

Przygotowanie do pracy zespołowej

Zespół ustala jedno konto sieciowe, na którym składuje zasoby będące wynikiem pracy. Członkowie zespołu mogą się na to jedno konto logować i pracować na dwóch komputerach, np. na jednym realizować projekt a na drugim redagować raport.

Uwagi

1. W nazwach zasobów pojawiających się w instrukcjach w miejsce xx należy wpisać numer konta/stanowiska, na którym jest realizowane zadanie (ten numer jest stały do końca semestru).
2. Szkielety raportów z badań są zapisane w ogólnie dostępnym katalogu sieciowym nauczyciela pod nazwą korelująca z tematyką ćwiczeń (np. *Asocjacje_rap.docx*). Te dokumenty należy umieszczać sukcesywnie (w miarę postępu ćwiczeń) w swoim folderze z raportami i zmieniać im nazwę tak, aby informowały na jakim koncie powstały, np. *Asocjacje_rap_02.docx* informuje, że zespół trzyma swoje zasoby na dysku sieciowym na koncie *mnd_s2*.
3. W trakcie pracy należy uzupełniać raport posługując się wynikami z poszczególnych węzłów diagramu projektu w SAS® EM™ lub wynikami uzyskanymi z programów w Pythonie.

Analiza asocjacji – identyfikacja preferencji zakupowych klienta marketu spożywczego

Raport z badania ma co najwyżej 2 strony.

Ćwiczenie 1

Duży sklep spożywczy zamawia analizę danych dotyczących zarejestrowanych zakupów swoich klientów w celu wykrycia zależności jakie grupy produktów łącznie klient zazwyczaj kupuje. W ten sposób sklep chce określić preferencje zakupowe swoich klientów. Jednym z celów jest taka ekspozycja towarów, aby sprowokować klienta do tzw. transakcji wiązanej. Dzięki temu istnieje szansa na zwiększenie obrotów sklepu. Wykonać analizę koszykową danych o zakupach klientów sklepu na podstawie informacji zawartej w pliku *assoc.xlsx*. Opracować raport z wykonanych badań.

Wykonanie

Diagram projektu zbudowany w środowisku EM zawiera cztery węzły realizujące zadania: dostarczenie danych, zastąpienie wartości elementów koszyka, pobranie danych transakcyjnych, tworzenie reguł asocjacyjnych oraz węzdo wstępnej eksploracji danych. W trakcie budowy diagramu należy łączyć ze sobą węzły, wskazując kierunek przepływu informacji z węzła do węzła. Zaznaczony węzeł można obsługiwać wykorzystując menu podręczne lub panel właściwości.

Uwaga. Diagram składa się z dwóch *work flow*, ponieważ SAS EM nie obsługuje węzła *Zastępowanie* dla importowanych danych transakcyjnych.

- Pozycja menu *Run* uruchamia przetwarzania danych w węźle zgodnie z zadanymi parametrami.

Uruchomić EM z menu systemu SAS. Utworzyć nowy projekt. W oknie tworzenia projektu podać nazwę *Asocjacje_xx*. Wskazać właściwy folder zapisu projektu. Nadać diagramowi nazwę *Koszyki*.

Dane do diagramu

1. Dane do realizacji zadania są importowane ze skoroszytu Excela z arkusza wskazanego przez nauczyciela. Wprowadzić do obszaru projektowego EM węzeł importu danych. Ustawić rolę zbioru danych na *Uczqce*. Uruchomić węzeł. Sprawdzić wyniki (porównać dane z zawartością pliku źródłowego). Jak EM importuje dane ze skoroszytu wieloarkuszowego?

Modyfikacja wartości zmiennej

1. Wprowadzić do diagramu węzeł *Zastępowanie*. Połączyć z tym węzłem węzeł danych.
2. W panelu właściwości w pozycji *Edytor zastąpień* grupy *Zmienne klasyfikujące* wprowadzić polskie odpowiedniki towarów oferowanych w sklepie, zgodnie z podaną tabelą.
3. Uruchomić węzeł i sprawdzić wyniki.

Zmienna	wartość sformatowana	Wartość zastępująca
PRODUCT	apples	jabłka
PRODUCT	artichok	karczoch
PRODUCT	avocado	awokado
PRODUCT	baguette	bagietka
PRODUCT	bordeaux	
PRODUCT	bourbon	
PRODUCT	chicken	kurczak
PRODUCT	coke	kola
PRODUCT	corned_b	chleb kukurydziany
PRODUCT	cracker	krakersy
PRODUCT	ham	szynka
PRODUCT	heineken	
PRODUCT	hering	śledzie
PRODUCT	ice_crea	lody
PRODUCT	olives	olowki
PRODUCT	peppers	papryczki
PRODUCT	sardines	sardynki
PRODUCT	soda	soda oczyszczona
PRODUCT	steak	steaki
PRODUCT	turkey	

Wstępna eksploracja danych

1. Zbadać ilu jest klientów i ile jest różnych towarów zarejestrowanych w zbiorze danych. Tę informację uzyskuje się określając liczbę poziomów zmiennych nominalnych.

Uwaga. SAS EM ma limit wyświetlania liczby poziomów do 512. Jeżeli zmienna ma więcej unikatowych wartości należy przededefiniować odpowiednią makrozmiennej. Kliknąć na nazwę projektu i z panelu właściwości wybrać opcję *Makrozmienne projektu*. Nadać makrozmiennej *EM_TRAIN_MAXLEVELS* wartość 5000. Zamknąć projekt, po czym otworzyć go ponownie.

2. Umieścić w obszarze projektu węzeł *Eksploracja statystyk*. Skierować do niego połączenie z węzła *Zastępowanie*. Uruchomić węzeł. Otworzyć z panelu właściwości okno *Eksportowane dane*. Zaznaczyć zbiór treningowy i obejrzeć plik (przycisk *Przeglądaj*).
3. Otworzyć wyniki węzła *Eksploracja statystyk*; menu podręczne i *Rezultaty*. Przewinąć okno *Wyniki* i poszukać zestawu statystyk opisujących zbiór. Odnaleźć informację o częstości wystąpienia poszczególnych towarów uruchamiając menu *Widok* → *Statystyki opisowe* → *Zmienne klasyfikujące*.

Wyznaczenie reguł asocjacyjnych

1. Sprawdzić jaką nazwę nosi plik eksportowany przez węzeł *Zastępowanie*. Ta nazwa jest potrzebna do wyznaczenia reguł asocjacyjnych. *przerwane*
2. Zdefiniować drugi *work flow*. Pierwszym jest węzeł *Dane wejściowe*. Ustawić właściwości węzła następująco.
 - *Wybór danych*: *Nowa tabela*.
 - *Nazwa tabeli*: wskazać zbiór danych będący wynikiem węzła zastępowanie z pierwszego *work flow*. Jakiego typu jest zbiór pobieranych danych?
 - *Rola*: *Transakcyjne*.
3. Określić właściwości zmiennych *Klient* i *REP_Produkt*: typ (*Poziom*) nominalnej i role odpowiednio *Id* i *Zmienna celu*. Odrzucić pozostałe zmienne.
4. Wprowadzić węzeł *Koszyk zakupów*. Sprawdzić domyślne ustawienia węzła i uruchomić go.
5. Reguły asocjacyjne oraz wartości ich miar są dostępne w oknie rezultatów z paska narzędzie i przycisku *Tabela*. Przeglądać pozostałe wyniki i zdecydować, które dodać do raportu.
6. Przeprowadzić w raporcie krótką analizę wyników.

Ćwiczenie 2

We właściwościach węzła *Koszyk zakupów* dobrać nowe ustawienia parametrów, tak aby można było zidentyfikować silne reguły związków trójelementowych. Wykorzystać grupę właściwości *Więzy*. Wykonać analizę otrzymanych reguł.

Na zakończenie, na podstawie obu eksperymentów, sformułować w raporcie wniosek końcowy określając preferencje zakupowe klienta hipermarketu spożywczego.