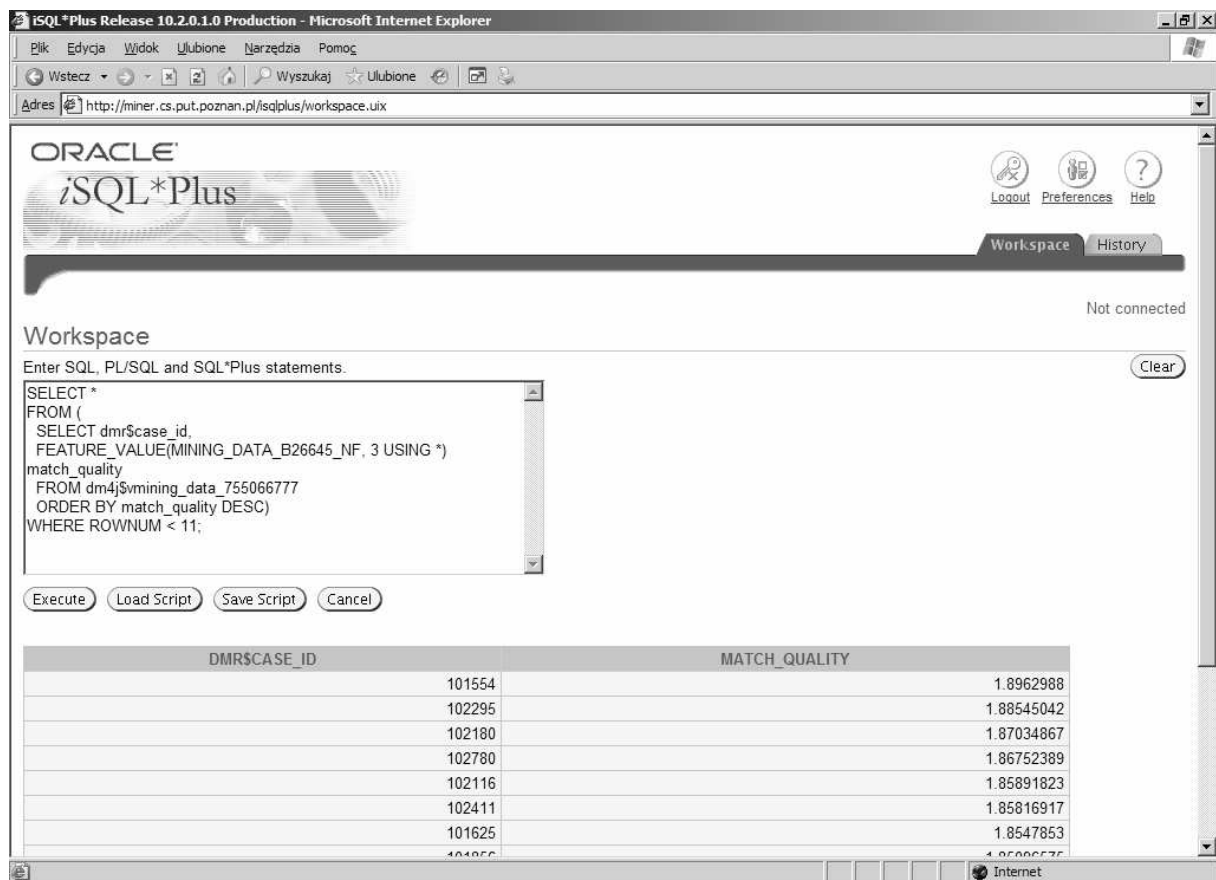
SELECT FEATURE_ID(MINING_DATA_B26645_NF USING *) AS feat,
       COUNT(*) AS cnt
FROM dm4j$vmining_data_755066777
GROUP BY FEATURE_ID(MINING_DATA_B26645_NF USING *)
ORDER BY cnt DESC;

```



17. Wyświetl dziesięciu klientów którzy są najlepiej dopasowani do cechy numer 3. Wykorzystaj poniższe zapytanie SQL.

```
SELECT *
FROM (
  SELECT dmr$case_id,
         FEATURE_VALUE(MINING_DATA_B26645_NF, 3 USING *) AS match_quality
  FROM dm4j$vmining_data_755066777
  ORDER BY match_quality DESC)
WHERE ROWNUM < 11;
```



The screenshot shows the iSQL*Plus web interface in a Microsoft Internet Explorer browser. The browser title is "iSQL*Plus Release 10.2.0.1.0 Production - Microsoft Internet Explorer". The address bar shows "http://miner.cs.put.poznan.pl/isqlplus/workspace.uix". The page header includes the Oracle logo and "iSQL*Plus" text. There are navigation links for "Logout", "Preferences", and "Help". Below the header, there are tabs for "Workspace" and "History". The main content area is titled "Workspace" and contains a text input field with the following SQL query:

```
SELECT *
FROM (
  SELECT dmr$case_id,
         FEATURE_VALUE(MINING_DATA_B26645_NF, 3 USING *)
         match_quality
  FROM dm4j$vmining_data_755066777
  ORDER BY match_quality DESC)
WHERE ROWNUM < 11;
```

Below the query input field are buttons for "Execute", "Load Script", "Save Script", and "Cancel". To the right of the input field is a "Clear" button. Below the buttons is a table with two columns: "DMR\$CASE_ID" and "MATCH_QUALITY". The table contains 10 rows of data, sorted by match quality in descending order.

DMR\$CASE_ID	MATCH_QUALITY
101554	1.8962988
102295	1.88545042
102180	1.87034867
102780	1.86752389
102116	1.85891823
102411	1.85816917
101625	1.8547853
101856	1.85006777

Ćwiczenie samodzielne

W Twoim schemacie znajduje się tabela zoo o następującym schemacie:

```
SQL> desc zoo
```

Nazwa	Wartość NULL?	Typ
NAME		VARCHAR2(10)
HAIR		CHAR(1)
FEATHERS		CHAR(1)
EGGS		CHAR(1)
MILK		CHAR(1)
AIRBORNE		CHAR(1)
AQUATIC		CHAR(1)
PREDATOR		CHAR(1)
TOOTHED		CHAR(1)
BACKBONE		CHAR(1)
BREATHES		CHAR(1)
VENOMOUS		CHAR(1)
FINS		CHAR(1)
LEGS		NUMBER
TAIL		CHAR(1)
DOMESTIC		CHAR(1)
CATSIZE		CHAR(1)
TYPE		VARCHAR2(12)

Zbuduj model identyfikujący dominujące cechy w zbiorze zwierząt. Wyświetl uzyskane cechy i spróbuj dokonać ich ręcznej ewaluacji. Następnie zastosuj uzyskany model do zawartości tabeli zoo. Napisz następujące zapytania SQL:

- wyświetl 3 dominujące cechy opisujące lwa
- wyświetl 10 zwierząt najlepiej opisanych przez cechę reprezentującą zwierzęta jadowite
- wyświetl 10 zwierząt opisywanych przez cechę wyrażającą zdolność do latania
- wyświetl rozkład cech opisujących ptaki

Do wykonania zadania możesz wykorzystać przykładowy skrypt nmmf.plsql