

Marcin Wardyński

Analiza ryzyka kredytowego

Rozkład danych z histogramów

- Ponad 1400 osób uzyskało dobry wskaźnik ryzyka kredytowego i 1000 osób zły
- Najliczniejszą grupą wiekową jest 25-30 lat, tylko minimalnie ustępuje jej grupa 30-35 lat, a w dalszej kolejności są na niemalże tym samym miejscu grupy 20-25 lat, oraz 35-40 lat. Dla pozostałych grup, czym starsza wiekowo grupa, tym mniej reprezentatów w zbiorze danych posiada.
- Większość osób uwzględnionych w badaniu posiada średni dochód, ok 1/3 osób ma wysoki dochód, a najmniej liczną grupą są osoby mało zarabiające
- ok 2/3 osób posiada pięć i więcej kart kredytowych, a 1/3 mniej
- w grupie badawczej jest właściwie tyle samo osób po College-u, co i po High school
- prawie 2/3 osób posiada dwie lub więcej pożyczek na samochód, a reszta jedną lub żadnej

Drzewo CHAID dla klasyfikacji

- niemalże wszyscy posiadający mniej niż pięć kart kredytowych mają dobrą ocenę zdolności kredytowej
- dla posiadaczy pięciu i więcej kart kredytowych, ci co mają wysoki dochód też mają raczej dobrą ocenę zdolności kredytowej
- dla pozostałych grup dominuje zła ocena zdolności kredytowej

Wykres ważności

Ukazuje, które zmienne mają największy wpływ na przyznaną ocenę zdolności kredytowej. Można powiedzieć, że wykres ten odpowiada naszej intuicji i wskazaniom drzewa CHAID

- największe znaczenie ma dochód osoby i to jest dość oczywiste, że ocena kredytowa w pierwszej kolejności zależy od zarobków
- następnie ilość posiadanych kart, czym więcej, tym najpewniej więcej zobowiązań na wydatki bieżące i trudniej o kredyt
- wiek, który ma pewne przełożenie na zarobki, ale i maksymalny czas spłaty nowego kredytu
- pożyczki na samochód – samochód jest często dobrem luksusowym, z którego w razie konieczności można zrezygnować i odstąpić od pożyczki
- edukacja według tego wykresu nie ma właściwie żadnego wpływu na ocenę zdolności kredytowej

Wynik po usunięciu zmiennej odpowiadającej edukacji

Usunięcie zmiennej odpowiadającej edukacji nie niesie za sobą niemalże żadnych widocznych zmian, gdyż jak stwierdziliśmy w poprzednim punkcie, jest zmienna o najmniejszej, niemalże zerowej, ważności. Toteż jej eliminacja nie wpływa na jakość klasyfikacji zdolności kredytowej, a zarazem zmniejsza wymiarowość danych wejściowych.