

White Paper - Analiza transakcji na stacjach benzynowych Statoil dla Premium Club

Wprowadzenie

Odbiorca analiz

Program Premium Club zarządzany jest przez firmę Premium Club Sp. z o.o. W programie tym uczestniczą firmy z różnych sektorów biznesu:

- sieć stacji benzynowych Statoil,
- restauracje Pizza Hut i KFC,
- sieć salonów EMPIK i minilaboratoria EMPIK Foto,
- Bank BPH.

W programie Premium Club prowadzony jest system kart PC, na które zbierane są punkty za dokonane zakupy. Punkty te mogą być wymienione na upominki.

W 2005 roku analitycy StatConsulting wykonali na zlecenie Premium Club projekt „Analiza transakcji na stacjach benzynowych Statoil dla Premium Club.”

Problem biznesowy

Po rozmowach z przedstawicielami PremiumClub ustalono, że celem projektu jest skonstruowanie analitycznego modelu danych opisującego zachowania klientów dokonujących zakupów na stacjach benzynowych.

Na stacjach Statoil nabywane są nie tylko produkty paliwowe, ale również czasopisma, napoje i produkty żywnościowe. Kierowca punkty należne za zakupy zbiera na karcie PC. Dzięki temu możliwe jest ustalenie, jakie produkty zostały nabyte. W wyniku analiz miała być ustalona struktura zakupów dokonywanych na stacjach benzynowych Statoil.

Drugim zadaniem postawionym przed firmą StatConsulting było zdobycie informacji na temat preferencji klientów odnośnie typów odwiedzanych stacji.

Rozwiązanie analityczne

Konsultanci firmy StatConsulting mieli do dyspozycji dane o zakupach dokonanych na stacjach Statoil oraz z powiązanych z nimi kart PC, pochodzące z okresu pół roku. Po rozpatrzeniu problemu analitycy StatConsultingu postanowili w oparciu o powyższe dane skonstruować trzy modele Data Mining. Każdy model dotyczył odrębnego aspektu analizowanego problemu biznesowego:

- Struktura zakupów na stacjach Statoil - model asocjacyjny
- Preferencje klientów odnośnie typów stacji benzynowych
 - klasyfikacja preferencji - ekspercki model klasyfikacyjny
 - grupowanie klientów według ich preferencji - automatyczny model segmentacyjny

Do analizy danych oraz budowy modeli Data Mining został wykorzystany System Górnik, oprogramowanie do Data Mining, stworzone przez Statconsulting.

Wstępna obróbka danych

Przygotowanie danych to jeden z najistotniejszych elementów procesu Data Mining. Od niego bowiem w bardzo dużym stopniu zależy jakość tworzonych później modeli, oraz wiarygodność wyprowadzanych wniosków. Dobry, wiarygodny model powstaje w oparciu o wartościowe, prawidłowo przygotowane i oczyszczone dane. Zwykle etap ten zajmuje od 60 do 80% czasu poświęcanego na proces Data Mining, co miało miejsce również w wypadku opisywanego tu projektu.

Dane, dostarczone przez firmę Premium Club zawierały opis transakcji dokonanych w ciągu pół roku przy użyciu kart PC przez klientów z całej Polski.

Na początku dokonano wstępnej analizy opisowej rozkładów i wykresów poszczególnych zmiennych. Na jej podstawie ustalono, jakim procedurom mają dalej podlegać dane.

Dane zostały oczyszczone i przygotowane do analizy Data Mining. Zostały z nich usunięte rekordy z brakami danych i błędy. Usunięto z nich również dane nie spełniające wymogów modelu i uzupełniono niektóre z brakujących danych. Następnie dane zostały zagregowane w taki sposób, by jeden rekord opisywał jedną kartę.

Modele

Analitycy firmy StatConsulting skonstruowali trzy modele analizy danych, mające umożliwić znalezienie odpowiedzi na następujące pytania:

- Jaka jest struktura zakupów dokonywanych na stacjach Statoil?
- Jakie są preferencje klientów odnośnie typów odwiedzanych stacji benzynowych?
- Na jakie grupy dzielą się klienci stacji Statoil, jeśli weźmiemy pod uwagę dostarczone dane, opisujące ich zachowanie?

Model asocjacyjny - Struktura zakupów

Opracowany model tworzył reguły i zależności dotyczące zakupów dokonywanych na stacjach Statoil przy użyciu modułu reguł asocjacyjnych w Systemie Górnik. Polegają one na wykryciu związków przyczynowo-skutkowych między zdarzeniami. Otrzymanym wynikiem jest wiedza w postaci prostych reguł, które mogą później być zweryfikowane przez eksperta.

Stworzony model analizował strukturę i wewnętrzne zależności zakupów dokonywanych na stacjach benzynowych Statoil.

Model klasyfikacyjny - Preferencje typów stacji

Dokonana została analiza częstości odwiedzania różnych typów stacji Statoil. Klienci zostali najpierw podzieleni na grupy przez analityków StatConsulting w oparciu między innymi o analizę histogramów, wykresów i tabel danych. Głównymi czynnikami branymi pod uwagę przy przypisywaniu klientów do grup były:

- częstość odwiedzin na poszczególnych typach stacji,
- ilość wydawanych tam pieniędzy.

Następnie utworzony podział został skonsultowany z ekspertem Premium Club i zmodyfikowany w oparciu o jego wnioski. Ten model dzielił klientów stacji Statoil w oparciu o ich preferencje odnośnie typów stacji benzynowych. W wyniku analiz wyróżniono trzy odrębne grupy klientów.

Model segmentacyjny - Grupy klientów

W celu sprawdzenia poprawności dokonanego podziału analitycy StatConsulting skonstruowali kolejny model. Był to model Data Mining automatycznie grupujący klientów w oparciu o moduł samoorganizujących się map (SOM) Systemu Górnik. Model ten dzielił klientów, bazując na dostarczonych danych. Również w nim zostały utworzone trzy grupy, niemal identyczne z uzyskanymi w pierwszym modelu (średnio ponad 79% zgodności pomiędzy odpowiadającymi sobie grupami), co potwierdza trafność i rzetelność dokonanych analiz.

Wyniki

W wyniku przeprowadzonych analiz zostały ustalone reguły, jakimi kierują się klienci dokonujący zakupów na stacjach benzynowych, a zwłaszcza reguły dotyczące nabywania produktów paliwowych.

Analiza Data Mining dokonana przy użyciu Systemu Górnik wykryła dziesiątki zależności pomiędzy produktami kupowanymi przez klientów Premium Club na stacjach benzynowych różnych typów. Zależności te dotyczyły zarówno prawidłowości w kupowaniu produktów paliwowych, jak też i innych produktów - sezonowych, napojów, słodyczy, prasy itp.

Klienci, w zależności od swoich preferencji dotyczących poszczególnych typów stacji benzynowych oraz częstości i intensywności korzystania z nich, zostali podzieleni na trzy odrębne grupy. Wiarygodność tego podziału została potwierdzona dwoma modelami analitycznymi (modelem klasyfikacyjnym oraz segmentacyjnym).

Uzyskane wyniki umożliwiły Premium Club utworzenie odrębnych strategii marketingowych skierowanych do poszczególnych grup konsumenckich, w oparciu o ich preferencje oraz strukturę dokonywanych zakupów.

Wyniki analiz Data Mining przedstawione przez firmę StatConsulting zostały przyjęte z uznaniem i są obecnie wykorzystywane przez Premium Club do tworzenia strategii marketingowych. Z tego powodu nie jest jeszcze dostępna całkowita ocena skuteczności skonstruowanych modeli na danych bieżących.

Firma Premium Club, usatysfakcjonowana realizacją projektu, złożyła firmie StatConsulting propozycję przedłużenia współpracy i wykonania dalszych analiz danych.