

Wysoka jakość opisu statystycznego może być osiągnięta przez eksploracyjną analizę danych, graficzne metody eksploracji danych oraz wyszukiwania wzorców i trendów. W przypadku estymacji analiza statystyczna dostarcza kilku narzędzi. Zawierają one szacowanie wartości punktu i przedziału ufności, prostą regresję liniową i korelację oraz regresję wielokrotną. Z kolei wszystkie metody i techniki wykorzystywane do estymacji mogą być użyte do predykcji. Obejmują one tradycyjne metody szacowania wartości punktu i przedziału ufności, prostą regresję liniową i korelację oraz regresję wielokrotną oraz metody eksploracji danych i odkrywania wiedzy, np. sieci neuronowe, drzewa decyzyjne i metoda k-najbliższych sąsiadów. Klasyfikacje zapewnia algorytm k-najbliższych sąsiadów, drzewa decyzyjne lub sieci neuronowe. Grupowanie oznacza grupowanie rekordów, a nawet obserwacji lub przypadków w klasy podobnych obiektów. Grupa jest zbiorem rekordów, które są podobne do siebie nawzajem i niepodobne do rekordów z innych grup. Grupowanie różni się od klasyfikacji tym, że w przypadku grupowania nie ma zmiennej celu. Do identyfikacji naturalnych (niepredefiniowanych) segmentów możemy stosować dwustopniową analizę skupień. Jest ona stosowana wówczas, gdy występuje konieczność wyróżnienia grupy obiektów podobnych, w przypadku, gdy obiekty te są opisane przez więcej niż jedną cechę. W tym celu często stosowana jest metoda niehierarchiczna, która polega na przenoszeniu obiektów z jednego skupienia do innego, w poszukiwaniu najlepszego zestawu skupień według zadanego kryterium (np. najmniejszej wariancji międzygrupowej). Zastosowana procedura bazująca na najmniejszych odległościach k-średnich pozwoli rozpoznać i sprofilować podstawowe i homogeniczne segmenty przestrzenne, np. w obrębie jednostek terytorialnych. Przy wprowadzeniu predyktorów uzyskujemy dla każdego powiatu segmenty, których homogeniczność wewnętrzną i heterogeniczność zewnętrzną określa miara Silhouette, wskazująca na poprawność procedury grupowania. Z kolei odkrywanie reguł polega na eksploracji nieodkrytych reguł do ilościowego określenia relacji pomiędzy dwoma lub więcej atrybutami. Funkcjonalne są algorytmy tworzenia reguł asocjacyjnych, algorytm a priori i algorytm GRI. W poniższej egzemplifikacji zaprezentujemy przypadek analityczny dotyczący prognozowania popytu zawodów prawniczych na regionalnym rynku pracy w celu prześledzenia procedur pozwalających na przetworzenie danych zgodnie z modelowaniem CRISP-DM oraz zastosowanie formalnego modelu krótkosezonowego.