

## Case study 1.6 (ZED – TPD 2008)

### Analiza danych telekomunikacyjnych i problemu odejść klientów – tzw. „churn” problem

Wyjaśnienie terminu angielskiego „churn” – wskazania klientów którzy mogą zrezygnować z korzystania z usług konkretnej firmy na rzecz innej, konkurencyjnej firmy.

Churn może być działaniem formalnym – np. złożenie przez klienta wniosku o rezygnację z usług lub działaniem nieformalnym – np. zaprzestanie płatności faktur bez wypowiedzenia umowy. Celem analizy Churn jest identyfikacja klientów, którzy zagrożeni są odejściem lub rezygnacją z usług, jak również klientów, których Churn spowoduje największe straty dla firmy.

Analiza Churn pozwala określić profil klientów skłonnych do rezygnacji z usług.

Wyodrębnienie grup klientów o wysokim prawdopodobieństwie odejścia pozwala firmie na wyprzedzenie działań tych klientów i podjęcie akcji prewencyjnych (np. atrakcyjne programy „antychurnowe”).

#### Opis danych:

Dane zebrano w USA, dotyczą klientów pewnej kompanii telekomunikacyjnej:

Zbiór danych obejmuje 3334 klientów opisanych 20 zmiennymi. Ostatnia kolumna ma charakter atrybutu decyzyjnego definiującego klasyfikację klientów – zwróć uwagę na niezrównoważenie licznosci klas.

Krótką charakterystyką danych podana jest poniżej:

State: categorical, for the 50 states and the District of Columbia

Account length: integer-valued, how long account has been active

Area code: categorical

Phone number: essentially a surrogate for customer ID

International Plan: dichotomous categorical, yes or no

VoiceMail Plan: dichotomous categorical, yes or no

Number of voice mail messages: integer-valued

Total day minutes: continuous, minutes customer used service during the day

Total day calls: integer-valued

Total day charge: continuous, perhaps based on foregoing two variables

Total evening minutes: continuous, minutes customer used service during the evening

Total evening calls: integer-valued

Total evening charge: continuous, perhaps based on foregoing two variables

Total night minutes: continuous, minutes customer used service during the night

Total night calls: integer-valued

Total night charge: continuous, perhaps based on foregoing two variables

Total international minutes: continuous, minutes customer used service to make international calls

Total international calls: integer-valued

Total international charge: continuous, perhaps based on foregoing two variables

Number of calls to customer service: integer-valued

Uwagi do analizy:

Warto zwrócić uwagę na

- (a) ocena jakości danych i ew. oczyszczanie pliku z niepoprawnych elementów,
- (b) redukcja zbioru zmiennych do podzbioru najważniejszych zapewniających satysfakcjonującą ocenę klasyfikacji klientów,
- (c) poszukiwanie zależności między wartościami wybranych zmiennych a globalną oceną stanu klienta,
- (d) konstrukcja tzw. klasyfikatora, czyli wskazań dla podejmowania końcowej decyzji – porównania wielu sposobów konstrukcji takich klasyfikatorów
- (e) uwzględnienie na nierównowagę licznosci klas decyzyjnych – ważniejsza jest identyfikacja klasy klientów porzucających firmę.