

Raport Końcowy

Bank Marketing Classification Task

Bartosz Jezierski i Michał Iwaniuk

21 kwietnia 2024

1 Wstęp

Celem tego projektu było wprowadzenie nas w świat ML i procesu tworzenia modeli maszynowych. Został on zrealizowany jako część kursu "Wstęp do uczenia maszynowego" na kierunku Inżynieria i Analiza Danych na wydziale MINI na Politechnice Warszawskiej.

1.1 Zadania

Nasze zadanie zostało podzielone na kolejne podzadania:

- EDA (Exploatory Data Analysis) - wejść w nasz dataset, zrozumieć jego cechy i znaleźć relacje je łączące za pomocą wizualizacji,
- Feature Engineering - przetworzyć oraz ewentualnie wyczyścić i wyselekcjonować cechy,
- Stworzenie modelu - przejście od prostych modeli do tych bardziej zaawansowanych oraz dobranie hiperparametrów w celu maksymalizacji metryk oceny modelu.

1.2 Walidacja

Przez cały okres tworzenia projektu walidowaliśmy projekt naszych kolegów oraz byliśmy przez kolejnych walidowani. Zadaniem walidatorów było ciągle sprawdzanie czy np. nie został popełniony gdzieś błąd logiczny w trakcie procesu oraz ewentualna pomoc w całym zagadnieniu. Ponadto otrzymali oni od nas 30% wszystkich danych, do których później nie mogliśmy zaglądać. Dane te miały posłużyć do ostatecznego testu naszego modelu.

1.3 Dane i cel

Naszą bazą były dane pochodzące z kampanii bezpośredniego marketingu prowadzonego przez portugalski bank. Oparto je o rozmowy telefoniczne prowadzone z potencjalnymi przyszłymi klientami.

Celem zadania była estymacja czy dany klient będzie chętny złożyć lokatę w

banku czy nie.
Link do danych - tu.

2 EDA

Nasza część datasetu składa się z 3164 obserwacji. Oto pierwszych pięć:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	37	entrepreneur	single	tertiary	no	1467	yes	yes	cellular	17	nov	349	1	-1	0	unknown	no
1	25	unemployed	single	secondary	no	34	no	yes	unknown	4	jul	316	1	-1	0	unknown	no
2	53	retired	married	secondary	no	1278	yes	no	telephone	15	jul	174	3	-1	0	unknown	no
3	31	housemaid	married	unknown	yes	-6	no	yes	telephone	7	jul	94	2	-1	0	unknown	no
4	43	unemployed	married	secondary	no	3529	no	no	cellular	5	feb	169	2	-1	0	unknown	no