

Analiza współpracy naukowców metodami statystycznej eksploracji danych

praca inżynierska, autor **Mateusz Ozimek**

opiekun: dr inż. Julian Sienkiewicz

Niniejsza praca dyplomowa zawiera wyniki analiz zbioru danych dotyczących publikacji naukowców utworzonego z bazy publikacji *Web of Science* przy wykorzystaniu metod statystycznej eksploracji danych. Dane były badane w celu uzyskania zależności pozycji uczelni w rankingu od wybranych parametrów opisujących uczelnię (np. liczby publikacji z określonej dziedziny nauki). Ranking uczelni użyty w wykorzystywanej bazie danych to *QS World University Rankings* z roku 2009. Baza danych wykorzystywana w niniejszej pracy umożliwia dostęp do następujących danych – rankingu uczelni, wpisów dotyczących każdej publikacji wydanej przez odpowiednią uczelnię: roku tejże publikacji, kategorii, do których została ona zaklasyfikowana oraz języka jej publikacji jak również danych o każdej publikacji takie jak nazwa uczelni oraz adres każdego ze współautorów.

Powyższe dane wykorzystane zostały do treningu i testowania klasyfikatorów. Zostały zastosowane w tym celu: liniowa i kwadratowa analiza dyskryminacyjna (LDA i QDA), metoda k-najbliższych sąsiadów (KNN), metoda wektorów wspierających (SVM), a także naiwny klasyfikator Bayesa. Do testów zostały wykorzystane metody sprawdzające przez powtarne podstawienie, krosvalidacja oraz metoda *leave-one-out*. Z otrzymanych wyników można wywnioskować, że uczelnie, które mają większą liczbę wszystkich publikacji oraz publikacji z fizyki mają większą szansę na zajęcie wyższej pozycji w rankingu. Ponadto ważnymi elementami składowymi pozycji w rankingu według przeprowadzonych badań są : liczba publikacji z najlepszą dziesiątką uczelni, średni ranking uczelni z którą prowadzona jest współpraca oraz liczba publikacji z uczelniami ze Stanów Zjednoczonych. Dodatkowo okazuje się, że uczelnie znajdujące się w czołówce wyróżniają się pod względem publikacji z fizyki spośród pozostałych w rankingu.