

*IBM SPSS Modeler Entity Analytics  
18.0 — podręcznik użytkownika*

**IBM**

**Uwaga**

Przed skorzystaniem z niniejszych informacji oraz produktu, którego one dotyczą, należy zapoznać się z informacjami zamieszczonymi w sekcji “Uwagi” na stronie 55.

**Informacje o produkcie**

Niniejsze wydanie publikacji dotyczy wersji 18, wydania 0, modyfikacji 0 produktu IBM SPSS Modeler oraz wszystkich następnych wersji i modyfikacji do czasu, aż w kolejnym wydaniu publikacji zostanie zawarta informacja o stosownej zmianie.

# Spis treści

<b>Przedmowa</b> . . . . .	<b>v</b>	Porównywanie nowych obserwacji w oparciu o repozytorium (węzeł przesyłania strumieniowego EA) . . . . .	25
<b>Rozdział 1. Analiza jednostek.</b> . . . . .	<b>1</b>	Mapowanie zmiennych wejściowych do cech (węzeł przesyłania strumieniowego EA) . . . . .	26
Informacje dotyczące analizy jednostek . . . . .	1	Wyświetlanie mapowań zmiennych i źródeł danych (węzeł przesyłania strumieniowego EA) . . . . .	27
Analiza jednostek i analiza predykcyjna . . . . .	2	Wyniki z węzła przesyłania strumieniowego EA . . . . .	28
<b>Rozdział 2. Analiza jednostek za pomocą programu IBM SPSS Modeler.</b> . . . . .	<b>5</b>	Korzystanie z programu IBM SPSS Modeler Entity Analytics z innymi produktami IBM SPSS . . . . .	28
Korzystanie z analizy jednostek w programie IBM SPSS Modeler . . . . .	5	Zadania administracyjne . . . . .	29
Etap 1: Wczytywanie danych źródłowych do programu SPSS Modeler . . . . .	6	Konfiguracja przypisań portów . . . . .	29
Etap 2: Tworzenie repozytorium . . . . .	6	Zarządzanie danymi uwierzytelniającymi administratora dla bazy danych repozytorium . . . . .	30
Etap 3: Łączenie programu SPSS Modeler z repozytorium. . . . .	7	Przenoszenie repozytorium do innego katalogu składowania . . . . .	30
Etap 4: Mapowanie zmiennych wejściowych do cech repozytorium. . . . .	7	Ustawianie właściwości strumienia dla zmiennych daty/czasu i znacznika czasu . . . . .	31
Etap 5: Eksport danych do repozytorium i rozdzielanie tożsamości . . . . .	7	Dostosowywanie ustawień limitu czasu . . . . .	31
Etap 6: Analiza rozdzielonych tożsamości . . . . .	9	Uruchomienie programu IBM SPSS Modeler Entity Analytics za pomocą klienta SPSS Modeler i SPSS Modeler Server w tym samym systemie Windows . . . . .	31
Etap 7: Rozdzielanie nowych obserwacji w oparciu o repozytorium. . . . .	9	Czyszczenie repozytorium jednostek . . . . .	32
Etap 8: Generowanie alertów . . . . .	10	Usuwanie nieużywanych źródeł danych z repozytorium . . . . .	32
		Usuwanie repozytorium jednostek . . . . .	32
		Usuwanie repozytorium bez połączenia . . . . .	33
<b>Rozdział 3. Zadania w zakresie analizy jednostek</b> . . . . .	<b>11</b>	<b>Rozdział 4. Przebieg analizy jednostek</b> . . . . .	<b>35</b>
Informacje o zadaniach . . . . .	11	Informacje o tym przykładzie . . . . .	35
Ustawianie repozytorium jednostek (węzeł Eksport EA) . . . . .	11	Pierwotny model . . . . .	35
Czyszczenie repozytorium jednostek . . . . .	11	Dodawanie analizy jednostek . . . . .	38
Łączenie ze źródłem danych . . . . .	12	Przeniesienie danych źródłowych do repozytorium . . . . .	38
Tworzenie repozytorium . . . . .	12	Odczytywanie rozdzielonych tożsamości . . . . .	39
Mapowanie zmiennych wejściowych do cech (węzeł Eksport EA) . . . . .	14	Porównywanie wyników analizy jednostek z pierwotnym modelem . . . . .	45
Wyświetlanie mapowań zmiennych (węzeł Eksport EA) . . . . .	15	Podsumowanie . . . . .	49
Konfigurowanie repozytorium jednostek . . . . .	16	<b>Dodatek. Właściwości skryptów dla programu IBM SPSS Modeler Entity Analytics.</b> . . . . .	<b>51</b>
Wyświetlanie zmapowanych źródeł danych . . . . .	17	Skrypty w IBM SPSS Modeler Entity Analytics . . . . .	51
Opracowywanie cech repozytorium . . . . .	17	Wspólne właściwości . . . . .	51
Dodawanie lub edycja cech . . . . .	18	Właściwości entityanalytics_exportnode . . . . .	51
Anonimizacja cech repozytorium . . . . .	19	Właściwości entityanalytics_sourcenode . . . . .	52
Opracowywanie typów jednostek . . . . .	20	Właściwości entityanalytics_processnode . . . . .	52
Ustawianie progu dopasowania jednostek . . . . .	22	<b>Uwagi</b> . . . . .	<b>55</b>
Ponowne użycie konfiguracji repozytorium . . . . .	22	Znaki towarowe . . . . .	56
Zapisywanie zmian w konfiguracji . . . . .	22	Warunki dotyczące dokumentacji produktu . . . . .	57
Zamykanie okna konfiguracji . . . . .	23	<b>Indeks</b> . . . . .	<b>59</b>
Analiza rozdzielonych tożsamości (węzeł źródłowy Entity Analytics (EA)) . . . . .	23		
Wybieranie źródła danych . . . . .	23		
Zmiana nazwy zmiennych zawierających dane . . . . .	24		
Ustawianie typu informacji dla zmiennych zawierających dane . . . . .	24		
Dodawanie węzłów do strumienia . . . . .	24		



---

## Przedmowa

IBM® SPSS Modeler to oferowane przez IBM Corp. zaawansowane środowisko eksploracji danych. SPSS Modeler pomaga przedsiębiorstwom i instytucjom w rozwijaniu relacji z klientami i obywatelami w oparciu o pogłębioną interpretację dostępnych danych. Organizacje korzystają z wiedzy uzyskanej dzięki programowi SPSS Modeler w bardzo szerokim spektrum zastosowań, m.in. do zatrzymywania najbardziej wartościowych klientów, określania możliwości sprzedaży wiązanej, przyciągania nowych klientów, wykrywania oszustw, ograniczania ryzyka i podnoszenia jakości usług publicznych.

Interfejs graficzny produktu SPSS Modeler zachęca użytkowników, aby wykorzystywali specjalistyczną wiedzę, dzięki której możliwe będzie opracowanie bardziej wydajnych modeli predykcyjnych i skrócenie czasu potrzebnego do uzyskania rozwiązania. SPSS Modeler oferuje wiele technik modelowania, takich jak predykcja, klasyfikacja, segmentacja i algorytmy do wykrywania związków. Po utworzeniu modeli program IBM SPSS Modeler Solution Publisher umożliwia udostępnienie ich osobom podejmującym decyzje w całym przedsiębiorstwie lub zapisanie w bazie danych.

## Informacje o programie IBM Business Analytics

Oprogramowanie IBM Business Analytics dostarcza kompletne, spójne i dokładne informacje, na których mogą polegać osoby decyzyjne chcące polepszyć wyniki biznesowe. Wszechstronne portfolio obejmujące moduły: analiza biznesowa, analiza prognostyczna, zarządzanie wynikami i strategiami finansowymi oraz aplikacje analityczne, zapewnia jasny, natychmiastowy i pozwalający na podjęcie działań wgląd w bieżące wyniki oraz daje możliwość przewidywania przyszłych wyników. W połączeniu z licznymi rozwiązaniami branżowymi, sprawdzonymi praktykami i profesjonalnymi usługami, organizacje o różnych rozmiarach mogą wspomagać najwyższą produktywność, w sposób pewny zautomatyzować decyzje i uzyskać lepsze wyniki.

Oprogramowanie IBM SPSS Predictive Analytics będące częścią tego portfolio wspomaga organizacje w zakresie przewidywania przyszłych zdarzeń oraz pozwala proaktywnie wpływać na ten wgląd z korzyścią dla wyników finansowych. Klienci komercyjni, rządowi i uczelnie na całym świecie polegają na technologii IBM SPSS zapewniającej przewagę konkurencyjną, dzięki której przyciągają, zatrzymują i pozyskują nowych klientów, walcząc z nieuczciwością i ograniczając ryzyko. Wdrażając oprogramowanie IBM SPSS do swojej codziennej działalności, organizacje stają się przewidującymi przedsiębiorstwami, zdolnymi do zarządzania i automatyzacji decyzji w celu realizacji celów biznesowych i osiągnięcia mierzalnej przewagi konkurencyjnej. W celu uzyskania dalszych informacji lub skontaktowania się z przedstawicielem, proszę wejść na stronę <http://www.ibm.com/spss>.

## Wsparcie techniczne

Wsparcie techniczne jest dostępne w celu zapewnienia klientom obsługi technicznej. Klienci mogą się kontaktować z działem Wsparcia technicznego w celu uzyskania pomocy dotyczącej korzystania z produktów IBM Corp. lub pomocy w instalacji dla jednego z obsługiwanych środowisk sprzętowych. Aby skontaktować się z działem Wsparcia technicznego, wejdź na stronę internetową IBM Corp. pod adresem <http://www.ibm.com/support>. W przypadku prośby o pomoc należy przygotować swoje dane identyfikacyjne, dane swojej organizacji, a także dane dotyczące usług wsparcia.



---

# Rozdział 1. Analiza jednostek

---

## Informacje dotyczące analizy jednostek

**IBM SPSS Modeler Entity Analytics** dodaje dodatkowy wymiar do analiz predykcyjnych wykonywanych przy użyciu produktu IBM SPSS Modeler. Analizy predykcyjne stanowią próbę przewidzenia przyszłych zachowań na podstawie danych z przeszłości, natomiast techniki Entity Analytics służą przede wszystkim poprawie spójności danych bieżących poprzez rozwiązywanie konfliktów dotyczących tożsamości w samych rekordach. Tożsamość może dotyczyć osoby, organizacji, obiektu lub dowolnej innej jednostki, względem której może istnieć niejednoznaczność. Ustalenie tożsamości może być istotne w przypadku wielu dziedzin, takich jak zarządzanie relacjami z klientami, wykrywanie oszustw, pranie pieniędzy oraz bezpieczeństwo w skali krajowej i międzynarodowej.

Założmy, że dostępne są rekordy klienta z dwóch różnych źródeł i nie ma pewności, czy dotyczą one tej samej osoby, czy dwóch różnych.

### Źródło 1

Record no.: 70001  
Name: Jon Smith  
Address: 123 Main Street  
Tax Reference: 555-00-1111  
Driv. License: 0001133107  
Cred. Card: 10229127

### Źródło 2

Record no.: 9103  
Name: JOHNATHAN Smith  
Date of Birth: 06/17/1934  
Telephone: 555-1212  
Cred. Card: 10229128  
Email: jls@mail.com  
IP address: 9.50.18.77

Nie ma dokładnych dopasowań pomiędzy tymi dwoma rekordami. Jednak wprowadzenie trzeciego źródła umożliwia znalezienie wspólnych atrybutów.

### Źródło 3

Record no.: 6251  
Name: Jon Smith  
Telephone: 555-1212  
Driv. License: 0001133107  
Cred. Card: 10229132

Numer prawa jazdy łączy się z rekordami w źródle 1 i źródle 3, a numer telefonu łączy się ze źródłami 2 i 3. Można by mieć pewność, że wszystkie trzy źródła odnoszą się do tej samej osoby.

Co zrobić, jeśli nie można przeprowadzić tak prostego rozróżnienia? Może być dostępna bardzo mała ilość danych będących podstawą do podjęcia decyzji. Należy wziąć pod uwagę następujące dwa rekordy.

### Źródło 4

Record no.: S45286  
Name: John T Smith Jr  
Address: 456 Main Street  
Telephone: 703-555-2000

Date of birth: 03/12/1984

Record no.: S45287  
Name: John T Smith  
Address: 456 Main Street  
Telephone: 703-555-2000  
Driv. License: 009900991

Najwyraźniej nie jest to ten sam pan Smith z wcześniejszego rekordu — różnice wystarczają, by wykluczyć taką możliwość. Wciąż jednak istnieje problem. Dwa różne rekordy z tego samego źródła danych wydają się dotyczyć tej samej osoby. Czy są to powielone rekordy? Nie można być tego pewnym, dopóki nie znajdzie się inny powiązany rekord zawierający więcej informacji, być może pochodzący z innego źródła.

#### Źródło 5

Record no.: 769582-2  
Name: John T Smith Sr  
Address: 456 Main Street  
Telephone: 703-555-2000  
Driv. License: 009900991  
Date of birth: 06/25/1959

To rozwiązuje problem. Dwa rekordy w źródle 4 nie są duplikatami, ale odnoszą się do ojca i syna o tym samym nazwisku, mieszkających pod tym samym adresem i korzystających z tego samego numeru telefonu. W systemie ręcznym znalezienie jednego rekordu rozdzielającego te tożsamości zajęłoby kilka tygodni. W zautomatyzowanym systemie analizy jednostek czas znalezienia rozwiązania został radykalnie skrócony.

---

## Analiza jednostek i analiza predykcyjna

Jeśli wszystkie dane dotyczą kompletnych i jednoznacznych rekordów z jednego źródła, rozwiązanie konfliktów tożsamości w IBM SPSS Modeler będzie dość prostym zadaniem. Korzystając z analizy predykcyjnej, można wczytać dane do programu IBM SPSS Modeler, przetworzyć je i uzyskać wiarygodne wyniki.

W warunkach rzeczywistych sytuacja różni się jednak znacząco. Dane są zwykle niekompletne, bywają dwuznaczne i są rozproszone w kilku różnych źródłach danych. Co więcej, mają wiele różnych atrybutów z niewielką liczbą nakładających się zmiennych. Wartość analizy jednostek wynika częściowo z połączenia danych z różnych źródeł w jednym centralnym miejscu zapisu zwanym **repozytorium**. System analizy jednostek następnie analizuje dane z dużą szczegółowością w celu rozwiązania konfliktów, dodając unikalny identyfikator do rekordów pochodzących od jednej osoby lub organizacji.

Poniższa tabela przedstawia różnice pomiędzy dwoma typami analiz.

*Tabela 1. Różnice między analizą jednostek i analizą predykcyjną.*

Cecha	Analiza predykcyjna	Analiza jednostek
Typy danych uczących	Na podstawie stosunkowo małych zbiorów i zakresów numerycznych	Może wykorzystywać duże zbiory (zmiennie bez typu), jak imiona i nazwiska czy adresy
Rozmiar danych uczących	Zwykle ignoruje duże zbiory (zmiennie bez typu)	Wszystkie wykorzystywane dane
Uogólnienie	Algorytm uogólnia dane uczące, tak aby utworzyć spójny model	Dane zachowane w strukturach odpowiednie do dopasowania jednostek i wykrywania relacji



Tabela 1. Różnice między analizą jednostek i analizą predykcyjną (kontynuacja).

<b>Cecha</b>	<b>Analiza predykcyjna</b>	<b>Analiza jednostek</b>
Wykrywanie oszustw	Rekordy oznaczane jako potencjalnie oszukańcze, jeśli mają typowe cechy oszukańczego wniosku kredytowego	Rekordy oznaczane jako potencjalnie oszukańcze, jeśli są powiązane ze znanymi oszukańczymi rekordami lub jeśli pochodzą od tej samej osoby z inną tożsamością



---

## Rozdział 2. Analiza jednostek za pomocą programu IBM SPSS Modeler

---

### Korzystanie z analizy jednostek w programie IBM SPSS Modeler

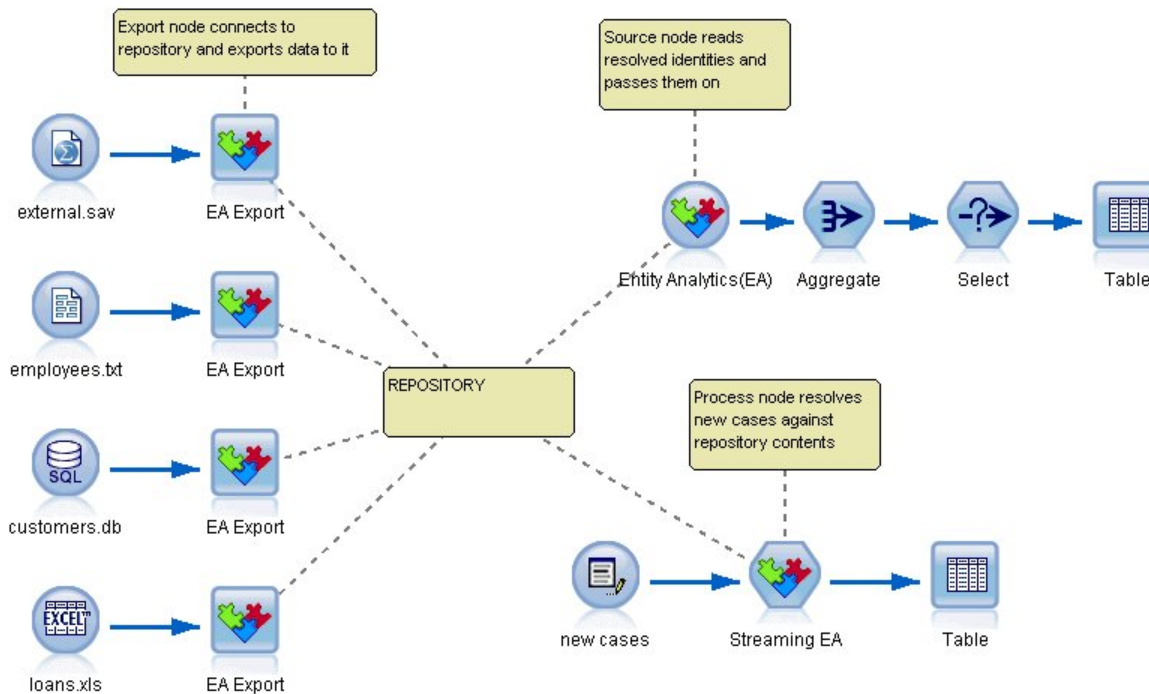
Istnieje podejrzenie, że w danych występują problemy związane z tożsamością. Na przykład poszczególne osoby mogą pojawiać się kilkakrotnie, a pojedyncze podmioty mogą być wyświetlane jako scalone lub może ich brakować. Jak można rozwiązać ten problem w programie IBM SPSS Modeler Entity Analytics? Poniżej opisana jest proponowana procedura. Można ją jednak modyfikować w zależności od indywidualnych potrzeb.

- Wczytaj dane źródłowe do programu IBM SPSS Modeler
- Utwórz repozytorium gotowe do przechowywania danych
- Połącz program IBM SPSS Modeler z repozytorium
- Zmapuj zmienne danych do cech repozytorium
- Wyeksportuj dane do repozytorium i rozdziel tożsamości
- Przeanalizuj rozdzielone tożsamości
- Rozdziel nowe obserwacje w oparciu o repozytorium
- Utwórz nowe wymagane alerty (wsadowe lub w czasie rzeczywistym)

Na tym etapie przychodzi czas, by zapoznać się z programem IBM SPSS Modeler. IBM SPSS Modeler jest narzędziem przyjaznym dla użytkownika wykorzystującym elementy graficzne obrazujące dane przepływające przez węzły. Każdy węzeł stanowi określony etap procesu.

IBM SPSS Modeler udostępnia szeroką gamę węzłów reprezentujących wszystkie standardowe funkcje eksploracji danych. Za pomocą programu IBM SPSS Modeler Entity Analytics można dodawać węzły używane w szczególności w analizie jednostek. Są to: węzeł Eksport EA, węzeł źródłowy Entity Analytics (EA), węzeł procesu przesyłania strumieniowego EA.

Na poniższym rysunku przedstawiony został przebieg procesu.



Rysunek 1. Proces analizy jednostek

## Etap 1: Wczytywanie danych źródłowych do programu SPSS Modeler

W pierwszej kolejności należy wczytać dane do programu SPSS Modeler, korzystając z jednego lub większej liczby węzłów źródłowych oznaczonych w SPSS Modeler za pomocą okrągłej ikony.

Dane mogą być w dowolnym formacie obsługiwany przez SPSS Modeler, np. pliki tekstowe, tabele baz danych, arkusze kalkulacyjne, pliki XML itd. Każdy format wymaga jednak odpowiedniego węzła źródłowego w programie SPSS Modeler. Na rysunku zaprezentowany został węzeł źródłowy bazy danych.

Każdy plik źródła danych musi zawierać co najmniej jedną zmienną, która w sposób unikalny identyfikuje każdy rekord. Jeśli w źródle danych nie znajduje się taka zmienna, łatwo będzie można ją dodać w programie SPSS Modeler. Więcej informacji można znaleźć w temacie “Dodawanie unikalnego identyfikatora rekordu” na stronie 12.

Więcej informacji można znaleźć w temacie “Łączenie ze źródłem danych” na stronie 12.

**Uwaga:** Znaki z alfabetu innego niż łaciński nie są obsługiwane. Jeśli dane uwzględniają kilka rekordów zawierających zestawy znaków łacińskich (np. zachodnioeuropejskie) i innych, rozwiązane zostaną wpisy zawierające znaki łacińskie.

## Etap 2: Tworzenie repozytorium

Głównym elementem wszystkich działań w zakresie analizy jednostek jest repozytorium, czyli centralny obszar, w którym przechowywane są wszystkie rekordy danych.

Aby utworzyć repozytorium, należy najpierw połączyć źródło danych z węzłem Eksport EA oznaczonym ikoną kwadratu.

Za pomocą węzła eksportu można utworzyć nowe repozytorium (lub wybrać istniejące), które od razu jest gotowe na odebranie danych.

Proces tworzenia repozytorium został dokładnie opisany w dalszej części. Więcej informacji można znaleźć w temacie “Ustawianie repozytorium jednostek (węzeł Eksport EA)” na stronie 11.

**Uwaga:** W przypadku włączania trybu zdalnego serwera repozytorium należy utworzyć na maszynie Modeler Server. Oznacza to, że podczas tworzenia repozytorium należy być podłączonym do oprogramowania Modeler Server z klienta Modeler. W ten sposób repozytorium EA zostanie utworzone na maszynie serwera.

Po skonfigurowaniu repozytorium można aktualizować jego zawartość na kilka sposobów. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

### Etap 3: Łączenie programu SPSS Modeler z repozytorium

Po utworzeniu repozytorium należy podłączyć do strumienia SPSS Modeler.

Więcej informacji można znaleźć w temacie “Opcje repozytorium jednostek” na stronie 13.

### Etap 4: Mapowanie zmiennych wejściowych do cech repozytorium

Źródła danych mogą zawierać wiele różnych rodzajów informacji o jednostkach. Wybrane typy informacji są wspólne dla większości źródeł danych dotyczących jednostek. Istnieją również takie, które są charakterystyczne wyłącznie dla określonych źródeł. W repozytorium jednostek różne typy informacji zwane są **cechami**. Repozytorium standardowo zawiera wiele cech; możliwe jest również tworzenie własnych.

Cecha repozytorium to indywidualny rodzaj informacji, który można wykorzystać ze źródłem danych jednostki. Wybranych cech (np. imię, nazwisko, data urodzenia itd.) można używać z wieloma różnymi źródłami danych. Innych natomiast można używać wyłącznie z określonym źródłem danych. Cecha jest zwykle tożsama ze zmienną w rekordzie danych lub kolumną w tabeli bazy danych.

Po utworzeniu i połączeniu z repozytorium należy zdefiniować jedną zmienną danych wejściowych jako zmienną **unikalny klucz** używaną w kolejnych analizach. Należy również zmapować zmienne danych wejściowych do odpowiednich cech w repozytorium. Mapowanie do wcześniej zdefiniowanych cech stanowi dla repozytorium jednostek informację dotyczącą zmiennych do porównania oraz sposobu porównania. Węzeł Eksport EA udostępni tabelę mapowania, w której można tworzyć mapowania.

Więcej informacji można znaleźć w temacie “Mapowanie zmiennych wejściowych do cech (węzeł Eksport EA)” na stronie 14.

### Etap 5: Eksport danych do repozytorium i rozdzielanie tożsamości

Każdy węzeł źródła danych wymaga własnego węzła Eksport EA. Jeśli więc dane pochodzą z wielu różnych źródeł, może wystąpić strumień z kilkoma źródłami danych, z których każdy jest połączony z odrębnym węzłem Eksport EA. Więcej informacji można znaleźć w temacie “Korzystanie z analizy jednostek w programie IBM SPSS Modeler” na stronie 5.

Jeśli występuje więcej niż jedno źródło, można wybrać odczytywanie rekordów z jednego, kilku lub wszystkich źródeł danych. System Entity Analytics analizuje wybrane rekordy i do każdego z nich dodaje zmienną identyfikatora o nazwie \$EA\_ID. Jeśli istnieje możliwość rozdzielania dwóch lub większej liczby rekordów powiązanych z wcześniej dwuznacznymi tożsamościami, identyfikatory dodane do tych rekordów będą unikalne w całym repozytorium. W systemie zostanie również dodana zmienna, aby wskazać źródło danych, z którego pochodzi dany rekord.

Każdy węzeł źródłowy bazy danych należy połączyć z odpowiednim węzłem Eksport EA, zmapować zmienne wejściowe do cech repozytorium, a następnie uruchomić strumień w celu wyeksportowania danych z SPSS Modeler do repozytorium i rozdzielenie dowolne tożsamości podczas jednej operacji. Aby zobrazować te czynności, wyobraźmy sobie, że następujące rekordy pochodzą z czterech różnych źródeł danych.

Dane zewnętrzne

Tabela 2. Dane zewnętrzne

Imię	Numer telefonu	Ryzyko kredytowe
Mike	555-1234	560
Joe	555-4567	780

Pracownicy

Tabela 3. Pracownicy

Imię	Adres	Numer telefonu
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Klienci

Tabela 4. Klienci

Imię	Adres	Oszczędności
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Pożyczki

Tabela 5. Pożyczki

Imię	Adres	Numer telefonu	Pożyczka
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

Jak można było zobaczyć, wszystkie źródła danych zostały kolejno wyeksportowane do repozytorium. W międzyczasie w repozytorium aktualizowane jest rozdzielanie poszczególnych rekordów. Każdy rekord w repozytorium poprzedza zmienna identyfikatora (o nazwie *SEA-ID*) i zmienna wskaźnika źródła (o nazwie *SEA-SRC*), która wskazuje źródło danych, z którego pochodzi rekord. Ta więc w naszym przykładzie po wyeksportowaniu wszystkich czterech źródeł danych zawartość repozytorium będzie wyglądała następująco.

Tabela 6. Przykładowa zawartość repozytorium po etapie eksportu.

SEA-ID	SEA-SRC	Imię	Numer telefonu	Adres	Ryzyko kredytowe	Oszczędności	Pożyczka
1	Pracownicy	Michael	555-1234	1234 5th St			
1	Zewnętrzne	Mike	555-1234		560		
2	Klienci	Joe		777 Oak St		\$5	
2	Zewnętrzne	Joe	555-4567		780		
2	Pożyczki	Joseph	555-4567	777 Oak St			\$50,000
3	Pracownicy	Fred	555-9876	543 1st Ave			
4	Klienci	Susan		1234 5th St		\$1234	
4	Pożyczki	Sue	555-1234	1234 5th St			\$10,000

W systemie Entity Analytics na podstawie takiego samego numeru telefonu określono, że *Mike* w zbiorze danych *Zewnętrzne* jest tą samą osobą co *Michael* w zbiorze danych *Pracownicy*. Osobie tej nadano identyfikator 1.

W przypadku *Joe'ego* w zbiorze danych *Zewnętrzne* sytuacja jest bardziej skomplikowana. Czy jest to ta sama osoba co *Joe* w zbiorze *Klienci*? Nie można tego stwierdzić na podstawie tych dwóch źródeł danych. Istnieje jednak trzecie źródło danych, *Pożyczki*, zawierające rekord *Joseph*. Mamy teraz dopasowanie: numer telefonu Josepha jest taki sam jak *Joe'ego* w zbiorze danych *Zewnętrzne*. Na tej podstawie system określa, czy jest to ta sama osoba, i nadaje jej identyfikator 2.

Dostępnych jest wiele rekordów *Fred* — osoba ta otrzymała identyfikator 3. *Susan* ze zbioru *Klienci* została zidentyfikowana jako ta sama osoba co *Sue* ze zbioru *Pożyczki*, ponieważ ma taki sam adres. Osobie tej został nadany identyfikator 4.

*Uwaga:* W tym przykładzie do celów ilustracji dopasowanie miało optymistyczny przebieg. Można również wybrać bardziej pesymistyczny zbiór reguł, w którym imię, numer telefonu lub adres same z siebie nie stanowią dokładnego dopasowania, a oba rekordy będą miały przypisany ten sam identyfikator.

## Etap 6: Analiza rozdzielonych tożsamości

Po rozdzieleniu tożsamości w repozytorium można przejść do dalszej analizy i przetwarzania wyników. Na przykład, w przypadku podejrzenia nieuczciwych działań związanych z istnieniem powielonych rekordów dla tej samej tożsamości przydatne może być utworzenie raportu z powielonymi rekordami.

W pierwszej kolejności należy utworzyć węzeł źródłowy Entity Analytics (EA) i połączyć go z repozytorium.

Wspólny wynik z węzła składa się z następujących zmiennych.

- Zmienna identyfikatora dodana przez system (*\$EA-ID* w przykładzie z etapu 5)
- Zmienna wskaźnika źródłowego dodana przez system (*\$EA-SRC* w przykładzie z etapu 5)
- Unikalna zmienna kluczowa wyznaczona na etapie 4

Dodatkowo w przypadku relacji generowane są następujące wyniki. Więcej informacji można znaleźć w temacie “Wybieranie źródła danych” na stronie 23.

- Stopień separacji pomiędzy jednostkami (*\$EA-DEGREE*)
- Zmienna nadrzędna (*\$EA-PARENT*)
- Zmienna podrzędna (*\$EA-CHILD*)
- Reguła identyfikująca relację (*\$EA-RULE*)

Aby wyświetlić wyniki w programie SPSS Modeler, można dołączyć węzeł wyników SPSS Modeler, taki jak węzeł tabeli lub węzeł raportu, i uruchomić tę część strumienia. Jeśli wymagane jest podsumowanie wyników, które mogą być dość pokaźne, można dodać węzły operacji na rekordach, takie jak węzły agregacji lub wyboru.

Węzeł źródłowy Entity Analytics (EA) został dokładnie opisany w dalszej części. Więcej informacji można znaleźć w temacie “Analiza rozdzielonych tożsamości (węzeł źródłowy Entity Analytics (EA))” na stronie 23.

## Etap 7: Rozdzielanie nowych obserwacji w oparciu o repozytorium

Rozdzielanie tożsamości we wszystkich rekordach w danych źródłowych. Co jednak stanie się, jeśli zostanie porównany nowy zbiór rekordów w celu wyświetlania relacji z istniejącymi informacjami, tak aby uzyskać lepszą ocenę? W takiej sytuacji przydatny będzie węzeł przesyłania strumieniowego EA.

W pierwszej kolejności należy dodać nowy węzeł źródła danych SPSS Modeler, tak aby wczytać nowe dane do strumienia. Następnie należy połączyć węzeł źródła z węzłem przesyłania strumieniowego EA. Aby wyświetlić wyniki, należy dodać węzeł tabeli tak jak poprzednio.

Po uruchomieniu tej części strumienia węzeł przesyłania strumieniowego EA odczyta każdy nowy rekord i porówna go z zawartością repozytorium. W przypadku znalezienia dopasowanych rekordów w repozytorium węzeł przesyłania strumieniowego EA wyświetli wszystkie dopasowane rekordy z nowym rekordem, do którego doda identyfikator i

zmienne wskaźnika źródła. Jeśli dopasowane rekordy nie zostaną odnalezione, węzeł procesu wyświetli tylko nowe rekordy z dodanym identyfikatorem i zmiennymi wskaźnika źródła.

Aby zobrazować ten przykład, wyobraźmy sobie, że repozytorium aktualnie składa się z zawartości, która została wygenerowana przez węzeł źródłowy Entity Analytics (EA). Patrz Tabela 6 na stronie 8.

Utworzone zostały następujące rekordy. Czy związane są z jakąkolwiek znajomą osobą?

Tabela 7. Nowe rekordy podlegające ocenie.

Imię	Adres	Numer telefonu	Pożyczka
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

Podczas porównywania nowych danych z istniejącą zawartością repozytorium węzeł przesyłania strumieniowego EA najpierw dopasowuje nowy rekord z osobą posiadającą identyfikator 4 w istniejących rekordach. Nie można jednak znaleźć dopasowania do drugiego nowego rekordu. W tym przypadku zostanie przypisany nowy, unikalny identyfikator 5.

Węzeł przesyłania strumieniowego EA dodaje identyfikator i zmienną wskaźnika źródła i generuje nowy rekord razem z wszystkimi dopasowanymi rekordami. Wygenerowane wyniki będą wyglądać następująco.

Tabela 8. Wyniki z węzła przesyłania strumieniowego EA.

SEA-ID	SEA-SRC	Imię	Numer telefonu	Adres	Ryzyko kredytowe	Oszczędności	Pożyczka
4	Cust	Susan		1234 5th St		\$1234	
4	Pożyczka	Sue	555-1234	1234 5th St			\$10,000
4	Nowa pożyczka	Suzan	555-1234	1234 5th Street			\$100,000
Źródło 5	Nowa pożyczka	Mark	555-9999	888 9th Ave			\$60,000

Wyniki można następnie zagregować za pomocą identyfikatora analizy jednostek jako zmienną agregacji i przekazać do innych węzłów podrzędnych do dalszego przetwarzania.

Węzeł przesyłania strumieniowego EA został dokładnie opisany w dalszej części.

## Etap 8: Generowanie alertów

Ponownie konieczna może być analiza podejrzanych działań. W tym przypadku osoba z identyfikatorem 4 wzięła już pożyczkę o wartości 10 000 USD, a teraz złożyła pod bardzo podobnym imieniem i nazwiskiem wniosek o kolejną, dziesięciokrotnie większą, pożyczkę. Może to być oczywiście akceptowalne działanie bez jakiegokolwiek zamiaru defraudacji. Jeśli jednak zgodnie z obowiązującymi zasadami biznesowymi takie działanie stanowi zachowanie podejrzane, warto wówczas przeprowadzić dokładniejszą analizę.

Można na przykład dołączyć i uruchomić węzeł tabeli lub węzeł raportu SPSS Modeler, wydrukować zawartość z okna wyników, przeanalizować ją i ewentualnie ręcznie utworzyć alerty. Opcjonalnie wyniki przesyłania strumieniowego EA można przekazać do modelu oceny ryzyka, który został wcześniej utworzony w IBM SPSS Modeler. Można w ten sposób wygenerować zestaw ocen, który jeszcze dokładniej odzwierciedla reguły biznesowe. Inną możliwością jest wyeksportowanie wyników do bazy danych lub innego miejsca w celu dalszego przetworzenia. IBM SPSS Modeler udostępnia wiele możliwości działania dostosowanych do określonych wymogów.



---

## Rozdział 3. Zadania w zakresie analizy jednostek

---

### Informacje o zadaniach

W tej sekcji znajduje się opis następujących zadań w zakresie analizy jednostek.

- Ustawianie repozytorium jednostek
- Konfigurowanie repozytorium jednostek
- Analizowanie rozdzielonych tożsamości
- Rozdzielenie nowych obserwacji w oparciu o repozytorium jednostek
- Czyszczenie repozytorium jednostek
- Usuwanie repozytorium jednostek
- Korzystanie z analizy jednostek z innymi produktami IBM SPSS
- Analiza jednostek administracyjnych

---

### Ustawianie repozytorium jednostek (węzeł Eksport EA)

Proces ustawiania repozytorium jednostek składa się z następujących zadań.

1. Połączenie ze źródłem danych. Więcej informacji można znaleźć w temacie “Łączenie ze źródłem danych” na stronie 12.
2. Utworzenie repozytorium. Więcej informacji można znaleźć w temacie “Tworzenie repozytorium” na stronie 12.
3. Mapowanie zmiennych wejściowych w źródle danych do cech w repozytorium. Więcej informacji można znaleźć w temacie “Mapowanie zmiennych wejściowych do cech (węzeł Eksport EA)” na stronie 14.

Po skonfigurowaniu mapowań można je wyświetlić dla bieżącego źródła danych lub dla wszystkich źródeł danych znanych w repozytorium. Więcej informacji można znaleźć w temacie “Wyświetlanie mapowań zmiennych (węzeł Eksport EA)” na stronie 15.

**Uwaga:** Poczynając od wersji 16 SPSS Entity Analytics obsługuje repozytoria w produkcie IBM DB2. Repozytorium jest specyficzne dla każdej wersji SPSS Modeler i nie można go importować z wcześniejszych wersji. Dlatego w przypadku posiadania repozytorium i aktualizacji wersji 16 SPSS Entity Analytics należy ponownie utworzyć repozytorium w nowej bazie danych DB2.

### Czyszczenie repozytorium jednostek

Repozytorium udostępnia centralny obszar zapisu, działający jak pamięć podręczna danych dla wszystkich informacji o jednostce. Ponieważ repozytorium działa w czasie rzeczywistym, występuje w jednym stanie i nie ma możliwości stworzenia wersji z repozytorium jednostek. W repozytorium znajdują się wszystkie aktualne dane wejściowe, a ich objętość może być bardzo duża.

Zawartość repozytorium można opracować, korzystając z łatwego w użyciu interfejsu graficznego. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

**Ważne:** Poczynając od wersji 16 SPSS Entity Analytics obsługuje repozytoria na serwerach IBM DB2; poprzednie wersje SPSS Entity Analytics obsługiwały repozytoria na serwerach IBM solidDB. W przypadku posiadania istniejącego repozytorium solidDB po aktualizacji programu SPSS Entity Analytics do wersji 16 lub późniejszej konieczne będzie odtworzenie tego repozytorium w nowej bazie danych DB2.

## Łączenie ze źródłem danych

W pierwszej kolejności należy wczytać dane źródłowe do SPSS Modeler za pomocą węzła źródłowego.

Aby połączyć ze źródłem danych

1. Na karcie Źródła na palecie węzłów w dolnej części głównego okna SPSS Modeler dwukrotnie kliknij ikonę odpowiadającą typowi danych źródłowych. W ten sposób węzeł źródłowy zostanie dodany do obszaru roboczego ekranu.
2. W obszarze roboczym ekranu dwukrotnie kliknij ikonę, aby otworzyć okno dialogowe.
3. W polu Plik wprowadź lokalizację i nazwę pliku danych źródłowych.
4. W razie potrzeby uzupełnij pozostałe elementy w oknie dialogowym (kliknij przycisk Pomoc, aby uzyskać więcej informacji), a następnie kliknij OK.
5. Jeśli plik danych źródłowych nie zawiera zmiennej, która w sposób unikalny identyfikuje każdy rekord, należy dodać taką zmienną przez węzeł wyliczeń. Więcej informacji można znaleźć w temacie “Dodawanie unikalnego identyfikatora rekordu”.

**Uwaga:** Znaki z alfabetu innego niż łański nie są obsługiwane. Jeśli dane uwzględniają kilka rekordów zawierających zestawy znaków łańskich (np. zachodnioeuropejskie) i innych, rozwiązane zostaną wpisy zawierające znaki łańskie.

### Dodawanie unikalnego identyfikatora rekordu

Każdy plik źródła danych przekazywany do repozytorium jednostek musi zawierać jedną zmienną, która w sposób unikalny identyfikuje każdy rekord. Jeśli plik źródła danych nie zawiera takiej zmiennej, można ją dodać poprzez węzeł wyliczeń SPSS Modeler.

Aby dodać unikalny identyfikator rekordu do pliku źródła danych

1. W obszarze roboczym ekranu kliknij węzeł źródłowy dodany w poprzednim zadaniu.
2. Na karcie **Zmienne** na palecie węzłów dwukrotnie kliknij ikonę **Wyliczanie**, aby dołączyć węzeł wyliczeń do węzła źródłowego.
3. W obszarze roboczym ekranu dwukrotnie kliknij węzeł wyliczeń, aby otworzyć okno dialogowe.
4. W polu **Wyliczanie** zamień domyślną nazwę na znaczącą nazwę (np. **ID**) dla dodawanej zmiennej identyfikatora.
5. Sprawdź, czy dla pola **Wylicz jako** została wybrana wartość **Formuła**.
6. W obszarze **Typ zmiennej** wybierz wartość **Ilościowa**.
7. W polu tekstowym **Formuła** wpisz **@INDEX** i kliknij **OK**.

## Tworzenie repozytorium

Repozytorium służy do przechowywania wszystkich danych wejściowych.

**Uwaga:** W przypadku włączania trybu zdalnego serwera repozytorium należy utworzyć na maszynie Modeler Server. Oznacza to, że podczas tworzenia repozytorium należy być podłączonym do oprogramowania Modeler Server z klienta Modeler. W ten sposób repozytorium EA zostanie utworzone na maszynie serwera.

Aby utworzyć repozytorium

1. Z karty Eksportuj na palecie węzłów SPSS Modeler przeciągnij węzeł Eksport EA do obszaru roboczego strumienia.

*Uwaga:* W przypadku tworzenia repozytorium po raz pierwszy należy użyć węzła Eksport EA i połączyć go z węzłem źródłowym SPSS Modeler zawierającym dane, które mają być przekazane do repozytorium (lub węzła wyliczeń, o ile został dodany w celu uzyskania zmiennej unikalnego identyfikatora). Aby połączyć węzły, należy wykonać następujące kroki.

- a. Kliknij prawym przyciskiem myszy węzeł źródłowy SPSS Modeler.
- b. Wybierz opcję Połącz.
- c. Kliknij węzeł Eksport EA.

2. Aby otworzyć okno dialogowe, dwukrotnie kliknij węzeł Eksport EA.
3. Kliknij listę **Repozytorium jednostek**.
4. Kliknij przycisk **<Przełączaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
5. W oknie dialogowym Instancje rozdzielania jednostek kliknij pole Nazwa repozytorium.
6. Wybierz polecenie **<Stwórz/Dodaj nowe repozytorium...>**, aby wyświetlić kreator tworzenia repozytorium.

## Kreator tworzenia repozytorium

### Krok 1

W tym miejscu można zdecydować o utworzeniu lokalnego repozytorium korzystającego z produktu IBM DB2 powiązanego z IBM SPSS Modeler Entity Analytics lub o wykorzystaniu dla repozytorium z zewnętrznej bazy danych.

**Utwórz lokalne repozytorium.** Określ nazwę użytkownika i hasło dla administratora bazy danych IBM DB2, gdzie będzie znajdować się tworzone repozytorium. Potwierdź hasło i kliknij przycisk **Dalej**.

**Uwaga:** W nazwie użytkownika nie można użyć myślnika ani podkreślnika.

Zadania administratora repozytorium w węzłach Entity Analytics, np. tworzenie lub niszczenie repozytorium, wymagają dodatkowych uprawnień. W systemie UNIX użytkownik zalogowany do serwera IBM SPSS Modeler Server musi należeć do grupy db2iadm1. W celu wykonywania zadań administratora użytkownik zalogowany do IBM SPSS Modeler Server w systemie Windows musi być członkiem grupy DB2ADMNS.

Jeśli konieczna będzie późniejsza zmiana danych uwierzytelniających administratora, można skorzystać z edytora wiersza komend dla bazy danych. Więcej informacji można znaleźć w temacie “Zarządzanie danymi uwierzytelniającymi administratora dla bazy danych repozytorium” na stronie 30.

**Uwaga:** Możliwe jest tylko jedno połączenie nazwy użytkownika i hasła. Wszyscy użytkownicy zalogowani do repozytorium współdzielą nazwę użytkownika i hasło.

**Dodaj zewnętrzne repozytorium.** Ta opcja umożliwia użycie zewnętrznej bazy danych w celu udostępnienia repozytorium. Wpisz lokalizację pliku .ini bazy danych w polu **Zaznacz plik repozytorium .ini:** i kliknij **Dalej**.

### Krok 2

**Nowa nazwa repozytorium.** Wpisz unikalną nazwę nowego repozytorium.

**Importuj konfigurację z.** (Wyłącznie lokalne repozytorium) Aby do konfiguracji wykorzystać konfigurację istniejącego repozytorium, w tym miejscu wybierz to repozytorium. W przeciwnym razie wybierz opcję **Domyślny**. Aby uzyskać więcej informacji, patrz “Konfigurowanie repozytorium jednostek” na stronie 16.

Po wybraniu istniejącego repozytorium wprowadź szczegóły dotyczące połączenia, jeśli różnią się od wprowadzonych na poprzednim ekranie.

Kliknij przycisk **OK**, aby utworzyć nowe repozytorium i wyświetlić okno dialogowe Instancje rozdzielania jednostek. Za pośrednictwem tego okna można połączyć się z repozytorium.

## Opcje repozytorium jednostek

Okno dialogowe Repozytoria jednostek zawiera wiele opcji umożliwiających tworzenie, łączenie, konfigurowanie i utrzymywanie repozytorium jednostek.

**Połącz się z repozytorium.** Ta opcja służy do tworzenia nowego repozytorium jednostek lub połączenia z istniejącym.

- **Nazwa repozytorium.** Wyświetla bieżące repozytorium jednostek (o ile istnieje). Aby wybrać inne repozytorium (jeśli istnieje kilka repozytoriów), należy wybrać je z listy.

Aby utworzyć nowe repozytorium, wybierz opcję **<Stwórz/Dodaj nowe repozytorium...>**. Spowoduje to uruchomienie kreatora, który przeprowadzi użytkownika przez proces tworzenia repozytorium.

- **Nazwa użytkownika.** Dla wybranego repozytorium wprowadź poprawną nazwę użytkownika.
- **Hasło.** Hasło dla tego użytkownika.
- **Połącz.** Kliknij, aby połączyć się z bieżącym repozytorium.

**Zarządzaj repozytorium.** W tabeli znajdują się źródła danych pobrane do bieżącego repozytorium (z którym użytkownik jest aktualnie połączony). Wyświetlana jest również liczba rekordów w każdym źródle danych.

- **Odśwież.** Aktualizuje źródło danych i wielkość informacji w tabeli, np. po dodaniu nowego źródła danych lub zmianie wielkości istniejącego.
- **Wyczyść wszystko.** Usuwa z repozytorium wszystkie dane źródłowe, ale zachowuje wszystkie dane konfiguracyjne. Ta opcja jest przydatna, jeśli dane konfiguracyjne są wciąż potrzebne, a użytkownik chce usunąć wszystkie rekordy danych z repozytorium. Więcej informacji można znaleźć w temacie “Czyszczenie repozytorium jednostek” na stronie 32.
- **Usuń nieużywane.** Usuwa z repozytorium wszystkie wyróżnione dane źródłowe, ale zachowuje wszystkie dane konfiguracyjne. Więcej informacji można znaleźć w temacie “Usuwanie nieużywanych źródeł danych z repozytorium” na stronie 32.
- **Zmień nazwę źródła.** Otwiera okno dialogowe, w którym można zmienić nazwę wyróżnionego źródła danych.

**Uwaga:** Opcja ta umożliwia zmianę nazwy wewnątrz repozytorium; użytkownik będzie musiał ponownie wybrać nową nazwę źródła danych w dowolnym dostępnym węźle eksportu lub przesyłania strumieniowego, do którego ta nazwa się odnosi.

**Zniszcz całe repozytorium.** Całkowicie niszczy zawartość bieżącego repozytorium i dane konfiguracyjne. Więcej informacji można znaleźć w temacie “Usuwanie repozytorium jednostek” na stronie 32.

**Konfiguruj repozytorium.** Wyświetla okno, w którym można konfigurować bieżące repozytorium. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

## Mapowanie zmiennych wejściowych do cech (węzeł Eksport EA)

Repozytorium zawiera standardowo wiele wcześniej zdefiniowanych cech. Różne źródła danych mogą korzystać z różnych nazw zmiennych (na przykład **Address1** lub **Address Line 1**) dla typów informacji odpowiadających tej samej cesze. Aby uniknąć powielania, konieczne jest zmapowanie zmiennych źródła danych wejściowych do określonych cech repozytorium. Nie ma konieczności mapowania każdej zmiennej w zbiorze danych. Wystarczy zmapować tylko te, które mogą odpowiadać tej samej cesze w innych zbiorach danych.

Jeśli źródło danych korzysta ze zmiennych odpowiadających innym typom informacji, które nie zostały wcześniej zdefiniowane, można utworzyć nowe cechy w oknie Konfiguracja repozytorium. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

Aby zmapować zmienne wejściowe do cech

1. Dołącz węzeł Eksport EA do węzła źródła danych w obszarze roboczym strumienia. Każdy używany węzeł źródła danych musi być dołączony do własnego węzła Eksport EA.
2. Otwórz węzeł Eksport EA, aby wyświetlić kartę Wejścia, która zawiera opcje mapowania zmiennych wejściowych. Więcej informacji można znaleźć w temacie “Opcje repozytoriów wejściowych w procesie mapowania” na stronie 15.
3. W węźle Eksport EA wybierz kartę Repozytorium, aby wyświetlić przypisania mapowania dla bieżącego źródła danych lub dla wszystkich źródeł danych (w przypadku korzystania z większej liczby źródeł danych).
4. Aby zapisać przypisania mapowania (np. w celu użycia z węzłem eksportu dla innego źródła danych), kliknij opcję **Eksportuj mapowanie**.

Po skończeniu mapowania pierwszego węzła źródła danych powtórz proces dla każdego innego węzła źródła danych, który zostanie użyty.

## Opcje repozytoriów wejściowych w procesie mapowania

Na karcie danych wejściowych znajdują się opcje mapowania zmiennych źródeł danych do cech repozytorium gotowych do eksportu do repozytorium. Na tej karcie skonfiguruj przydziały mapowania; opcjonalnie kliknij kartę Repozytorium, aby wyświetlić mapowania dla innych źródeł danych, a następnie kliknij polecenie **Wykonaj**, aby wyeksportować dane do repozytorium.

Jeśli zbiór mapowań jest już zapisany w pliku XML, można je wykorzystać, klikając opcję **Importuj mapowanie**.

**Dominanta.** Pozostaw domyślny wybór **Dodaj do repozytorium**, aby dodać rekordy pliku źródłowego do istniejącej zawartości repozytorium. Aby usunąć zawartość repozytorium i zachować informacje na temat konfiguracji przed dodaniem rekordów źródłowych, wybierz polecenie **Wyczyść repozytorium przed wyeksportowaniem**.

**Repozytorium jednostek.** Wyświetla bieżące repozytorium jednostek (o ile istnieje). Aby wybrać inne repozytorium (jeśli istnieje kilka repozytoriów), należy wybrać je z listy. Aby utworzyć nowe repozytorium, wybierz przycisk **<Przełączaj...>**, aby wyświetlić okno dialogowe, w którym można utworzyć repozytorium. Więcej informacji można znaleźć w temacie “Opcje repozytorium jednostek” na stronie 13.

**Mapuj na typ jednostki.** Lista typów jednostek (tj. zbiory cech) zdefiniowanych w repozytorium. Wybierz jeden z listy lub wybierz opcję **<Dodaj nowy typ jednostki...>**, aby wyświetlić okno konfiguracji repozytorium, w którym można zdefiniować nowy typ jednostek. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

**Znacznik źródłowy.** Lista znaczników wskazujących źródła danych aktualnie znanych w repozytorium. Wybierz jeden z listy lub wybierz opcję **<Dodaj nowy znacznik źródła...>**, aby utworzyć znacznik dla nowego źródła danych.

**Unikalny klucz.** (wymagane) Zmienna wejściowa używana dla unikalnych identyfikatorów rekordów danych.

**Tabela mapowania.** W tej tabeli można mapować każdą zmienną wejściową do odpowiedniej cechy w repozytorium. Jeśli odpowiednia cecha nie istnieje w wybranym typie jednostki, nową cechę można utworzyć w tym miejscu.

- **Zmienna.** Zbiór zmiennych wejściowych w wybranym źródle danych. Każda zmienna zawiera ikonę wskazującą poziom pomiaru (tj. typ danych) dla zmiennej.
- **Mapowanie do cechy.** Aby zmapować zmienną do cechy, dwukrotnie kliknij tę kolumnę (lub naciśnij klawisz spacji) w wierszu zmiennej i wybierz cechę z listy. Jeśli odpowiednia cecha nie jest dostępna, wybierz opcję **<Dodaj nową cechę...>**, aby wyświetlić okno konfiguracji repozytorium, w którym można zdefiniować nową cechę dla tego typu jednostki. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.
- **Użycie.** Wskazuje kontekst poszczególnych zmiennych, jeśli możliwy jest więcej niż jeden kontekst, np. numer telefonu domowego i służbowego. Dla cechy ADDRESS i PHONE dostępne są wstępnie ustawione typy użycia. Można również utworzyć własny typ użycia dla wszystkich cech. Aby ustawić użycie inne niż zdefiniowane (**Automatycznie**), kliknij tę kolumnę w wybranym wierszu i wybierz jeden z istniejących typów użycia (jeśli występują) lub kliknij polecenie **<Dodaj użycie...>**, aby utworzyć nowe. Więcej informacji można znaleźć w temacie “Opracowywanie typów jednostek” na stronie 20.

**Importuj mapowanie.** Importuje wcześniej wyeksportowany zbiór mapowań zmiennej do cechy z zewnętrznego pliku XML. Opcja ta może być przydatna w przypadku różnych źródeł danych z tymi samymi wymaganiami. Pozwala na uniknięcie powtórzonego definiowania tych samych mapowań dla różnych źródeł.

**Eksportuj mapowanie.** Eksportuje do zewnętrznego pliku XML zbiór mapowań zmiennej do cechy wyświetlony w tabeli mapowania.

## Wyświetlanie mapowań zmiennych (węzeł Eksport EA)

Na karcie Repozytorium kliknij przycisk **Odśwież**, aby wyświetlić cechy repozytorium posiadające zmapowane zmienne wejściowe. Cechy te można wyświetlać dla bieżącego źródła danych (kontrolowanego przez węzeł źródłowy dołączony do tego węzła eksportu) lub wszystkich źródeł danych.

**Pokaż wejścia dla.** Wybierz opcję wyświetlania mapowań dla bieżącego źródła danych lub dla wszystkich źródeł danych znanych w repozytorium.

**Odśwież.** Aktualizuje wyświetlane elementy dla wybranej opcji danych wejściowych.

**Cechy.** Lista wszystkich cech z mapowaniami w wyświetlonych źródłach danych. Cechy niezmapowanie nie są wyświetlane.

**<Źródło danych>.** Każda kolumna zawiera zmapowane zmienne w poszczególnych źródłach danych dla każdej cechy ze zdefiniowanym mapowaniem.

---

## Konfigurowanie repozytorium jednostek

Zawartość repozytorium można konfigurować w oknie Konfiguracja repozytorium udostępniającym łatwy w użyciu interfejs wizualny dla całego repozytorium.

W przypadku korzystania z jednego lub większej liczby repozytoriów o takiej samej lub podobnej konfiguracji można wybrać podstawową konfigurację i wyeksportować ją do pliku, który można następnie zaimportować do innych repozytoriów. Więcej informacji można znaleźć w temacie “Ponowne użycie konfiguracji repozytorium” na stronie 22.

**Uwaga:** Począwszy od wersji 16 SPSS Entity Analytics obsługuje repozytoria w produkcie IBM DB2. Repozytorium jest specyficzne dla każdej wersji SPSS Modeler i nie można go importować z wcześniejszych wersji. Dlatego w przypadku posiadania repozytorium i aktualizacji wersji 16 SPSS Entity Analytics należy ponownie utworzyć repozytorium w nowej bazie danych DB2.

### UWAGA:

**Po zmianie i zapisaniu konfiguracji repozytorium zawierającego już dane może pojawić się monit o usunięcie zawartości repozytorium i ponowne wczytanie danych. W ten sposób można uniknąć niespójności w repozytorium.**

Aby skonfigurować repozytorium

1. Otwórz węzeł Entity Analytics.
2. Kliknij listę **Repozytorium jednostek**.
3. Kliknij przycisk **<Przeglądaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
4. W oknie dialogowym Instancje rozdzielania jednostek kliknij listę **Nazwa repozytorium**.
5. Wybierz repozytorium, które ma zostać skonfigurowane.
6. Jeśli nie zostało nawiązane połączenie, wprowadź nazwę użytkownika i hasło dla administratora i kliknij polecenie **Połącz**.
7. Jeśli przycisk **Konfiguruj repozytorium** jest włączony, kliknij go, aby wyświetlić okno Konfiguracja repozytorium.
8. Utwórz szczegóły konfiguracji zgodnie z opisem w poniższych sekcjach.

Okno nawigacji po lewej stronie okna Konfiguracja repozytorium zawiera strukturę drzewa, która umożliwia zarządzanie różnymi cechami repozytorium.

Tabela 9. Główne elementy okna Konfiguracja repozytorium.

Sekcja	Opis	
Źródła danych	Wyświetlanie mapowań wszystkich źródeł danych do różnych cech repozytorium.	Więcej informacji można znaleźć w temacie “Wyświetlanie zmapowanych źródeł danych” na stronie 17.
Cechy	Tworzenie nowych cech i duplikowanie, edytowanie lub usuwanie istniejących.	Więcej informacji można znaleźć w temacie “Opracowywanie cech repozytorium” na stronie 17.

Tabela 9. Główne elementy okna Konfiguracja repozytorium (kontynuacja).

Sekcja	Opis	
Typy jednostek	Tworzenie nowego typu jednostek lub zarządzanie istniejącymi typami (duplikowanie, zmiana nazwy, dołączanie lub usuwanie cech, usuwanie).	Więcej informacji można znaleźć w temacie "Opracowywanie typów jednostek" na stronie 20.
Reguły rozdzielania	Ustawianie progu dopasowania jednostek.	Więcej informacji można znaleźć w temacie "Ustawianie progu dopasowania jednostek" na stronie 22.

## Wyświetlanie zmapowanych źródeł danych

W sekcji Źródła danych w oknie Konfiguracja repozytorium wpis Wszystkie źródła zawiera listę tylko do odczytu z mapowaniami wszystkich źródeł danych do różnych cech repozytorium.

Kliknij przycisk **Odśwież**, aby zaktualizować listę po dodaniu do repozytorium nowych źródeł danych.

*Uwaga:* W tym miejscu nie można dodawać do repozytorium źródła danych. Źródła danych można dodawać wyłącznie poprzez utworzenie węzła źródłowego SPSS Modeler i połączenie go z węzłem eksportu Entity Analytics. Więcej informacji można znaleźć w temacie "Łączenie ze źródłem danych" na stronie 12.

## Opracowywanie cech repozytorium

Cecha repozytorium to indywidualny rodzaj informacji, który można wykorzystać ze źródłem danych jednostki. Wybranych cech (np. imię, nazwisko, data urodzenia itd.) można używać z wieloma różnymi źródłami danych. Innych natomiast można używać wyłącznie z określonym źródłem danych. Cecha może zawierać jeden lub większą liczbę elementów; każdy element odpowiada zwykle zmiennej w rekordzie danych lub kolumnie w tabeli bazy danych.

W sekcji Cechy okna Konfiguracja repozytorium wpis Wszystkie cechy umożliwia opracowanie wszystkich cech repozytorium. Można wykonać następujące czynności.

- Tworzenie nowej cechy
- Powielenie istniejącej cechy (np. w celu utworzenia nowej cechy w oparciu o istniejącą)
- Edytowanie istniejącej cechy
- Usuwanie istniejącej cechy

Instrukcje dotyczące wykonywania tych zadań znajdują się w dalszej części tej sekcji.

Lista cech zawiera wszystkie cechy zdefiniowane w repozytorium. W kolumnach na liście wyświetlane są różne właściwości cechy.

**Cecha.** Nazwa cechy. Symbol kłódki obok nazwy cechy wskazuje, że cecha jest zablokowana. Zablokowanych cech nie można usunąć ani zduplikować. Jedyną zmianą, którą można zapisać, jest zmiana atrybutu anonimizacji.

**Częstość.** Wskazuje liczbę jednostek z taką samą wartością dla tej cechy. Poprawne wartości to **Jedna** (np. numer paszportu), **Kilka** (np. adres) lub **Wiele** (np. data urodzenia).

**Wyłączność.** Wskazuje jednostkę, która powinna zwykle zawierać tylko jeden rodzaj tej cechy. Na przykład data urodzenia lub krajowy numer identyfikacyjny będzie w tym miejscu zawierać wartość **Tak**, a adres lub numer karty kredytowej będzie miał wartość **Nie** (ponieważ jednostka może mieć wiele adresów lub kart kredytowych).

**Stabilność.** Wskazuje wartość stabilności tej cechy (tj. czy istnieje *małe* prawdopodobieństwo, że wartość ta ulegnie zmianie w ciągu życia jednostki). Na przykład cecha daty urodzenia będzie miała wartość **Tak**, ponieważ nigdy się nie zmieni. Cecha adresu może mieć wartość **Nie**, ponieważ wartość ta może się zmienić i jest zatem mniej stabilna.

*Uwaga:* Płeć jest zwykle stabilną wartością w ciągu całego życia, jednak ze względu na częste ustalanie niepoprawnej płci w oparciu o niewłaściwe dane w domyślnej konfiguracji wybrana została wartość **Nie**.

**Anonimizacja.** Wskazuje, czy cecha została anonimizowana. Wpisy mają wartość **Tak** lub **Nie**. Więcej informacji można znaleźć w temacie “Anonimizacja cech repozytorium” na stronie 19.

Aby utworzyć nową cechę

1. Wykonaj jedną z czynności.
  - Kliknij przycisk Stwórz nową cechę (górny przycisk po prawej stronie ekranu).
  - Kliknij prawym przyciskiem myszy opcję **Wszystkie cechy** w panelu nawigacji po lewej stronie ekranu i wybierz opcję **Nowa cecha**.
2. Uzupełnij okno dialogowe Dodaj/Edytuj cechę. Więcej informacji można znaleźć w temacie “Dodawanie lub edycja cech”.

Aby zduplikować istniejącą cechę

1. W kolumnie **Cecha** w tabeli po prawej stronie ekranu wybierz cechę, która ma być zduplikowana.
2. Kliknij przycisk Duplikuj wybraną cechę (drugi przycisk po prawej stronie ekranu).
3. Uzupełnij okno dialogowe Dodaj/Edytuj cechę. Więcej informacji można znaleźć w temacie “Dodawanie lub edycja cech”.

Aby dokonać edycji istniejącej cechy

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.

1. W kolumnie **Cecha** w tabeli po prawej stronie ekranu wybierz cechę do edycji. *Uwaga:* Można edytować tylko te cechy, które zostały utworzone przez użytkownika, a nie cechy systemowe.
2. Kliknij przycisk Edytuj wybrane cechy (trzeci przycisk po prawej stronie ekranu).
3. Uzupełnij okno dialogowe Dodaj/Edytuj cechę. Więcej informacji można znaleźć w temacie “Dodawanie lub edycja cech”.

Aby usunąć istniejącą cechę

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.

1. W kolumnie **Cecha** w tabeli po prawej stronie ekranu wybierz cechę, która ma zostać usunięta. *Uwaga:* Można usuwać tylko te cechy, które zostały utworzone przez użytkownika, a nie cechy systemowe.
2. Wykonaj jedną z czynności.
  - Kliknij przycisk Usuń zaznaczoną cechę (dolny przycisk po prawej stronie ekranu).
  - Kliknij prawym przyciskiem myszy opcję **Wszystkie cechy** w panelu nawigacji po lewej stronie ekranu i wybierz opcję **Usuń**.
3. Kliknij przycisk **Kontynuuj**, aby potwierdzić usuwanie cechy.

**UWAGA:**

**Nie można cofnąć usunięcia cechy.**

## Dodawanie lub edycja cech

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.



W oknie dialogowym Dodaj/Edytuj cechę można utworzyć nową cechę repozytorium lub zduplikować istniejącą bądź dokonać jej edycji.

*Uwaga:* Jeśli istniejąca cecha jest zablokowana, nie można edytować jej szczegółów w tym oknie dialogowym.

**Typ cechy.** Etykieta wskazująca typ informacji, z którym związana jest cecha. Etykieta tworzy pierwszą część identyfikatora cechy.

**Opis.** Krótki opis tekstowy typu cechy, wyłącznie do celów informacyjnych.

**Częstość.** Wskazuje liczbę jednostek z taką samą wartością dla tej cechy. Poprawne wartości to **Jedna** (np. numer paszportu), **Kilka** (np. adres) lub **Wiele** (np. data urodzenia).

**Wyłączność.** Wskazuje jednostkę, która powinna zwykle zawierać tylko jeden rodzaj tej cechy. Na przykład data urodzenia lub krajowy numer identyfikacyjny będzie w tym miejscu zawierać wartość **Tak**, a adres lub numer karty kredytowej będzie miał wartość **Nie** (ponieważ jednostka może mieć wiele adresów lub kart kredytowych).

**Stabilność.** Wskazuje wartość stabilności tej cechy (tj. czy istnieje *małe* prawdopodobieństwo, że wartość ta ulegnie zmianie w ciągu życia jednostki). Na przykład cecha daty urodzenia będzie miała wartość **Tak**, ponieważ nigdy się nie zmieni. Cecha adresu może mieć wartość **Nie**, ponieważ wartość ta może się zmienić i jest zatem mniej stabilna.

*Uwaga:* Płeć jest zwykle stabilną wartością w ciągu całego życia, jednak ze względu na częste ustalanie niepoprawnej płci w oparciu o niewłaściwe dane w domyślnej konfiguracji wybrana została wartość **Nie**.

**Tabela elementów.** Lista elementów składających się na cechę.

- **Element.** Nazwa elementu.
- **Opis.** Krótki opis elementu.
- **Typ danych.** Typ danych, jaki można wykorzystać dla tego elementu. Dostępne typy to: łańcuch, liczba całkowita, liczba rzeczywista i data.

**Przycisk dodawania nowego elementu.** Dodaje nowy wiersz do tabeli elementów, umożliwiając zdefiniowanie nowego elementu.

**Przycisk usuwania elementu.** Usuwa wybrany wiersz z tabeli elementów. Nie można cofnąć tej czynności.

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.

**Anonimizacja.** W celu ochrony danych można wykonać anonimizację danych podczas dodawania ich do repozytorium; aby przeprowadzić anonimizację, wybierz opcję **Tak**. Więcej informacji można znaleźć w temacie “Anonimizacja cech repozytorium”.

## Anonimizacja cech repozytorium

W ramach ochrony danych można wykonać anonimizację danych dodawanych do repozytorium. Pozwoli to na zmniejszenie ryzyka ujawnienia informacji umożliwiających ustalenie tożsamości.

Podczas eksportowania anonimizowanych danych do repozytorium wymagana jest metoda anonimizacji umożliwiająca rozdzielanie jednostek na podstawie anonimizowanych danych. Na przykład, jeśli dwa rekordy danych karty kredytowej jednej osoby są anonimizowane jako "anon\_s21" i "anon\_s9271", utracona zostaje relacja pomiędzy tymi rekordami. Jeśli jednak zostanie utrzymane ogólne wewnętrzne połączenie pomiędzy rekordami, w systemie będzie można odczytać, że jedna nazwa jest skrótem drugiej nazwy.

Ogólne połączenie i identyfikatory umożliwiające łączenie anonimizowanych danych generowane są podczas tworzenia repozytorium i są unikalne dla danego repozytorium. Szyfrowane dane zapisywane są wewnętrznie, a następnie odczytywane po połączeniu strumienia z repozytorium.

Podczas konfigurowania repozytorium można określić, czy poszczególne cechy mają być anonimizowane. Jeśli cecha jest anonimizowana, anonimizowane są wszystkie jej elementy, a anonimizacja jest wykonywana bez względu na typ użycia. Więcej informacji można znaleźć w temacie “Dodawanie lub edycja cech” na stronie 18.

**Uwaga:** Należy pamiętać, aby nie anonimizować wszystkich zmiennych dla produktu SPSS Entity Analytics. Uniemożliwi to identyfikację powracających danych. Zaleca się zostawić jedną zmienną (nawet jeśli jest to numer wiersza) bez anonimizacji, tak aby kontrolować późniejsze ponowne łączenie z pierwotnymi danymi.

W kolumnie na liście cech w oknie Konfiguracja repozytorium wyświetlane są cechy, które zostały ustawione jako anonimizowane. Wpisy mają wartość **Tak** lub **Nie**.

**Uwaga:** Jeśli istniejące repozytorium będzie zawierać dane, zanim cechy zostaną anonimizowane, należy najpierw wyczyścić wszystkie dane. W przeciwnym razie nie wystąpią żadne dopasowania pomiędzy anonimizowanymi i nieanonimizowanymi cechami.

## Opracowywanie typów jednostek

**Typ jednostek** to posiadający nazwę zbiór cech repozytorium logicznie ze sobą połączonych. Na przykład typ jednostek używany ze zbiorem danych klientów może zawierać takie cechy jak imię i nazwisko, data urodzenia, płeć, adres, numer telefonu itd.

Repozytorium IBM SPSS Modeler Entity Analytics zawiera standardowy zbiór typów jednostek, do których można dodawać własne.

Sekcja Typy jednostek w oknie Konfiguracja repozytorium zawiera typy jednostek utworzone przez użytkownika. Można wykonać następujące czynności.

- Utworzyć nowy typ jednostek
- Powielić istniejący typ jednostek (np. w celu utworzenia nowego typu jednostek w oparciu o istniejący)
- Przyłączyć cechy do typu jednostek
- Usunąć cechy z typu jednostek
- Zmienić nazwę typu jednostek
- Usunąć typ jednostek

**Typ jednostek.** Nazwa wybranego typu jednostek.

**Cecha.** Lista poprawnych cech uwzględnionych w tym typie jednostek.

**Typ użycia.** (Opcja) Wskazuje różne konteksty możliwego wykorzystania tej cechy. Dwukrotnie kliknij tę kolumnę, aby dodać lub dokonać edycji typu użycia, oddzielając typy wykorzystania przecinkiem i spacją. Wartości określone w tym miejscu definiują wartości wyświetlone w węźle Eksport EA lub węźle przesyłania strumieniowego EA po kliknięciu kolumny Użycie dla cechy na karcie Wejścia. Więcej informacji można znaleźć w temacie “Opcje repozytoriów wejściowych w procesie mapowania” na stronie 15.

Informacje ogólne na temat typów użycia:

- Typy użycia są arbitralnymi etykietami.
- Typ użycia można utworzyć prawie z każdego wpisu tekstowego; nie można jednak wprowadzać spacji i niepoprawnych znaków.
- Wprowadzany tekst jest automatycznie zmieniany na tekst pisany wielką literą.
- Dozwolona jest dowolna liczba typów użycia.
- Typy użycia nie muszą być znaczące. Pomocne jednak będzie, jeśli podczas mapowania użytkownik będzie korzystał z konwencji nazw zrozumiałej dla siebie i innych użytkowników.
- Podczas mapowania wyświetlone zostanie ostrzeżenie napisane czerwoną czcionką, jeśli użytkownik zmapował elementy jednego typu, a inny użytkownik innego typu.

Zwykle wyświetlony zostaje komunikat błędu, jeśli użytkownik próbuje mapować dwie zmienne do tej samej cechy/elementu. Typy użycia to sposób na mapowanie dwóch lub większej liczby zmiennych do tej samej cechy/elementu i znalezienie dopasowań pomiędzy nimi.

Na przykład po zdefiniowaniu dwóch odrębnych cech: *HOMEADDRESS* i *WORKADDRESS* nie będzie można ich dopasować. Jeśli jedna jednostka zawiera cechę *HOMEADDRESS*, która jest taka sama jak cecha *WORKADDRESS* innej jednostki, nie będzie istniało między nimi dopasowanie, ponieważ są to różne cechy. Jednak w przypadku ponownego użycia jednej cechy z różnymi typami użycia, wówczas według rozdzielenia cecha *ADDRESS.WORK* jest taka sama jak cecha *ADDRESS.HOME*.

Typy użycia można ponownie wykorzystywać dla różnych cech lub stosować nowe; np. *HM* i *WK* jako telefon i *HOME* i *WORK* jako adres. Nie ma to znaczenia, ponieważ nie szuka się dopasowań wśród numerów telefonów w oparciu o adresy. Jeśli jednak można zachować spójność, łatwiej będzie później identyfikować i grupować zmienne.

Jeśli do jednego repozytorium wczytanych zostanie wiele typów jednostek, to typ użycia nie ma znaczenia (o ile użytkownik korzysta z tej samej cechy). Na przykład po zdefiniowaniu *WK* i *HM* jako typów użycia cechy *ADDRESS* dla typu jednostki *COMPANY*, dopasowanie zostanie przeprowadzone w oparciu o typ użycia *WORK* i *HOME* dla cechy *ADDRESS* dla *PERSON*.

Aby utworzyć nowy typ jednostek

1. Kliknij prawym przyciskiem myszy opcję **Typy jednostek** w panelu nawigacji po lewej stronie ekranu.
2. Wybierz **Nowy typ jednostki**.
3. Wprowadź unikalną nazwę typu jednostki i kliknij OK.
4. Przyłącz cechy do typu jednostki (patrz następna sekcja).

Aby przyłączyć cechy do typu jednostek

1. Wybierz typ jednostek w panelu nawigacji po lewej stronie ekranu.
2. Kliknij przycisk Załącz cechy (górny przycisk po prawej stronie ekranu).
3. Z listy dostępnych cech wybierz jedną lub więcej (użyj klawisza Ctrl-kliknięcie, aby wybrać wiele cech) i kliknij OK.

Aby usunąć cechy z typu jednostek

1. Wybierz typ jednostek w panelu nawigacji po lewej stronie ekranu.
2. Wybierz jedną lub większą liczbę cech z tabeli dołączonych cech po prawej stronie ekranu. Użyj klawisza Ctrl-kliknięcie, aby wybrać wiele cech.
3. Kliknij przycisk Odłącz cechę (dolny przycisk po prawej stronie ekranu).

Aby zduplikować istniejący typ jednostek

1. W oknie nawigacji po lewej stronie ekranu kliknij prawym przyciskiem myszy typ jednostki, który chcesz zduplikować.
2. Wybierz opcję **Duplikuj typ jednostki**.
3. Wprowadź unikalną nazwę nowego typu jednostki i kliknij OK.
4. Dołącz cechy do typu jednostki lub usuń je z typu jednostki zgodnie z wymogami (patrz wcześniejsze instrukcje).

Aby zmienić nazwę typu jednostek

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.

1. W oknie nawigacji po lewej stronie ekranu kliknij prawym przyciskiem myszy typ jednostek, którego nazwę chcesz zmienić.
2. Wybierz opcję **Zmień nazwę**.

3. Wprowadź nową nazwę typu jednostki i kliknij OK.

Aby usunąć typ jednostek

**UWAGA** Podczas edytowania, usuwania lub anonimizowania cech lub elementu cechy, gdy repozytorium zawiera już dane, należy w następnej kolejności wyczyścić repozytorium i ponownie załadować dane. Pozwoli to na uniknięcie niespójności w repozytorium.

1. W oknie nawigacji po lewej stronie ekranu kliknij prawym przyciskiem myszy typ jednostki, który chcesz usunąć.
2. Wybierz polecenie **Usuń**.
3. Kliknij **OK**, aby potwierdzić usuwanie typu jednostki.

**UWAGA:**

**Nie można cofnąć czynności usuwania typu jednostki.**

## Ustawianie progu dopasowania jednostek

W sekcji Reguły rozdzielania w oknie Konfiguracja repozytorium wybierz próg, przy którym będzie miało miejsce dopasowywanie jednostek.

Podczas tworzenia repozytorium dopasowanie ma miejsce przy domyślnym progu.

Wybierz opcję **Ustaw jako rozdzielanie agresywne**, jeśli nie można znaleźć dostatecznej liczby dopasowań w rekordach w celu wykonania rozdzielania jednostki.

Wybierz opcję **Ustaw jako rozdzielanie domyślne**, aby wrócić do domyślnego progu z innych ustawień.

Wybierz opcję **Ustaw jako rozdzielanie zachowawcze**, jeśli znajduje się zbyt wiele dopasowań.

Aby utworzyć repozytorium dla obu jednostek i relacji, wybierz opcję **Uwzględnij relacje**.

## Ponowne użycie konfiguracji repozytorium

Jeśli użytkownik chce wykorzystać wcześniejszą konfigurację w innym repozytorium, istniejącą konfigurację można wyeksportować do pliku XML i zaimportować plik do innego (docelowego) repozytorium. Ta czynność jest możliwa wyłącznie w ramach istniejących ustawień. Na przykład nie można przenieść konfiguracji repozytorium z jednej wersji IBM SPSS Modeler do innej lub z jednego typu bazy danych do innego.

Aby ponownie użyć istniejącej konfiguracji

1. Wyświetl okno Konfiguracja repozytorium dla repozytorium, którego konfiguracja zostanie użyta. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.
2. Z menu w tym oknie wybierz **Konfiguracja > Eksportuj konfigurację**.
3. W oknie dialogowym Zapisz jako wybierz nazwę i lokalizację wyeksportowanego pliku XML.
4. Wyświetl okno Konfiguracja repozytorium dla docelowego repozytorium.
5. Z menu w tym oknie wybierz **Konfiguracja > Importuj konfigurację**.
6. W oknie dialogowym Otwórz wybierz nazwę i lokalizację wcześniej wyeksportowanego pliku XML i kliknij przycisk **Otwórz**.

## Zapisywanie zmian w konfiguracji

Aby zapisać zmiany w konfiguracji

Z menu w oknie Konfiguracja repozytorium wybierz

**Plik > Zapisz.**

## Zamykanie okna konfiguracji

Aby wyjść z okna konfiguracji

Z menu w oknie Konfiguracja repozytorium wybierz

**Plik > Zakończ.**

W przypadku niezapisanych zmian w konfiguracji kliknij przycisk **OK**, aby zapisać zmiany i wyjść, lub kliknij **Anuluj**, aby wyjść bez zapisywania.

---

## Analiza rozdzielonych tożsamości (węzeł źródłowy Entity Analytics (EA))

Po wyeksportowaniu danych do repozytorium węzeł źródłowy Entity Analytics (EA) może posłużyć do przekazania rozdzielonych tożsamości do innych węzłów IBM SPSS Modeler do dalszej analizy lub przetwarzania, np. do utworzenia raportu zawierającego rozdzielone tożsamości.

Aby przeanalizować rozdzielone tożsamości

1. Dodaj węzeł źródłowy Entity Analytics (EA) do strumienia.
2. Otwórz węzeł Entity Analytics (EA).
3. Na karcie Dane wybierz repozytorium jednostek oraz jedno lub większą liczbą źródeł danych wejściowych (kliknij przycisk **Odśwież** w celu zaktualizowania liczebności rekordów). Więcej informacji można znaleźć w temacie “Wybieranie źródła danych”.
4. Dodaj kolejne węzły do strumienia w celu wykonania wymaganych czynności przetwarzania. Więcej informacji można znaleźć w temacie “Dodawanie węzłów do strumienia” na stronie 24.

## Wybieranie źródła danych

Na karcie Dane wybierz co najmniej jedno źródło danych w repozytorium, w którym będzie wykonywane dalsze przetwarzanie. Aby zaktualizować liczebność rekordów dla wyświetlonych źródeł danych, kliknij przycisk **Odśwież**.

**Repozytorium jednostek.** Wyświetla bieżące repozytorium jednostek (o ile istnieje). Aby wybrać inne repozytorium (jeśli istnieje kilka repozytoriów), należy wybrać je z listy. Aby utworzyć nowe repozytorium, wybierz przycisk **<Przełóżaj...>**, aby wyświetlić okno dialogowe, w którym można utworzyć repozytorium. Więcej informacji można znaleźć w temacie “Opcje repozytorium jednostek” na stronie 13.

**Uwzględnij rekordy ze źródeł danych.** Ta tabela uwzględnia różne źródła danych, które zostały dodane do repozytorium, oraz liczbę rekordów w każdym źródle. Zaznacz pole wyboru **Uwzględnij** dla źródeł danych, które mają być uwzględnione podczas dalszej analizy i przetwarzania. Aby wybrać lub anulować wybór wszystkich źródeł danych, kliknij odpowiednio **Uwzględnij wszystkie** lub **Wyklucz wszystkie**.

**Relacje.** Wybierz typ relacji, który ma być uwzględniony w repozytorium. Opcja ta jest dostępna wyłącznie, jeśli uwzględnianie relacji zostało przewidziane w konfiguracji repozytorium.

- **Brak relacji.** Szczegóły relacji nie są wykorzystywane.
- **Bliskie relacje.** Wybierane są tylko blisko powiązane jednostki. Bliskość relacji zależy od wielu zmiennych, np. właściwości mapowanych cech, udostępnionych cech oraz tego, czy rozwiązanie ma być zachowawcze, czy agresywne.
- **Wszystkie relacje.** Wybiera wszystkie powiązane jednostki.

**Maks. stopień separacji.** Opcja dostępna wyłącznie po wybraniu ustawień **Bliskie relacje** lub **Wszystkie relacje**. Wybierz liczbę stopni separacji, która zostanie użyta do identyfikowania relacji. Na przykład, jeśli Ann i Bob nie znają się, ale John zna zarówno Ann, jak i Boba, Ann i Bob są związani przez dwa stopnie separacji.

**Wyjściowy typ jednostki.** Jeśli repozytorium zawiera szczegółowe informacje, ta opcja wyświetla domyślnie pierwszy typ jednostki znajdujący się w repozytorium. Jeśli w repozytorium występuje więcej niż jeden typ jednostki, wybór typu jednostki zmienia cechy wyświetlone na karcie Filtr, tak aby uwzględnić cechy dla tego typu. Można wybierać dowolne typy jednostek używane w repozytorium.

## Zmiana nazwy zmiennych zawierających dane

Karta Filtr służy do zmiany nazwy dowolnych zmiennych zawierających rozdzielone tożsamości przekazanych do dalszego przetwarzania. Można również zmienić nazwę zmiennej zawierającej rozdzielone tożsamości na przykład w celu utrzymania zgodności nazwy zmiennej podczas łączenia z innym zbiorem danych.

Zmienne z pierwotnymi nazwami zostały przedstawione poniżej.

Tabela 10. Zmienne zawierające rozdzielone tożsamości

Zmienna	Opis
\$EA-ID	Identyfikator jednostek
\$EA-SRC	Znacznik źródłowy identyfikujący źródło danych, z którego pochodzi rekord
\$EA-KEY	Zmienna wyznaczona jako unikalna zmienna w pliku źródła danych

*Uwaga:* Chociaż możliwe jest użycie karty Filtr w celu odfiltrowania zmiennych, nie należy z niej korzystać w tym miejscu, ponieważ zmienne zawierające rozdzielone tożsamości to minimum konieczne do przetworzenia analizy jednostek.

## Ustawianie typu informacji dla zmiennych zawierających dane

Na karcie Typy można wyświetlić lub zmienić różne właściwości zmiennych zawierających rozdzielone tożsamości, które zostały przekazane do dalszego przetwarzania.

Właściwości, które można zmienić, są takie same jak na karcie Typy w typowym węźle SPSS Modeler. Właściwości te zostały przedstawione poniżej.

Tabela 11. Właściwości typu dla zmiennych

Właściwość	Opis
<b>Poziom</b>	Poziom pomiaru (tj. typ danych) stosowany do opisu cech danych w zmiennej.
<b>Wartości</b>	Udostępnia opcje odczytu wartości danych ze zbioru danych.
<b>Braki</b>	Służy do określenia sposobu obsługi braków danych dla zmiennej.
<b>Sprawdź</b>	Opcje walidacji w celu zapewnienia, że wartości zmiennych są zgodne z określonymi wartościami lub zakresami.
<b>Rola</b>	Określa sposób użycia zmiennej w przypadku przekazania danych do węzła modelowania lub modelu użytkowego.

## Dodawanie węzłów do strumienia

Istnieje możliwość dodania do strumienia różnych węzłów SPSS Modeler w celu przeprowadzenia analizy lub przetworzenia wyników z węzła źródłowego Entity Analytics (EA). Na przykład można dodać jeden lub większą liczbę poniższych węzłów.

- Węzeł agregacji lub węzeł powtórzeń w celu podsumowania wyników, które mogą mieć bardzo duży rozmiar
- Węzeł selekcji w celu wybrania podzbioru wyników
- Węzeł tabeli w celu wyświetlenia wyników z węzła źródłowego Entity Analytics (EA)

- Węzeł raportu w celu wydrukowania wyników w formie raportu
- Węzeł eksportu SPSS Modeler w celu wyeksportowania wyników do innego formatu, np. arkusza kalkulacyjnego lub bazy danych

Więcej informacji znajduje się w sekcji poświęconej węzłom operacji rejestrowania, wyników i eksportu w podręczniku *IBM SPSS Modeler — węzły źródłowe, procesowe i wyników*.

---

## Porównywanie nowych obserwacji w oparciu o repozytorium (węzeł przesyłania strumieniowego EA)

Po wykonaniu rozdzielania tożsamości w repozytorium można za pomocą węzła przesyłania strumieniowego EA porównać nowe obserwacje dostrzeżone następnie w zawartości repozytorium. Ten węzeł przetwarza rekordy z nowego źródła danych, porównuje je z rozdzielonymi tożsamościami znajdującymi się już w repozytorium i przekazuje dopasowane rekordy do dalszego przetwarzania. Dopasowania można skonfigurować jako dokładne lub swobodnie związane z istniejącymi jednostkami.

Podobnie jak węzeł Eksport EA węzeł przesyłania strumieniowego EA potrzebuje jednego węzła źródłowego SPSS Modeler dostarczającego dane wejściowe. Węzeł przesyłania strumieniowego EA różni się jednak w następujący sposób. Węzeł eksportu generuje rekordy dla wszystkich jednostek związanych z rekordami wejściowymi, a węzeł przesyłania strumieniowego EA generuje rekordy wyłącznie dla tych jednostek, które są związane z już rozdzielonymi tożsamościami w repozytorium. Więcej informacji można znaleźć w temacie “Wyniki z węzła przesyłania strumieniowego EA” na stronie 28.

Aby porównać nowe obserwacje w oparciu o repozytorium

1. Połącz się ze źródłem danych zawierającego nowe rekordy, które mają być porównane z istniejącymi jednostkami. Więcej informacji można znaleźć w temacie “Łączenie ze źródłem danych” na stronie 12.
2. Na karcie Rekordy dołącz węzeł przesyłania strumieniowego EA do węzła źródła danych.
3. Aby otworzyć okno dialogowe, dwukrotnie kliknij węzeł eksportu Entity Analytics.
4. Kliknij listę **Repozytorium jednostek**.
5. Kliknij przycisk **<Przeglądaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
6. W oknie dialogowym Instancje rozdzielania jednostek kliknij pole Nazwa repozytorium.
7. Kliknij nazwę repozytorium, które zostanie użyte.
8. Wprowadź nazwę użytkownika i hasło dla tego repozytorium i kliknij **Połącz**. Kliknij przycisk **OK** po połączeniu repozytorium.
9. W oknie dialogowym Przesyłanie strumieniowe EA wybierz Typ jednostki, który ma być mapowany. Więcej informacji można znaleźć w temacie “Opracowywanie typów jednostek” na stronie 20.
10. Zmapuj zmienne wejściowe w źródle danych do cech w repozytorium. Więcej informacji można znaleźć w temacie “Mapowanie zmiennych wejściowych do cech (węzeł przesyłania strumieniowego EA)” na stronie 26.
11. Opcjonalnie można skonfigurować rekordy w repozytorium w czasie rzeczywistym podczas oceny danych. Więcej informacji można znaleźć w temacie “Mapowanie zmiennych wejściowych do cech (węzeł przesyłania strumieniowego EA)” na stronie 26.
12. Kliknij kartę **Zmienne wyjściowe**, aby wyświetlić szczegóły różnych źródeł danych wprowadzonych do repozytorium i ustawić kryteria wyboru w celu pobrania istniejących jednostek. Więcej informacji można znaleźć w temacie “Wyświetlanie mapowań zmiennych i źródeł danych (węzeł przesyłania strumieniowego EA)” na stronie 27.
13. Kliknij kartę **Filtr**, aby wyświetlić szczegóły zmiennych wejściowych i cech przechowywanych w repozytorium. Cechy, które nie zostały zmapowane w węzle, zostaną domyślnie odfiltrowane; w razie potrzeby można jednak zmienić to ustawienie.
14. Kliknij przycisk **OK**, kiedy węzeł zostanie poprawnie skonfigurowany.
15. Dołącz węzeł tabeli do węzła przesyłania strumieniowego EA i uruchom strumień.

Okno wyników węzła tabeli zawiera listę wszystkich pobranych jednostek zgodnych z nowymi rekordami w źródle danych. Zmienne wyjściowe mają dodany przedrostek **SEA-**. Więcej informacji można znaleźć w temacie “Wyniki z węzła przesyłania strumieniowego EA” na stronie 28.

**Uwaga:** Podczas uruchamiania węzła przesyłania strumieniowego EA można napotkać na błąd o treści **Niepoprawną liczbę zmiennych odkryto w modelu danych.** Ten błąd może wystąpić na skutek edycji konfiguracji repozytorium po utworzeniu węzła przesyłania strumieniowego EA. Edytowanie konfiguracji w tej sytuacji może spowodować zmianę liczby i nazw zmiennych wychodzących z węzła. Aby rozwiązać ten problem, otwórz węzeł przesyłania strumieniowego EA i kliknij przycisk **Odśwież**. Spowoduje to ponownie przeliczenie liczby i nazw zmiennych wyjściowych.

## Mapowanie zmiennych wejściowych do cech (węzeł przesyłania strumieniowego EA)

Karta Wejścia zawiera opcje mapowania zmiennych w danych wejściowych tego węzła wejściowego na cechy repozytorium. Ustaw przypisania mapowania na tej karcie lub wybierz kartę **Widok**, aby wyświetlić szczegóły wszystkich źródeł danych w repozytorium, a następnie kliknij przycisk **OK**.

Jeśli zbiór mapowań jest już zapisany w pliku XML, można je wykorzystać, klikając opcję **Importuj mapowanie**.

**Repozytorium jednostek.** Wyświetla bieżące repozytorium jednostek (o ile istnieje). Aby wybrać inne repozytorium (jeśli istnieje kilka repozytoriów), należy wybrać je z listy. Aby utworzyć nowe repozytorium, wybierz przycisk **<Przełóżaj...>**, aby wyświetlić okno dialogowe, w którym można utworzyć repozytorium. Więcej informacji można znaleźć w temacie “Opcje repozytorium jednostek” na stronie 13.

**Mapuj na typ jednostki.** Lista typów jednostek (tj. zbiory cech) zdefiniowanych w repozytorium. Wybierz jeden z listy lub wybierz opcję **<Dodaj nowy typ jednostki...>**, aby wyświetlić okno konfiguracji repozytorium, w którym można zdefiniować nowy typ jednostek. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.

**Zachowaj wyszukiwania.** Wybierz tę opcję, aby zaktualizować rekordy w repozytorium w czasie rzeczywistym podczas oceny danych.

**Znacznik źródłowy.** Pozycja dostępna wyłącznie po wyborze opcji **Zachowaj wyszukiwania**. Lista znaczników wskazujących źródła danych aktualnie znanych w repozytorium. Wybierz jeden z listy lub wybierz opcję **<Dodaj nowy znacznik źródła...>**, aby utworzyć znacznik dla nowego źródła danych.

**Unikalny klucz.** Pozycja dostępna wyłącznie po wyborze opcji **Zachowaj wyszukiwania**. Zmienna wejściowa używana dla unikalnych identyfikatorów rekordów danych.

**Tabela mapowania.** W tej tabeli można mapować każdą zmienną wejściową do odpowiedniej cechy w repozytorium. Jeśli odpowiednia cecha nie istnieje w wybranym typie jednostki, nową cechę można utworzyć w tym miejscu.

- **Zmienna.** Zbiór zmiennych wejściowych w wybranym źródle danych. Każda zmienna zawiera ikonę wskazującą poziom pomiaru (tj. typ danych) dla zmiennej.
- **Mapowanie do cechy.** Aby zmapować zmienną do cechy, dwukrotnie kliknij tę kolumnę (lub naciśnij klawisz spacji) w wierszu zmiennej i wybierz cechę z listy. Jeśli odpowiednia cecha nie jest dostępna, wybierz opcję **<Dodaj nową cechę...>**, aby wyświetlić okno konfiguracji repozytorium, w którym można zdefiniować nową cechę dla tego typu jednostki. Więcej informacji można znaleźć w temacie “Konfigurowanie repozytorium jednostek” na stronie 16.
- **Użycie.** Wskazuje kontekst poszczególnych zmiennych, jeśli możliwy jest więcej niż jeden kontekst, np. numer telefonu domowego i służbowego. Więcej informacji można znaleźć w temacie “Opracowywanie typów jednostek” na stronie 20.



**Importuj mapowanie.** Importuje wcześniej wyeksportowany zbiór mapowań zmiennej do cechy z zewnętrznego pliku XML. Opcja ta może być przydatna w przypadku różnych źródeł danych z tymi samymi wymaganiami. Pozwala na uniknięcie powtórnego definiowania tych samych mapowań dla różnych źródeł.

**Eksportuj mapowanie.** Eksportuje do zewnętrznego pliku XML zbiór mapowań zmiennej do cechy wyświetlony w tabeli mapowania.

## Wyświetlanie mapowań zmiennych i źródeł danych (węzeł przesyłania strumieniowego EA)

Karta Wyniki zawiera szczegóły różnych źródeł danych, które zostały przekazane do repozytorium. Są to źródła danych będące podstawą przetwarzania danych wejściowych przekazanych do tego węzła. Celem tej operacji jest wyszukiwanie i pobieranie dopasowanych jednostek. Kliknij przycisk **Odśwież**, aby zaktualizować liczebność rekordów.

**Uwzględnij dopasowania ze źródeł danych.** Ta tabela uwzględni różne źródła danych dostępne w repozytorium oraz liczbę rekordów w każdym źródle.

**Zgodności.** Te opcje określają, z jaką dokładnością informacje o mapowaniu zmiennej do pola podane na karcie Wejścia mają być dopasowywane na podstawie potencjalnych rekordów (tj. całej zawartości repozytorium). Im dokładniejsze kryterium dopasowania, tym mniej jednostek zostanie pobranych.

**Uwaga:** W przypadku znalezienia więcej niż 20 dopasowań zwróconych zostanie tylko pierwszych 20.

- **Uwzględnij tylko dokładne dopasowania.** To jest najdokładniejsze kryterium mapowania i generuje najmniejszą liczbę wybranych rekordów. Za pomocą tej opcji można zwracać tylko te jednostki, które są uznawane za dokładne dopasowania.
- **Uwzględnij możliwe dopasowania.** To ustawienie służy do zwracania zarówno dopasowanych jednostek, jak i jednostek mających ten sam identyfikator (jednostki z cechami skonfigurowanymi z wartością częstości Jedna, np. dopasowane numery kart kredytowych, numery identyfikacji podatkowej itd.).
- **Uwzględnij wszystkie dopasowania.** Ta opcja umożliwia wyświetlenie możliwie największej liczby jednostek w repozytorium mających wspólne cechy. To jest najluźniejsze kryterium mapowania i daje najmniejszą liczbę wybranych rekordów. Ta opcja zwraca dokładne dopasowania i jednostki współdzielące prawie każdą cechę (zwykle cechy z wartością częstości Jedna lub Kilka). Na przykład uwzględnione zostaną obie jednostki z tym samym numerem identyfikacji podatkowej i jednostki z podobnym adresem.

**Relacje.** Opcja dostępna wyłącznie, jeśli uwzględnianie relacji zostało przewidziane w konfiguracji repozytorium. Wybierz typ relacji, który ma być uwzględniony w danych wynikowych.

- **Brak relacji.** Szczegóły relacji nie są wykorzystywane.
- **Bliskie relacje.** Wybierane są tylko blisko powiązane jednostki. Bliskość relacji zależy od wielu zmiennych, np. właściwości mapowanych cech, udostępnionych cech oraz tego, czy rozwiązanie ma być zachowawcze, czy agresywne.
- **Wszystkie relacje.** Wybiera wszystkie powiązane jednostki.

**Maks. stopień separacji.** Opcja dostępna wyłącznie po wybraniu ustawień **Bliskie relacje** lub **Wszystkie relacje**. Wybierz liczbę stopni separacji, która zostanie użyta do identyfikowania relacji. Na przykład, jeśli Ann i Bob nie znają się, ale John zna zarówno Ann, jak i Boba, Ann i Bob są związani przez dwa stopnie separacji.

**Wyjściowy typ jednostki.** Jeśli repozytorium zawiera szczegółowe informacje, ta opcja wyświetla domyślnie pierwszy typ jednostki znajdujący się w repozytorium. Jeśli w repozytorium występuje więcej niż jeden typ jednostki, wybór typu jednostki zmienia cechy wyświetlone na karcie Filtr, tak aby uwzględnić cechy dla tego typu. Można wybierać dowolne typy jednostek używane w repozytorium.

## Wyniki z węzła przesyłania strumieniowego EA

Wyniki z węzła przesyłania strumieniowego EA składają się z następujących zmiennych dla każdego pobranego rekordu.

Zmienna	Opis
<i>Field1</i> [, <i>Field2</i> [, ... <i>FieldN</i> ]]	Zmienne ze źródła danych zawierającego nowe rekordy.
\$SEA-ID	Identyfikator jednostek dla tego rekordu w repozytorium.
\$SEA-SRC	Znacznik źródłowy identyfikujący źródło danych, z którego pochodzi ten rekord.
\$SEA-KEY	Wartość unikalnego klucza dla tego rekordu w pliku źródła danych.
\$SEA-SC	Zmienna wskazująca bliskość dopasowania pomiędzy tym rekordem a obserwowaną jednostką w repozytorium; wartość od 1.0 (słabe dopasowanie) do 10.0 (dobre dopasowanie).
\$SEA-Feature1[, \$SEA-Feature2[, ... \$SEA-FeatureN]]	Wartości zmapowanych cech dla tego rekordu w repozytorium.

Jeśli zmienne relacji są włączone w repozytorium i stopień separacji wynosi więcej niż zero na karcie Wyniki, to wyniki z węzła przesyłania strumieniowego EA również zawierają następujące zmienne dla każdego pobranego rekordu.

Zmienna	Opis
\$SEA-DEGREE	Stopień separacji
\$SEA-PARENT	Identyfikator rekordu, od którego obliczana jest separacja
\$SEA-CHILD	Identyfikator rekordu, do którego obliczana jest separacja
\$SEA-RULE	

---

## Korzystanie z programu IBM SPSS Modeler Entity Analytics z innymi produktami IBM SPSS

Dostępne są instalatory umożliwiające korzystanie z programu IBM SPSS Modeler Entity Analytics razem z następującymi produktami:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch dla systemu Windows
- IBM SPSS Modeler Solution Publisher

Aby możliwe było korzystanie z funkcji IBM SPSS Modeler Entity Analytics z tymi produktami, należy w pierwszej kolejności uruchomić instalatory. Więcej informacji znajduje się w przewodniku *Podręcznik instalowania programu IBM SPSS Modeler Premium*.

Po instalacji należy za pomocą klienta IBM SPSS Deployment Manager utworzyć definicję serwera repozytorium analizy jednostek. Czynność ta wymagana jest w celu korzystania ze strumienia IBM SPSS Modeler zawierającego węzeł analizy jednostek w zadaniu IBM SPSS Collaboration and Deployment Services (innymi słowy, w celu uruchomienia strumienia Entity Analytics w IBM SPSS Collaboration and Deployment Services). Definicja serwera musi być zgodna z nazwą repozytorium w strumieniu; za pomocą tej definicji do strumienia przekazywane są informacje na temat lokalizacji repozytorium oraz wymagane dane dotyczące połączenia.

---

## Zadania administracyjne

W przypadku repozytoriów utworzonych za pomocą węzła Entity Analytics nowa usługa bazy danych jest tworzona za pomocą produktu IBM DB2. Dostępnych jest kilka zadań administracyjnych związanych z DB2; zadania te są zwykle wykonywane przez administratora bazy danych lub administratora systemu. Do zadań tych należą:

- Konfiguracja przypisań portów
- Zarządzanie danymi uwierzytelniającymi administratora dla bazy danych repozytorium

Inne zadania administracyjne, których wykonanie może być konieczne, dotyczą wszystkich repozytoriów i uwzględniają następujące kroki.

- Przenoszenie repozytorium do innego katalogu składowania
- Ustawianie właściwości strumienia dla zmiennych daty/czasu i znacznika czasu
- Dostosowywanie ustawień limitu czasu
- Uruchomienie programu IBM SPSS Modeler Entity Analytics za pomocą klienta SPSS Modeler i SPSS Modeler Server w tym samym systemie Windows
- Czyszczenie repozytorium jednostek
- Usuwanie repozytorium jednostek
- Usuwanie repozytorium bez połączenia

## Konfiguracja przypisań portów

Każdej usłudze bazy danych DB2 należy przydzielić port, który nie może być przydzielony do innych usług uruchomionych na komputerze. Usługi bazy danych zlokalizowane są na tej samej maszynie, na której uruchomiony jest program IBM SPSS Modeler Server (lub kiedy program IBM SPSS Modeler jest używany bez połączenia z IBM SPSS Modeler Server, maszyna, na której uruchomiony jest IBM SPSS Modeler).

Domyślnie w programie Entity Analytics porty przypisywane są w zakresie od 1320 do 1520, rozpoczynając od portu 1320 dla pierwszego utworzonego repozytorium. W przypadku konfliktu można skonfigurować przypisanie portów poprzez edycję pliku: `<modeler_server_installation_path>/ext/bin/pasw.entityanalytics/ea.cfg` i wybranie odpowiednich wartości dla ustawień `min_port` i `max_port`. Domyślna zawartość tego pliku została przedstawiona poniżej:

```
# port range configuration for entity analytics
#
#   this port range controls which ports DB2 databases
#   (created to store Entity Analytics Repositories in)
#   may use.  Configure this if the default port range will
#   introduce a conflict on your system.
#
# default min_port = 1320
# default max_port = 1520
min_port, 1320
max_port, 1520
```

## Zarządzanie danymi uwierzytelniającymi administratora dla bazy danych repozytorium

Nazwy użytkownika i hasło administratora do bazy danych DB2 udostępniającej repozytorium jednostek są definiowane podczas tworzenia repozytorium. Jeśli znane są bieżące dane uwierzytelniające, można zmienić te elementy za pomocą edytora DB2 SQL.

### Aby uruchomić edytor DB2 SQL

1. Otwórz wiersz komend na komputerze klienckim.
2. Wprowadź:  

```
cd modeler_install_dir\ext\bin\pasw.entityanalytics\DB2\bin
```

gdzie *modeler\_install\_dir* to katalog, w którym zainstalowano SPSS Modeler.
3. Wprowadź:  

```
solsql -c "C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\version\EA\repositories\repos_name"
```

gdzie *version* to numer wersji instalacji SPSS Modeler, a *repos\_name* to nazwa repozytorium.
4. Po wyświetleniu monitu wprowadź aktualną nazwę użytkownika i hasło administratora bazy danych. Zostanie wyświetlony monit `solsql>`.

### Aby zmienić hasło administratora bazy danych

1. Po wyświetleniu monitu `solsql>` wprowadź:  

```
alter user username identified by password;  
commit work;
```

gdzie *username* to aktualna nazwa użytkownika administratora, a *password* to nowe hasło.
2. Wprowadź polecenie `exit`, aby zamknąć edytor.
3. Ponownie uruchom klienta SPSS Modeler.

Więcej informacji o innych zadaniach administracyjnych w bazie danych DB2 znajduje się w dokumentacji poświęconej odpowiedniej wersji IBM DB2 na stronie <http://publib.boulder.ibm.com/>.

## Przenoszenie repozytorium do innego katalogu składowania

Pliki repozytorium są domyślnie przechowywane w katalogu o nazwie *EA* w następujących lokalizacjach:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\*version*\EA (systemy Windows)
- *modeler\_install\_directory*/ext/bin/pasw.entityanalytics/EA (systemy UNIX)

Pliki służące do przechowywania repozytorium mogą mieć bardzo duży rozmiar, dlatego konieczne może być przeniesienie ich na inny dysk lub partycję w celu zwolnienia miejsca.

Aby przenieść repozytorium do innego katalogu

1. Wyjdź z programu SPSS Modeler.
2. Przenieś katalog *EA* z pierwotnej lokalizacji (wskazanej powyżej) do nowej lokalizacji. Na przykład w systemie Windows katalog można przenieść do nowej lokalizacji, tj. *F:\data\EA*.
3. Dokonaj edycji pliku `<modeler_server_installation_path>/ext/bin/pasw.entityanalytics/ea.cfg`, aby dodać następujące opcje:  

```
repository_data_directory, new_location
```

gdzie *new\_location* to katalog, do którego został przeniesiony katalog EA, np. *F:\data\EA*.

## Ustawianie właściwości strumienia dla zmiennych daty/czasu i znacznika czasu

Jeśli źródło danych zawiera zmienne uwzględniające dane daty/czasu lub znacznika czasu, należy dopilnować, aby właściwości odpowiedniego strumienia zostały ustawione w formacie rozpoznawanym przez produkt IBM SPSS Modeler Entity Analytics.

Aby ustawić format właściwości strumienia

1. Z menu głównego SPSS Modeler wybierz:  
**Narzędzia > Właściwości strumienia > Opcje.**
2. Wybierz **Data/Czas**.
3. Ustaw **Format daty** jako **RRRR-MM-DD**.
4. Ustaw **Format czasu** jako **GG:MM:SS**.
5. Kliknij przycisk **OK**.

## Dostosowywanie ustawień limitu czasu

Jeśli w wolnych lub przeciążonych systemach występują błędy podczas tworzenia lub dostępu do repozytoriów, konieczne może być zwiększenie limitu czasu dla uruchamiania i zatrzymywania silnika analizy jednostek lub serwera bazy danych analizy jednostek.

Aby dostosować limit czasu dla silnika analizy jednostek

1. Wyjdź z programu SPSS Modeler.
2. Dokonaj edycji pliku `<modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg`, aby zwiększyć wartość następującej opcji:

`timeout, value`

gdzie *value* to wartość limitu czasu w sekundach dla silnika analizy jednostek (wartość domyślna to 60).

Aby dostosować limit czasu dla serwera bazy danych analizy jednostek (tylko DB2)

1. Wyjdź z programu SPSS Modeler.
2. Dokonaj edycji pliku `<modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg`, aby zwiększyć wartość następującej opcji:

`timeout, value`

gdzie *value* to wartość limitu czasu w sekundach dla serwera bazy danych DB2 analizy jednostek (wartość domyślna to 100).

## Uruchomienie programu IBM SPSS Modeler Entity Analytics za pomocą klienta SPSS Modeler i SPSS Modeler Server w tym samym systemie Windows

Po zainstalowaniu programu IBM SPSS Modeler Entity Analytics zarówno na kliencie SPSS Modeler, jaki i SPSS Modeler Server w tym samym systemie Windows, klient i serwer będą domyślnie współużytkować to samo repozytorium. Aby udostępnić odrębne repozytoria, należy dokonać edycji pliku *ea.cfg* w **jednym** z systemów i skonfigurować go tak, aby korzystał z innego zakresu portów i folderu repozytorium.

*Uwaga:* Procedurę tę należy wykonać w szczególności w przypadku użycia 32-bitowego klienta SPSS Modeler i 64-bitowego serwera SPSS Modeler Server (lub odwrotnie).

1. Otwórz plik `<modeler [server] installation path>/ext/bin/pasw.entityanalytics/ea.cfg` w celu edycji.
2. Zmień ustawienia `min_port` i `max_port`, wybierając inne porty dla drugiego systemu. Więcej informacji można znaleźć w temacie “Konfiguracja przypisań portów” na stronie 29.
3. Zmień ustawienia `repository_data_directory`, wybierając inny katalog dla drugiego systemu.

4. Zapisz i zamknij plik *ea.cfg*.

## Czyszczenie repozytorium jednostek

Aby usunąć rekordy danych z repozytorium jednostek, ale zachować informacje o konfiguracji, można usunąć dane z repozytorium.

Aby usunąć wszystkie dane z repozytorium:

1. Otwórz węzeł Entity Analytics.
2. Kliknij listę **Repozytorium jednostek**.
3. Kliknij przycisk **<Przeglądaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
4. W oknie dialogowym Instancje rozdzielania jednostek kliknij listę **Nazwa repozytorium**.
5. Wybierz repozytorium, z którego zostaną usunięte dane.
6. Jeśli nie zostało nawiązane połączenie, wprowadź nazwę użytkownika i hasło dla administratora i kliknij polecenie **Połącz**.
7. Jeśli przycisk **Wyczyść wszystko** jest aktywny, kliknij go.
8. W oknie dialogowym czyszczenia wszystkich źródeł danych kliknij polecenie **Wyczyść**, aby potwierdzić czyszczenie repozytorium.

## Usuwanie nieużywanych źródeł danych z repozytorium

Nieużywane lub niepotrzebne źródło danych można usunąć z repozytorium jednostek. Do usunięcia można wybrać jedno lub większą liczbę źródeł danych.

Aby usunąć wybrane źródło danych z repozytorium:

1. Otwórz węzeł Entity Analytics.
2. Kliknij listę **Repozytorium jednostek**.
3. Kliknij przycisk **<Przeglądaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
4. W oknie dialogowym Instancje rozdzielania jednostek kliknij listę **Nazwa repozytorium**.
5. Wybierz repozytorium, z którego będzie usuwane źródło danych.
6. Jeśli nie zostało nawiązane połączenie, wprowadź nazwę użytkownika i hasło dla administratora i kliknij polecenie **Połącz**.
7. Na liście **Zarządzaj repozytorium** wybierz źródło danych do usunięcia. W razie potrzeby użyj klawisza Ctrl i kliknięcia, aby wybrać dodatkowe źródła danych.
8. Jeśli przycisk **Usuń nieużywane** jest aktywny, kliknij go.
9. W oknie dialogowym usuwania nieużywanych źródeł danych kliknij polecenie **Wyczyść**, aby potwierdzić czyszczenie repozytorium.

## Usuwanie repozytorium jednostek

Jeśli repozytorium nie jest już potrzebne, można całkowicie je usunąć.

### UWAGA:

**Ta opcja działa dokładnie zgodnie ze swoją nazwą. Nie można cofnąć tej czynności.**

W przypadku wątpliwości należy użyć przycisku **Wyczyść**, aby usunąć wszystkie dane źródłowe. Ta czynność nie powoduje usunięcia konfiguracji repozytorium. Więcej informacji można znaleźć w temacie “Czyszczenie repozytorium jednostek”.

**Uwaga:** W procedurze założono możliwość połączenia się z repozytorium z produktu SPSS Modeler. Użytkownik zna również nazwę użytkownika i hasło administratora do bazy danych, w której znajduje się repozytorium. Jeśli nie stanowi to problemu, wykonaj następującą procedurę usuwania repozytorium bez połączenia z tym repozytorium. Więcej informacji można znaleźć w temacie “Usuwanie repozytorium bez połączenia” na stronie 33.

## Aby usunąć repozytorium

1. Otwórz węzeł Entity Analytics.
2. Kliknij listę **Repozytorium jednostek**.
3. Kliknij przycisk **<Przeglądaj...>**, aby wyświetlić okno dialogowe Instancje rozdzielania jednostek.
4. W oknie dialogowym Instancje rozdzielania jednostek kliknij listę **Nazwa repozytorium**.
5. Wybierz repozytorium, które zostanie usunięte.
6. Jeśli nie zostało nawiązane połączenie, wprowadź nazwę użytkownika i hasło dla administratora i kliknij polecenie **Połącz**.
7. Jeśli przycisk **Usuń całe repozytorium** jest aktywny, kliknij go.
8. Kliknij przycisk **Usuń**, aby potwierdzić usuwanie repozytorium.
9. Kliknij przycisk **OK**, aby potwierdzić pomyślne usunięcie.

## Usuwanie repozytorium bez połączenia

Poniższa procedura służy do usuwania całego repozytorium jednostek w przypadku braku połączenia z tym repozytorium. Sytuacja taka występuje przez problemy łączności z SPSS Modeler lub gdy użytkownik zapomniał nazwę użytkownika lub hasło.

Procedurę tę należy wykonać na maszynie, na której znajduje się baza danych repozytorium.

## Systemy Windows

1. Otwórz okno Wiersz komend.
2. Wprowadź:

```
cd modeler_install_dir
cd ext\bin\pasw.entityanalytics
delete_repository.bat repos_name
```

gdzie *modeler\_install\_dir* to katalog, w którym SPSS Modeler jest zainstalowane, a *repos\_name* to nazwa repozytorium.

**Uwaga:** W nazwie repozytorium rozróżniana jest wielkość liter.
3. Kontynuuj od kroku "Kończenie procedury" znajdującego się w dalszej części tej sekcji.

## Systemy UNIX

1. Otwórz powłokę.
2. Wprowadź:

```
cd modeler_server_install_dir
cd ext/bin/pasw.entityanalytics
./delete_repository.sh repos_name
```

gdzie *modeler\_server\_install\_dir* to katalog, gdzie zainstalowany jest program SPSS Modeler Server, a nazwa *repos\_name* to nazwa repozytorium..

**Uwaga:** W nazwie repozytorium rozróżniana jest wielkość liter.

## Kończenie procedury (wszystkie systemy)

1. Po wyświetleniu monitu potwierdź usuwanie poprzez wprowadzenie symbolu Y.
2. Usuń katalog o takiej samej nazwie co usunięte repozytorium. Jeśli nie można usunąć katalogu, należy ponownie uruchomić komputer i spróbować ponownie.





---

## Rozdział 4. Przebieg analizy jednostek

---

### Informacje o tym przykładzie

W tym przykładzie pokazany zostanie wpływ, jaki ma analiza jednostek na poprawę już zdumiewających wyników uzyskanych dzięki programowi SPSS Modeler.

W tym przykładzie wykorzystany jest strumień *loan\_entity\_analytics.str*, który dotyczy pliku danych *loan\_applications.csv*. Pliki te dostępne są w katalogu *Demos* w dowolnej instalacji programu SPSS Modeler z zainstalowanym produktem SPSS Entity Analytics. Dostęp do katalogu *Demos* można uzyskać z grupy programu SPSS Modeler w menu Start systemu Windows. Plik *loan\_entity\_analytics.str* znajduje się w katalogu *Entity\_Analytics*.

**Uwaga:** Przed wykonaniem tego przykładowego strumienia należy utworzyć repozytorium w swoim systemie. Repozytorium należy utworzyć przed podjęciem kolejnych czynności w tym przykładzie. Więcej informacji można znaleźć w temacie “Tworzenie repozytorium” na stronie 12.

Zajmijmy się zatem znajomą sytuacją – kierownicy banku zastanawiają się, czy klienci, którzy złożyli wnioski kredytowe będą spłacać swoje zobowiązania. Dział IT banku już od dłuższego czasu korzysta z programu SPSS Modeler. Pracownicy już wcześniej utworzyli strumień i zbudowali model predykcyjny na podstawie istniejących danych dotyczących około 700 kredytów, jakich bank dotychczas udzielił. Kredyty te zostały spłacone bądź wystąpiły dla nich zaległości w spłacie.

---

### Pierwotny model

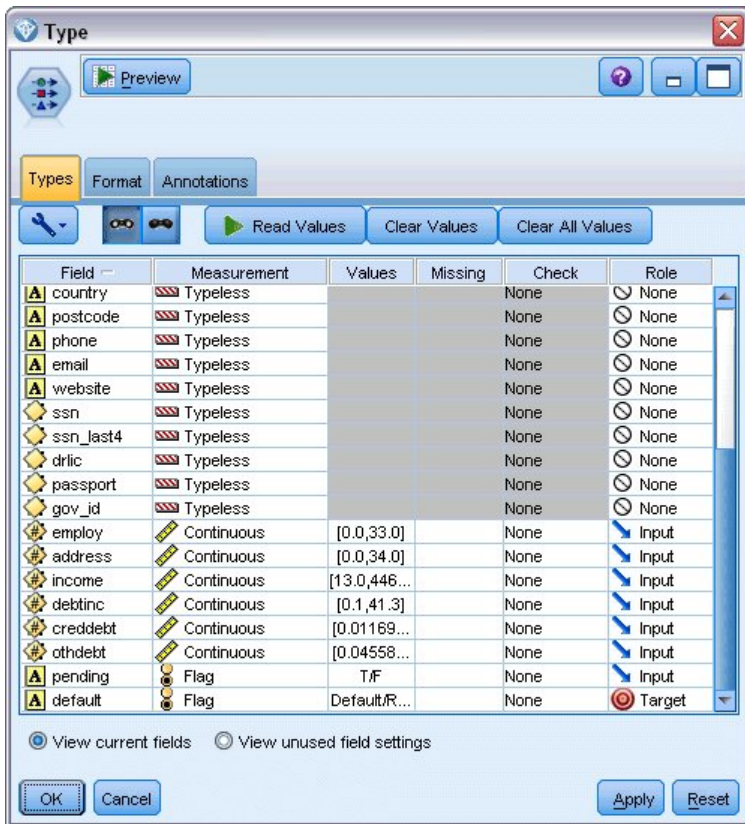
Oto w jaki sposób pracownicy banku zbudowali własny model i do jakich wniosków doszli na podstawie tego modelu.



Rysunek 2. Początkowy strumień z węzłem modelowania

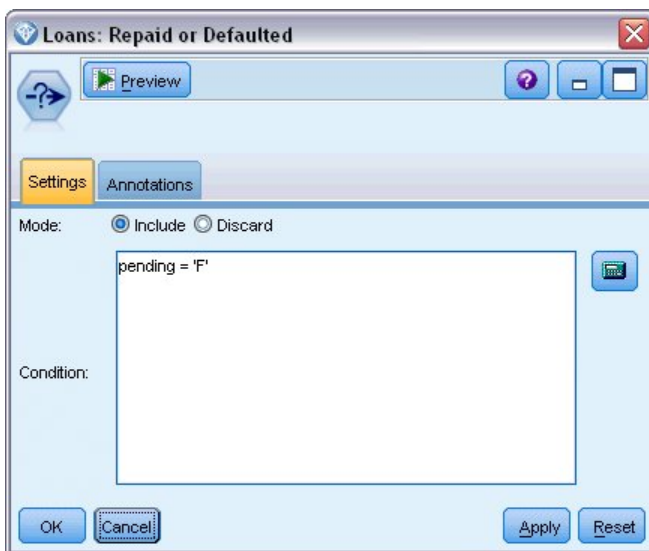
Zbiór danych *loan\_applications.csv* zawiera zarówno dane dotyczące udzielonych kredytów, jak i szczegóły dotyczące 150 klientów, których wnioski kredytowe nadal oczekują na decyzję kredytową. Łącznie dostępnych jest 850 rekordów.

Nie wszystkie zmienne ze zbioru danych będą pomocne przy tworzeniu predykcji, np. zmienne nazwy można zignorować. Węzeł Typy filtruje zmienne, które można zignorować. Wystarczy wybrać wartość **Brak** jako rolę zmiennych. Dla zmiennych, które zostaną użyte w predykcji, należy ustawić rolę **Dane wejściowe**, a dla zmiennej, której wartość będzie przewidywana w modelu, należy ustawić rolę **Zmienna przewidywana**.



Rysunek 3. Role zmiennych ustawiane w węźle Typy

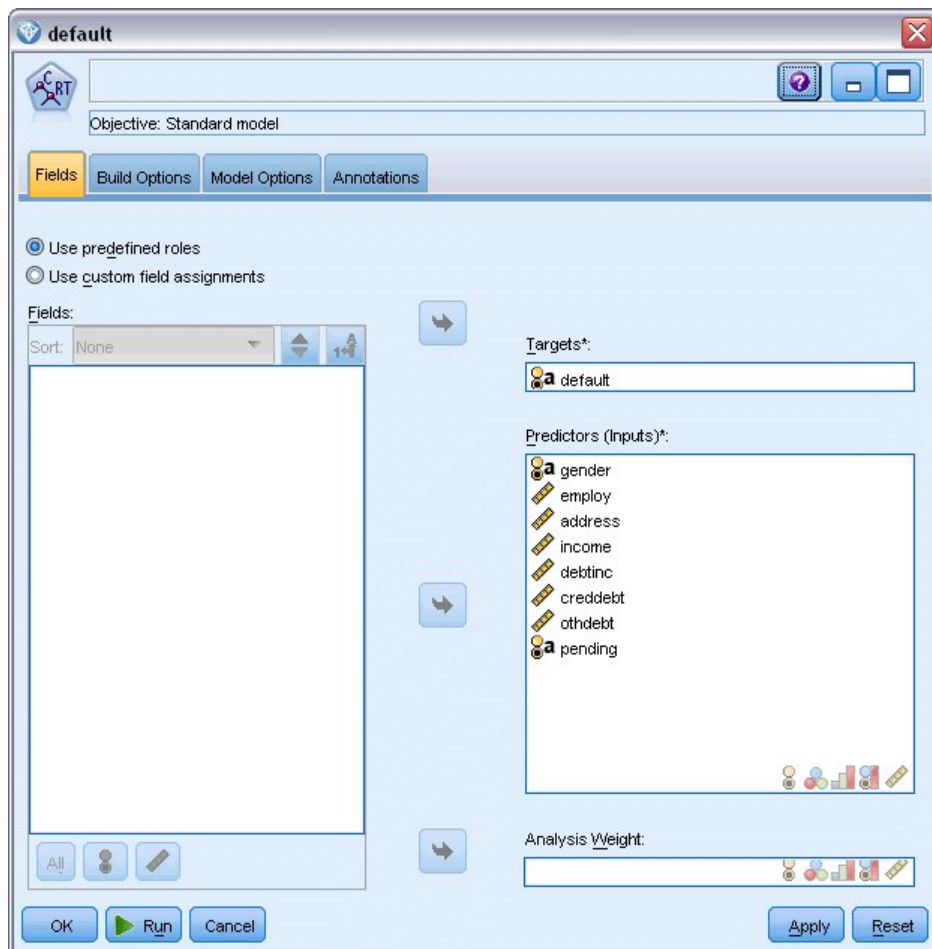
Model tworzy predykcje wyłącznie na podstawie wcześniejszych danych, dlatego strumień uwzględni węzeł selekcji zawierający kredyty, które *nie* zostały oznaczone jako oczekujące. W ten sposób zostało odrzuconych 150 oczekujących wniosków kredytowych.



Rysunek 4. Odrzucanie oczekujących wniosków kredytowych

Po odrzuceniu oczekujących wniosków kredytowych do węzła modelowania zostało przekazanych 700 kredytów, które zostały spłacone lub dla których wystąpiły opóźnienia w spłacie. W celu utworzenia poprawnego modelu bank mógł użyć jednego z wielu algorytmów SPSS Modeler. W tym przypadku zastosowano węzeł Drzewo C&R, który zostanie

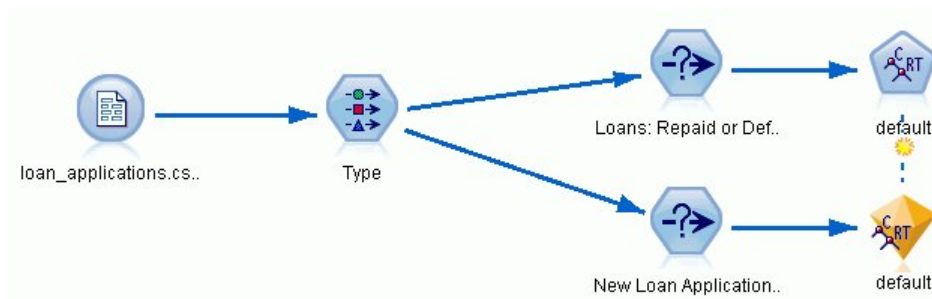
użyty do budowania modelu przewidującego kredytobiorców zalegających ze spłatą w oparciu o historię kredytową klientów banku.



Rysunek 5. Przypisywanie predyktorów i zmiennych przewidywanych

Zmienne użyte do utworzenia predykcji są oznaczone jako predyktory, a zmienna, którą model będzie przewidywał — **domyślnie** w tym przypadku — została ustawiona jako zmienna przewidywana zgodnie z ustawieniem w węźle typu.

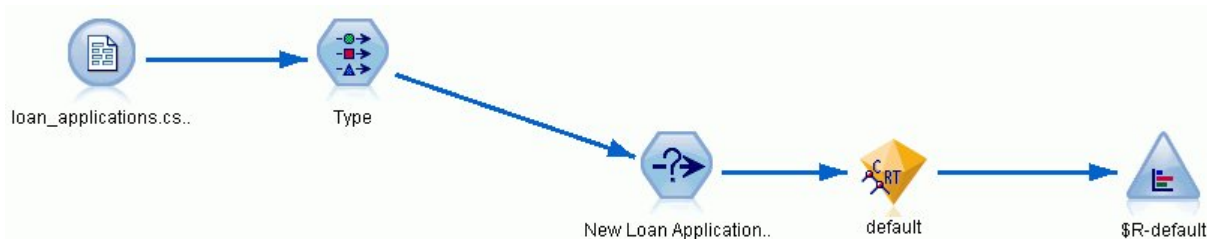
Uruchomienie tego strumienia spowoduje utworzenie modelu użytkowego zawierającego model utworzony ze zmiennych predykcyjnych.



Rysunek 6. Strumień z dodanym modelem użytkowym

Teraz analityk bankowy może użyć modelu do rozpoczęcia przewidywania, czy klienci z nieterminową spłatą zobowiązań mogą mieć zaległości płatnicze i w tym przypadku. Korzystając z pierwotnego zbioru danych, analityk

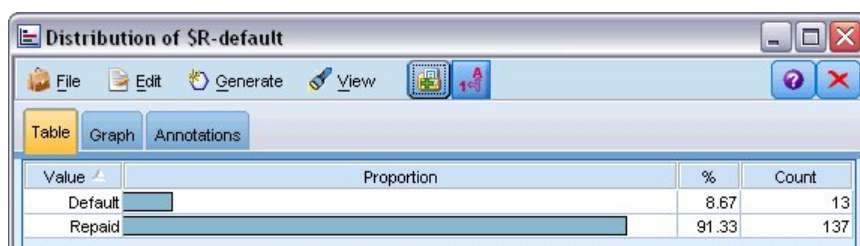
dodaje węzeł selekcji, który teraz zawiera 150 wcześniej odrzuconych rekordów kredytowych oznaczonych jako oczekujące. Analityk przekazuje te rekordy bezpośrednio do modelu, dodając węzeł rozkładu w celu graficznego przedstawienia predykcji modelu.



Rysunek 7. Strumień umożliwiający wybór nowych wniosków kredytowych z dodanym węzłem rozkładu

Węzeł rozkładu przedstawia rozkład wartości zmiennej  $\$R$ -default w modelu. Ta zmienna jest dodawana do modelu danych przez uruchomiony węzeł Drzewo C&R. Zmienna zawiera predykcję, czy każdy nowy wnioskujący będzie terminowo spłacał kredyt lub czy będzie miał zaległości w spłacie. Zmienna ta zostanie użyta później w celu porównania efektu dodania analizy jednostek.

Po wykonaniu tej części strumienia analityk dowie się na podstawie wyników z węzła rozkładu, że 137 ze 150 nowych wnioskujących będzie spłacało swoje kredyty. Przewiduje się, że pozostałych 13 wnioskodawców nie będzie regularnie spłacać kredytu, tak więc analityk będzie prawdopodobnie rekomendować, aby bank odrzucił ich wnioski.



Rysunek 8. Wyniki z węzła rozkładu bez analizy jednostek

## Dodawanie analizy jednostek

Zobaczmy teraz, czy można poprawić sytuację poprzez dodanie do równania analizy jednostek. Wyobraź sobie, że jesteś specjalistą w zakresie analizy jednostek wezwanym do banku w celu analizy możliwych oszukańczych wpisów w rekordach klientów w danych źródłowych. Mogą wystąpić powielone rekordy spowodowane błędami we wpisach danych, ale istnieje prawdopodobieństwo, że wnioskujący o kredyt może starać się ukryć swoją tożsamość. W każdym przypadku bank musi mieć prawdziwy obraz sytuacji.

Na potrzeby tego przykładu założymy, że repozytorium jednostek zostało utworzone wcześniej. Więcej informacji można znaleźć w temacie “Tworzenie repozytorium” na stronie 12.

## Przeniesienie danych źródłowych do repozytorium

W pierwszej kolejności należy dodać węzeł Eksport EA do węzła źródłowego danych. Umożliwi to eksport danych źródłowych do repozytorium jednostek.

Przed eksportem danych należy zmapować zmienne w źródle danych do cech w repozytorium jednostek. Jest to konieczne, ponieważ różne źródła danych mogą korzystać z różnych nazw zmiennych dla tego samego typu informacji. Repozytorium jednostek zawiera standardowy zbiór typów informacji (zwanymi “cechami”) umożliwiającą unikanie powielenia.

W węzle Eksport EA skonfiguruj szczegółowe informacje dotyczące repozytorium: szczegóły połączenia, znacznik źródłowy (do identyfikowania źródła danych – w tym przypadku **TEST**), typ jednostek (zbiór aktualnie używanych cech pod nazwą **PERSON**) oraz unikalną zmienną kluczową (w sposób unikalny identyfikująca każdy rekord). W tym przypadku jako unikalnego klucza użyj zmiennej **kluczowej**.

Teraz można skonfigurować mapowania. W aktualnie używanym zbiorze cech dostępne są cechy odpowiadające zmiennym *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* i *passport*.

Na początek należy skonfigurować mapowanie dla cechy *fname*. Dwukrotnie kliknij kolumnę **Mapowane do cech** w tabeli w wierszu *fname*, przewiń w dół do wpisu **NAME.GIVEN\_NAME** i kliknij go, aby utworzyć mapowanie.

Zmapuj teraz pozostałe zmienne z odpowiadającymi cechami, tak aby cały zbiór mapowań wyglądał w następujący sposób.

Tabela 12. Zmienne zmapowane do cech repozytorium.

Zmienna	Mapowanie do cech
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>generation</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passport</i>	PASSPORT.ID_NUM

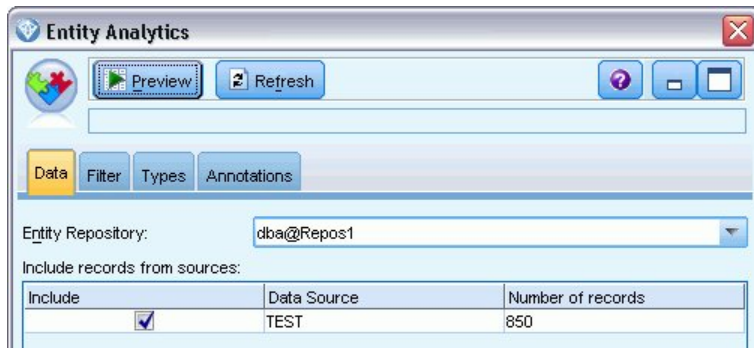
Kliknij polecenie **Wykonaj**, aby wyeksportować dane do repozytorium. Czynność ta może potrwać dłuższą chwilę. Eksport zostaje zakończony po zamknięciu okna dialogowego Informacje o stanie procesu.

## Odczytywanie rozdzielonych tożsamości

Podczas eksportu danych do repozytorium system analizy jednostek rozpoczyna rozwiązywanie konfliktów tożsamości, przypisując unikalny identyfikator jednostki, które będzie później widoczny jako zmienna *\$EA-ID*. (*Uwaga:* Ta zmienna nie jest taką samą zmienną co Unikalny klucz w węzle Eksport EA. Ta zmienna służy do identyfikowania w sposób unikalny rekordów danych źródłowych).

Pierwszym krokiem do odczytu rozdzielonych tożsamości jest dodanie węzła źródłowego Entity Analytics (EA) do strumienia. Tego węzła źródłowego nie należy na tym etapie łączyć z żadnym elementem.

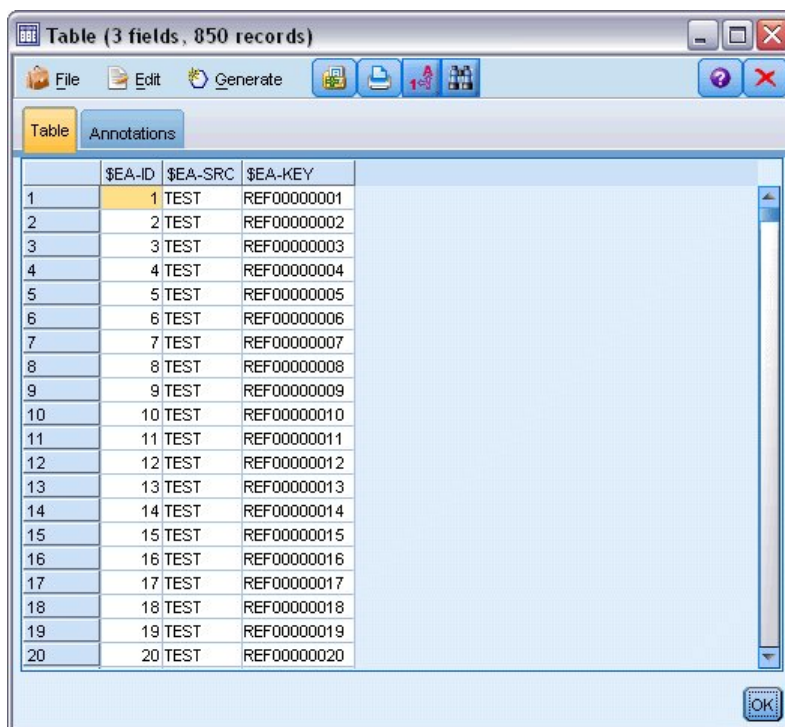
Otwórz węzeł źródłowy Entity Analytics (EA) i ustaw szczegółowe dane dotyczące repozytorium jednostek. Zostanie wyświetlona lista źródeł danych wyeksportowanych do repozytorium. W tym przypadku istnieje tylko jedno takie źródło.



Rysunek 9. Wybieranie źródła danych w repozytorium

Zaznacz pole wyboru dla źródła danych **TEST** i kliknij OK.

Przyjrzyjmy się, jaki wpływ system analizy jednostek ma na dane. Dołącz węzeł tabeli do węzła źródłowego Entity Analytics (EA), otwórz węzeł tabeli i kliknij polecenie **Wykonaj**, aby wyświetlić okno wyników węzła tabeli.



Rysunek 10. Wyniki z węzła tabeli

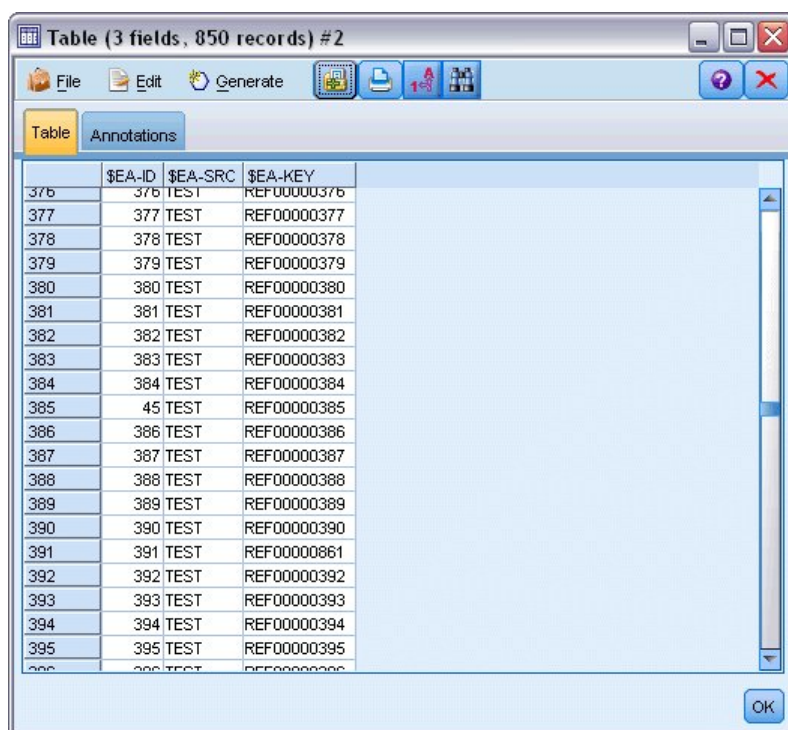
Tylko jedna zmienna wygląda znajomo – oznaczona etykietą *\$EA-KEY*. Jest to zmienna *kluczowa* z danych źródłowych i znajduje się w tym miejscu, ponieważ została wybrana jako zmienna Unikalny klucz w węźle Eksport EA.

W systemie zostały dodane również dwie inne zmienne. Zmienna *\$EA-ID* to unikalny identyfikator rozdzielonych tożsamości, a nie rekordów źródłowych. Różnica będzie widoczna za chwilę. Zmienna *\$EA-SRC* identyfikuje miejsce pochodzenia danych – w tym miejscu jest to **TEST**, ponieważ jest to znacznik źródłowy przypisany w węźle Eksport EA.

Co stało się ze wszystkimi innymi zmiennymi w danych źródłowych? Wciąż znajdują się w repozytorium. Jednak ze względów wydajnościowych węzeł źródłowy Entity Analytics (EA) przekazuje tylko minimalny zbiór zmiennych do dalszego przetwarzania.



Teraz przewiń wyniki węzła tabeli w dół do wiersza 385.

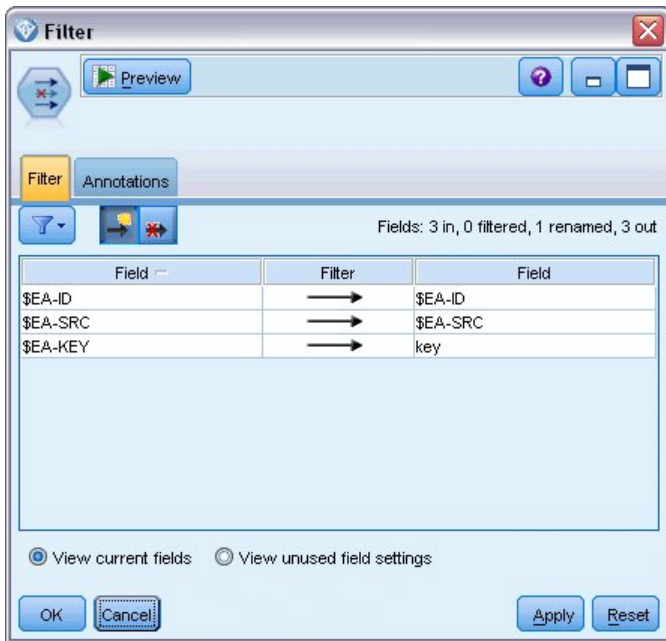


	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395
396	396	TEST	REF00000396

Rysunek 11. Różnice pomiędzy wierszami wyników węzła tabeli a liczbami \$EA-ID

Zwróć uwagę, że liczby *\$EA-ID* wydają się występować tu w przypadkowej kolejności. System analizy jednostek określił, że rekord REF00000385 dotyczy osoby zidentyfikowanej jako jednostka 45, która ma rekord REF00000045. Przewijając wyniki w dół, można zauważyć, że występuje tu więcej liczb z przypadkową kolejnością, np. w wierszach 485, 517, 520 itd. Przyjrzyjmy się temu bliżej.

Na początek podkreślmy fakt, że zmienna *\$EA-KEY* zawiera dane ze zmiennej *kluczowej* w danych źródłowych, zmieniając jej nazwę na *key*. Dołącz węzeł filtrowania do węzła źródłowego Entity Analytics (EA) i otwórz węzeł filtrowania. Dwukrotnie kliknij łańcuch *\$EA-KEY* w drugiej kolumnie **Zmienne** i wpisz *key*.



Rysunek 12. Zmiana nazwy zmiennej \$EA-KEY

Kliknij przycisk **OK**, aby zamknąć węzeł filtrowania.

Teraz należy rosnąco posortować identyfikatory jednostek *\$EA-ID*. Dołącz węzeł sortowania do węzła filtrowania. Otwórz węzeł sortowania, kliknij górny przycisk znajdujący się obok tabeli **Sortuj według**, wybierz **\$EA-ID** i kliknij **OK**.



Rysunek 13. Sortowanie identyfikatorów jednostek w porządku rosnącym

Pozostaw wartość **Rosnąco** jako kolejność sortowania i kliknij przycisk **OK**.

Teraz należy utworzyć dodatkową zmienną wskazującą, czy rekord jest unikalny czy zduplikowany. Dołącz węzeł wyliczeń do węzła sortowania. Otwórz węzeł wyliczeń i ustawia nazwę **Zmienna wyliczana** jako **IsDuplicate**. Z listy **Wylicz jako** wybierz wartość **Flaga**. Spowoduje to również wybranie dla pozycji **Typ zmiennej** wartości **Flaga**. Ustaw dla zmiennej **Wartość prawdziwa** wartość **Duplicate** i dla zmiennej **Wartość fałszywa** wartość **Unique**.



Aby znaleźć zduplikowane rekordy, należy użyć specjalnej funkcji sekwencji o nazwie @OFFSET dostępnej w programie SPSS Modeler.

Wpisz następujące wartości w polu **warunku**:

'\$EA-ID' = @OFFSET('\$EA-ID',1) or '\$EA-ID' = @OFFSET('\$EA-ID',-1))



Rysunek 14. Ustawianie warunku w węźle wyliczeń

Po posortowaniu identyfikatorów jednostek w kolejności rosnącej funkcja @OFFSET sprawdza, czy sąsiadujące identyfikatory jednostek są takie same. Jeśli tak, oznacza to, że są to duplikaty. W takim przypadku dla ich wartości *IsDuplicate* zostaje ustawiona wartość *Duplicate*, w innej sytuacji wybrana zostaje wartość *Unique*.

Kliknij przycisk **OK**, aby zamknąć węzeł.

Aby zobaczyć efekt działania węzła wyliczeń, dołącz węzeł tabeli do węzła wyliczeń, otwórz węzeł tabeli i kliknij polecenie **Wykonaj**. Przewiń wyniki w oknie wyników węzła tabeli do wiersza 45.

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Rysunek 15. Wyniki z węzeł wyliczeń

Czy przypominasz sobie wyniki pochodzące bezpośrednio z węzła źródłowego Entity Analytics (EA)? System już zidentyfikował, że rekord REF00000385 odnosi się do tej samej osoby co jednostka 45. Teraz wykonany został kolejny krok i oznaczony został fakt, że rekord REF00000045 i REF00000385 są duplikatami, ponieważ oba odnoszą się do jednostki 45.

Przewin wyniki w oknie dalej w dół. Wyświetlone zostaną inne rekordy oznaczone jako duplikaty.

Aby uzyskać raport zawierający zduplikowane rekordy, dołącz węzeł raportu (z karty węzłów wynikowych na palecie węzłów) do węzła wyliczeń *IsDuplicate*. Otwórz węzeł raportu, skopiuj następujący tekst do zmiennej wejściowej na karcie Szablon i kliknij polecenie **Wykonaj**.

```
<html>
<h1>List of duplicate customer records.

<h2>This report was generated: [@TODAY]

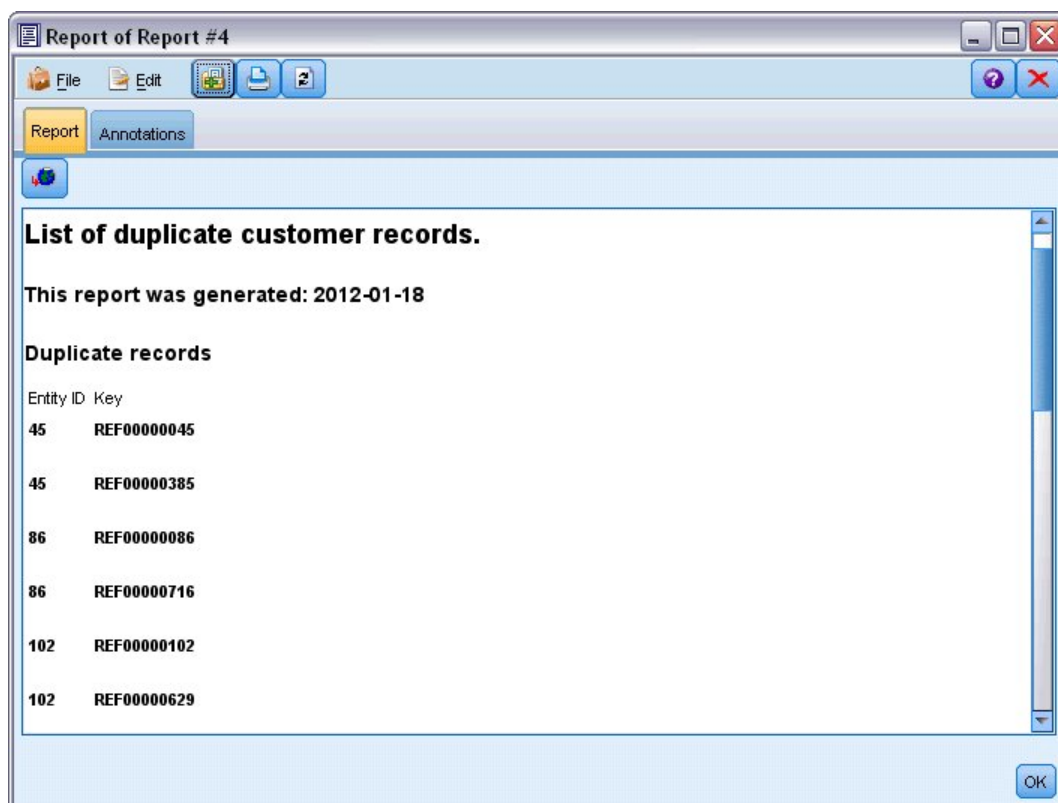
<h2>Duplicate records
<table>
  <tr>
    <td>Entity ID</td>
    <td>Key</td>
  </tr>

#WHERE IsDuplicate = "Duplicate"

  <tr>
    <td>Key</td>
    <td>Key</td>
  </tr>
```

```
#  
</table>  
</html>
```

Uzyskane zostaną następujące wyniki.



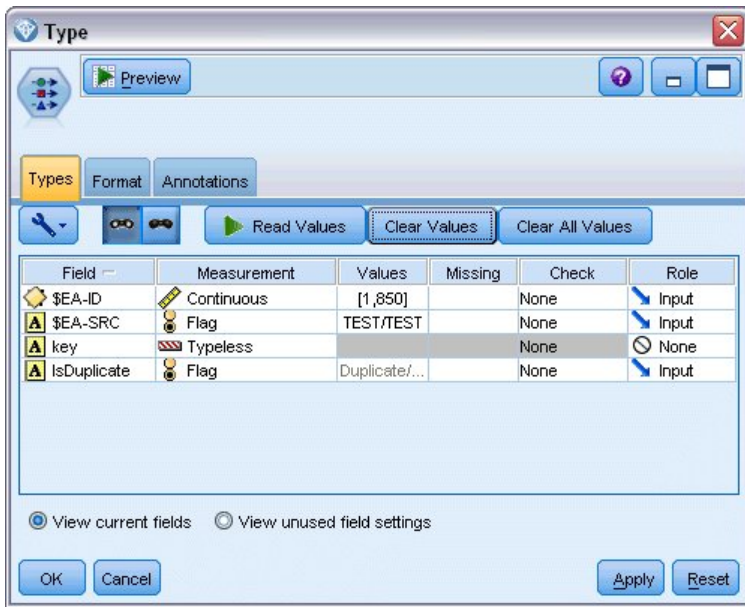
Rysunek 16. Wyniki z węzła raportu

W tym przypadku raport korzysta z formatu HTML, chociaż możliwe jest również użycie formatu XML lub ASCII.

## Porównywanie wyników analizy jednostek z pierwotnym modelem

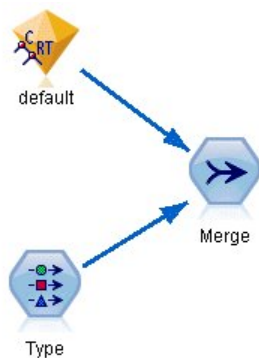
Ostatnim etapem w tym przykładzie jest sprawdzenie, czy dodanie analizy jednostek ma wpływ na pierwotne predykcje banku. W pierwotnym modelu przewidywanych było 13 potencjalnych dłużników na 150 oczekujących wniosków kredytowych. Wyniki z tego modelu można połączyć z powielonymi rekordami z analizy jednostek za pomocą węzła łączenia. W ten sposób będzie można zobaczyć ewentualny wpływ na pierwotne predykcje.

Na początek należy sprawdzić, czy nowe zmienne dodane przez analizę jednostek mają poprawne typy danych lub *poziomy pomiarów* znane w programie SPSS Modeler. Dołącz węzeł Typy do węzła wyliczeń **IsDuplicate**, otwórz węzeł Typy i kliknij przycisk **Odczytaj wartości**.



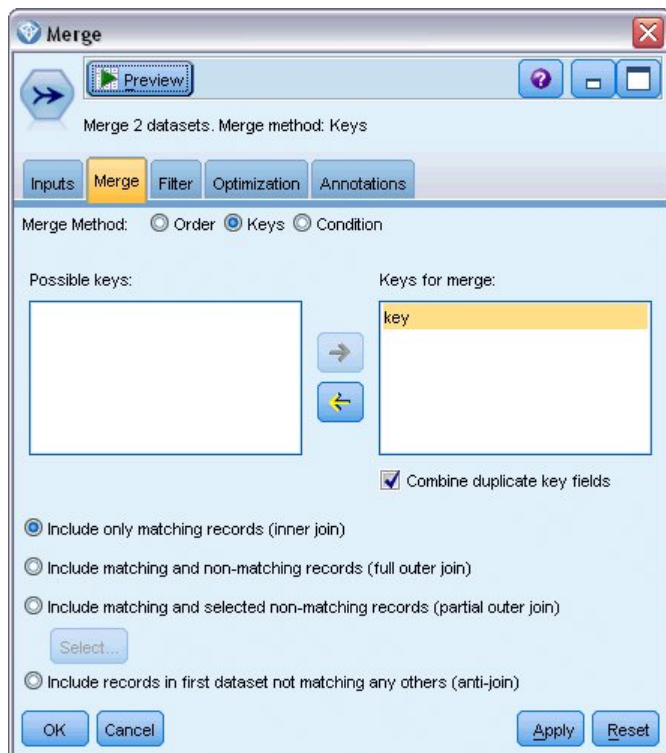
Rysunek 17. Ustawienia węzła Typy

Teraz można dodać węzeł łączenia. Dołącz go do węzła Typy i połącz z modelem użytkowym zawierającym pierwotny model. W tym celu dwukrotnie kliknij model użytkowy, wybierz opcję **Połącz**, a następnie kliknij węzeł łączenia, który teraz powinien zawierać dwie strzałki wejścia.



Rysunek 18. Dane wejściowe do węzła łączenia

Otwórz węzeł łączenia, dla pozycji **Metoda łączenia** wybierz wartość **Klucze** i kliknij przycisk strzałki w prawo, aby przenieść zmienną **kluczową** z pola **Dostępne klucze** do **Użyte klucze**, a następnie kliknij **OK**.



Rysunek 19. Określanie zmiennej kluczowej do operacji łączenia

Zostały wykonane prawie wszystkie czynności umożliwiające porównywanie. Jeśli jednak na tym etapie zostałyby dołączony i uruchomiony węzeł rozkładu, wówczas nie byłoby żadnej różnicy w stosunku do pierwotnej predykcji. Chociaż strumień łączy teraz wyniki z pierwotnego modelu użytkowego z nowymi zmiennymi utworzonymi przez analizę jednostek, sama zmienna predykcyjna (*\$R-default*) w modelu danych nie została zaktualizowana w oparciu o nowe informacje.

W tym celu należy użyć węzła wypełniania, który może zastąpić wartości zmiennej. Dołącz węzeł wypełniania do węzła łączenia i otwórz węzeł wypełniania.

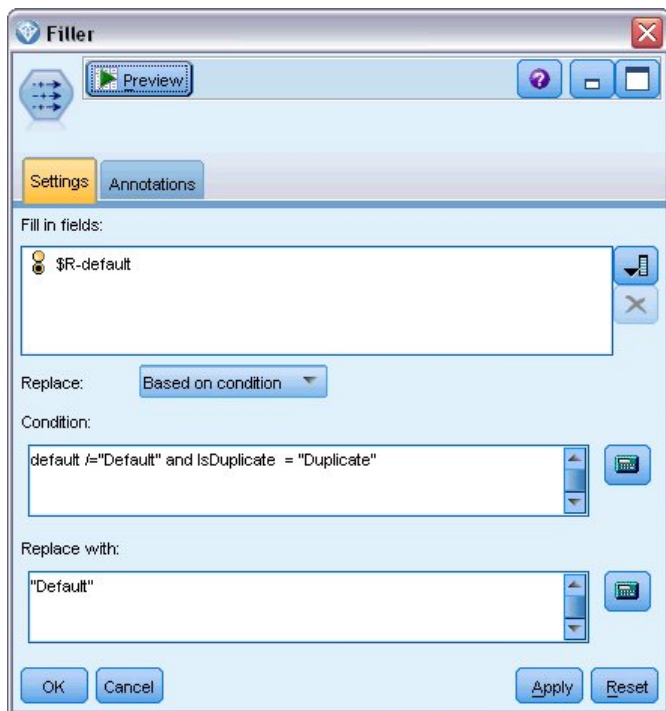
Kliknij górny przycisk po prawej stronie polecenia **Wypełnij zmienne**, przewiń w dół listy, wybierz wartość **\$R-default** i kliknij **OK**. To jest pole, którego wartości zostaną zmienione, jeśli warunek określony w dalszej części okna dialogowego zostanie spełniony.

Aby określić warunek, sprawdź, czy dla polecenia **Zamień** wybrano wartość **W oparciu o warunek**, a następnie w polu **Warunek** wprowadź:

```
default != "default" and IsDuplicate = "Duplicate"
```

W polu **Zamień na** wprowadź:

```
"default"
```



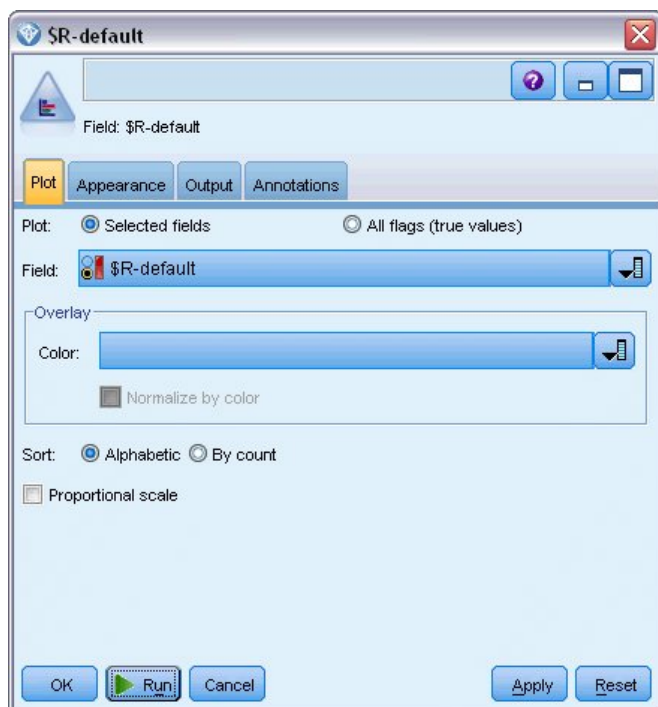
Rysunek 20. Określanie warunku zamiany wartości zmiennych

Te ustawienia wymagają wyjaśnienia. Warunek stwierdza, że dla każdego rekordu, gdzie wartość zmiennej *default* z pierwotnego zbioru danych nie jest równa wartości **Default** i rekord został oznaczony jako duplikat, wartość zmiennej *\$R-default* w modelu zostaje ustawiona jako wartość **Default**.

Zmienna *\$R-default* to zmienna z modelu zawierającego predykcję określającą, czy istnieje prawdopodobieństwo, że klient nie będzie spłacał kredytu. W ten sposób klienci z powielonymi rekordami są dodawani do modelu jako osoby potencjalnie niespłacające kredytu.

Kliknij **OK**, aby zamknąć węzeł wypełniania.

Teraz można wyświetlić różnice, jakie wywołała analiza jednostek. Na palecie Wykresy dołącz węzeł rozkładu do węzła wypełniania i otwórz węzeł rozkładu. Kliknij listę **Zmienna** i wybierz **\$R-default**.



Rysunek 21. Ustawienia węzła rozkładu

Kliknij polecenie **Wykonaj**, aby utworzyć wykres z nową predykcją.

Value	Proportion	%	Count
Default	10.67		16
Repaid	89.33		134

Rysunek 22. Wyniki z węzła rozkładu po analizie jednostek

Widocznych jest teraz 16 zamiast 13 ryzykownych wniosków kredytowych. Jeśli się potwierdzą, dodatkowe aplikacje mogą mieć bardzo duże konsekwencje. Można więc w sposób graficzny przedstawić bankowi korzyść dodania analizy jednostek do operacji oceny ryzyka.

## Podsumowanie

W tym przykładzie pokazano, jak poprzez zastosowanie analizy jednostek można wyeliminować powielane rekordy w danych dotyczących osób lub organizacji, podnosząc w ten sposób jakość predykcji.

*Uwaga:* W idealnym przypadku powielone rekordy eliminuje się przed przystąpieniem do dalszych czynności przetwarzania. Można również skorzystać z węzła automatycznego przygotowania danych do analizy danych i zidentyfikowania problemów do skorygowania, wyeliminowania problematycznych lub nieefektywnych zmiennych, ewentualnego wyliczenia nowych atrybutów i zwiększenia wydajności poprzez inteligentne techniki monitorowania.

Połączenie analizy jednostek i automatycznego przygotowania danych może pomóc w uzyskaniu możliwie jak najczystszych danych.





---

## Dodatek. Właściwości skryptów dla programu IBM SPSS Modeler Entity Analytics

---

### Skrypty w IBM SPSS Modeler Entity Analytics

Skrypty w SPSS Entity Analytics są potężnym narzędziem automatyzacji procesów interfejsu użytkownika. Skrypty mogą wykonywać takie same działania, jakie wykonujesz, używając myszy lub klawiatury, możesz je wykorzystać do zautomatyzowania wysoce powtarzalnych zadań lub takich zadań, których wykonanie ręczne wymaga dużo czasu. Wyjaśnienie, w jaki sposób używać skryptów, można znaleźć w podręczniku *ModelerScriptingAutomation.pdf* dostępnym razem z produktem SPSS Modeler.

---

### Wspólne właściwości

Właściwości wspólne dla węzłów IBM SPSS Modeler Entity Analytics znajdują się w poniższej tabeli. Informacje dotyczące określonych węzłów znajdują się w poniższych sekcjach.

Tabela 13. Wspólne właściwości

Nazwa właściwości	Typ danych	Opis właściwości
entity_repository	[ <i>'field'</i> , <i>'field'</i> , ... , <i>'field'</i> ]	Połączenie z repozytorium. Format: [ <i>'reposname'</i> , <i>'username'</i> , <i>'password'</i> ]  Przykład: entity_repository = [ 'repos1', 'dba', 'psw1' ]
entity_type	<i>string</i>	Typ jednostek (zestaw cech), który zostanie użyty.  Przykład: entity_type = 'PERSON'

---

### Właściwości entityanalytics\_exportnode



Węzeł Eksport EA to węzeł końcowy odczytujący ze źródła danych dane dotyczące jednostek i eksportujący dane do repozytorium w celu rozdzielenia jednostek.

Tabela 14. Właściwości entityanalytics\_exportnode

Właściwości entityanalytics_exportnode	Typ danych	Opis właściwości
mode	Add PurgeFirst	Tryb eksportu Add dodaje rekordy pliku źródłowego do istniejącej zawartości repozytorium; PurgeFirst usuwa istniejącą zawartość przed eksportem.
source_tag	<i>string</i>	Identyfikator źródła danych  Przykład: source_tag = 'CUST'
unique_key_field	<i>string</i>	Zmienna wejściowa używana dla unikalnych identyfikatorów dla rekordów danych.  Przykład: unique_key_field = 'ID'

Tabela 14. Właściwości *entityanalytics\_exportnode* (kontynuacja)

Właściwości <i>entityanalytics_exportnode</i>	Typ danych	Opis właściwości
field_mapping	[ [ 'field_name' 'feature.element' 'usage_type' ]... ]	Mapuje zmienne wejściowe do odpowiednich cech w repozytorium.  Przykład: field_mapping = [[ 'fname' 'NAME.GIVEN_NAME' " ] [ 'addr1' 'ADDRESS.ADDR1' 'PRIMARY' ] ] Uwaga: Aby ustawić dla <i>usage_type</i> wartość odpowiadającą wartości "(Auto)", należy użyć znaku ", tak jak w pierwszym przykładzie powyżej.

## Właściwości *entityanalytics\_sourcenode*



Węzeł źródłowy Entity Analytics (EA) odczytuje z repozytorium rozdzielone jednostki i przekazuje te informacje do strumienia w celu dalszego przetworzenia, np. formatowania w raporcie.

Tabela 15. Właściwości *entityanalytics\_sourcenode*

Właściwości <i>entityanalytics_sourcenode</i>	Typ danych	Opis właściwości
source_tags	list	Lista znaczników źródeł danych, które mogą być pobrane z repozytorium.  Przykład: source_tags=[ 'LOANS', 'CUSTOMERS' ]
relationships	None Close All	Kryterium dopasowania pobierające z repozytorium szczegóły dotyczące relacji. <b>None</b> nie zwraca żadnych relacji. <b>Close</b> zwraca bliskie dopasowania w zależności od szczegółów, takich jak stopień separacji. <b>All</b> zwraca wszystkie możliwe relacje.
max_degree_separation	integer	Minimum 0, maksimum 3.
output_entity_type	string	Lista typów jednostek używanych w repozytorium.

## Właściwości *entityanalytics\_processnode*



Węzeł przesyłania strumieniowego EA porównuje nowe obserwacje w oparciu o dane jednostek w repozytorium.

Tabela 16. Właściwości *entityanalytics\_processnode*

Właściwości <i>entityanalytics_processnode</i>	Typ danych	Opis właściwości
match	Exact ByIdentifier All	Kryterium dopasowania pobierające jednostki z repozytorium. <b>Exact</b> zwraca tylko dokładne dopasowania. <b>ByIdentifier</b> zwraca dokładne dopasowania oraz jednostki z takimi samymi identyfikatorami. <b>All</b> zwraca wszystkie możliwe dopasowania.

Tabela 16. Właściwości *entityanalytics\_processnode* (kontynuacja)

Właściwości <b>entityanalytics_processnode</b>	Typ danych	Opis właściwości
save_search_records	<i>boolean</i>	
relationships	None Close All	Kryterium dopasowania pobierające z repozytorium szczegóły dotyczące relacji. <b>None</b> nie zwraca żadnych relacji. <b>Close</b> zwraca bliskie dopasowania w zależności od szczegółów, takich jak stopień separacji. <b>All</b> zwraca wszystkie możliwe relacje.
max_degree_separation	<i>integer</i>	Minimum 0, maksimum 3.
output_entity_type	<i>string</i>	Lista typów jednostek używanych w repozytorium.



---

## Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. Materiał ten jest również dostępny w IBM w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem, że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na towary i usługi, o których mowa w niniejszej publikacji. Przedstawienie niniejszej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przysyłać na adres:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
U.S.A.*

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan, Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokio 103-8510, Japonia*

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (RĘKOJMIĘ RÓWNIEŻ WYŁĄCZA SIĘ), WYRAŹNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA TA NIE NARUSZA PRAW OSÓB TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w niniejszej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych innych podmiotów zostały wprowadzone wyłącznie dla wygody użytkownika i w żadnym wypadku nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przysłanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjobiorcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) wdrożenia wymiany informacji między niezależnie utworzonymi programami i innymi programami (łącznie z tym opisywanym) oraz (ii) wspólnego wykorzystywania wymienianych informacji, powinni skontaktować się z:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
U.S.A.*

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z tymi produktami. Pytania dotyczące możliwości produktów innych podmiotów należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennej pracy. W celu kompleksowego ich zilustrowania, podane przykłady zawierają nazwiska osób prywatnych, nazwy przedsiębiorstw oraz nazwy produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

---

## Znaki towarowe

IBM, logo IBM i [ibm.com](http://ibm.com) są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp. zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW IBM, w sekcji "Copyright and trademark information" (Informacje o prawach autorskich i znakach towarowych), pod adresem [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych Intel Corporation w Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym The Open Group w Stanach Zjednoczonych i/lub w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące języka Java są znakami towarowymi lub zastrzeżonymi znakami towarowymi Oracle i/lub przedsiębiorstw afiliowanych.

---

## **Warunki dotyczące dokumentacji produktu**

Zezwolenie na korzystanie z tych publikacji jest przyznawane na poniższych warunkach.

### **Zakres stosowania**

Niniejsze warunki stanowią uzupełnienie warunków używania serwisu WWW IBM.

### **Użytek osobisty**

Użytkownik ma prawo kopiować te publikacje do własnego, niekomercyjnego użytku pod warunkiem zachowania wszelkich uwag dotyczących praw własności. Użytkownik nie ma prawa dystrybuować ani wyświetlać tych publikacji czy ich części, ani też wykonywać na ich podstawie prac pochodnych bez wyraźnej zgody IBM.

### **Użytek służbowy**

Użytkownik ma prawo kopiować te publikacje, dystrybuować je i wyświetlać wyłącznie w ramach przedsiębiorstwa Użytkownika pod warunkiem zachowania wszelkich uwag dotyczących praw własności. Użytkownik nie ma prawa wykonywać na podstawie tych publikacji ani ich fragmentów prac pochodnych, kopiować ich, dystrybuować ani wyświetlać poza przedsiębiorstwem Użytkownika bez wyraźnej zgody IBM.

### **Prawa**

Z wyjątkiem zezwoleń wyraźnie udzielonych w niniejszym dokumencie, nie udziela się jakichkolwiek innych zezwoleń, licencji ani praw, wyraźnych czy domniemanych, odnoszących się do tych publikacji czy jakichkolwiek informacji, danych, oprogramowania lub innej własności intelektualnej, o których mowa w niniejszym dokumencie.

IBM zastrzega sobie prawo do anulowania zezwolenia przyznanego w niniejszym dokumencie w każdej sytuacji, gdy, według uznania IBM, korzystanie z tych publikacji jest szkodliwe dla IBM lub jeśli IBM uzna, że warunki niniejszego dokumentu nie są przestrzegane.

Użytkownik ma prawo pobierać, eksportować lub reeksportować niniejsze informacje pod warunkiem zachowania bezwzględnej i pełnej zgodności z obowiązującym prawem i przepisami, w tym ze wszelkimi prawami i przepisami eksportowymi Stanów Zjednoczonych.

IBM NIE UDZIELA JAKICHKOLWIEK GWARANCJI, W TYM TAKŻE RĘKOJMI, DOTYCZĄCYCH TREŚCI TYCH PUBLIKACJI. PUBLIKACJE TE SĄ DOSTARCZANE W STANIE, W JAKIM SIĘ ZNAJDUJĄ ("AS-IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (RĘKOJMIĘ RÓWNIEŻ WYŁACZA SIĘ), WYRAŹNYCH CZY DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU CZY NIENARUSZANIA PRAW OSÓB TRZECICH.





# Indeks

## A

- analiza jednostek
  - korzystanie z IBM SPSS Modeler 5
  - korzystanie z programu z innymi produktami IBM SPSS 28
  - porównanie z analizą predykcyjną 2
  - zdefiniowane 1
- anonimizacja cech
  - repozytorium jednostek 19

## C

- cechy
  - repozytorium jednostek 7, 14, 15, 17, 18, 26, 27
- czyszczenie
  - repozytorium jednostek 32

## D

- dane uwierzytelniające administratora
  - zarządzanie analizą jednostek 30
- dane źródłowe, wybieranie do analizy jednostek 23
- dopasowanie jednostek, ustawianie progu 22

## E

- eksportowanie
  - dane do repozytorium jednostek 7

## K

- konfiguracja
  - repozytorium jednostek 16, 22, 23

## M

- mapowanie zmiennych
  - do cech repozytorium jednostek 7, 14, 15, 17, 26, 27

## N

- nowe obserwacje, porównywanie w oparciu o repozytorium analizy jednostek 25

## P

- progu dopasowania jednostek, analiza jednostek 22
- przypisania portów
  - konfiguracja analizy jednostek 29

## R

- reguły rozdzielania, analiza jednostek 22

- repozytorium
    - analiza jednostek 6, 7, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 25, 26, 27, 32, 33
    - analiza jednostek administracyjnych 29
    - catalog składowania dla analizy jednostek, zmiana 30
  - repozytorium jednostek 11
  - anonimizacja 19
  - cechy 18
  - czyszczenie 32
  - konfiguracja przypisań portów 29
  - konfigurowanie 16, 22, 23
  - łączenie z IBM SPSS Modeler 7
  - opcje 13
  - opracowywanie 17
  - porównywanie nowych obserwacji w oparciu o 25
  - przeniesienie do innego katalogu składowania 30
  - tworzenie 6, 12, 13
  - ustawianie 11
  - ustawianie właściwości strumienia 31
  - usuwanie 32, 33
  - usuwanie nieużywanych danych 32
  - zadania administracyjne 29
  - zarządzanie danymi uwierzytelniającymi administratora 30
- rozdzielanie tożsamości, analiza jednostek 7
  - rozdzielone tożsamości, analizowanie za pomocą analizy jednostek 23

## T

- tagi źródłowe
  - repozytorium jednostek 15
- tworzenie skryptów
  - właściwości 51
- typ informacji, ustawianie dla analizy jednostek 24
- typy jednostek
  - analiza jednostek 20
  - repozytorium jednostek 15
- typy użycia, analiza jednostek 20

## U

- unikalne klucze
  - analiza jednostek 7
  - repozytorium jednostek 15
- usuwanie
  - repozytorium jednostek 32, 33
- usuwanie nieużywanych danych
  - repozytorium jednostek 32

## W

- węzeł Eksport EA, analiza jednostek 11
- węzeł przesyłania strumieniowego EA, analiza jednostek 25

- węzeł źródłowy Entity Analytics (EA) 23
- węzły
  - dodawanie do strumienia analizy jednostek 24
- węzły eksportu
  - analiza jednostek 7, 11
- węzły przetwarzania
  - analiza jednostek 9, 25
- węzły źródłowe
  - analiza jednostek 9, 23
- właściwości
  - tworzenie skryptów 51
- właściwości strumienia
  - ustawienia dla analizy jednostek 31
- wyniki
  - z analizy jednostek 28

## Z

- zmiana nazwy
  - zmiennicze zawierające dane 24

## Ź

- źródła danych
  - łączenie z analizą jednostek 6, 12
  - wyświetlanie dla analizy jednostek 13, 27







Drukowane w USA