

Data Analysis of Free Agent potential customers

Adam Babs

adam.babsm@gmail.com

Tasks

- Examine the provided data. Are there any differences in properties of behaviour between those customers that subscribe and those that do not?
- Construct a model to predict if a customer will subscribe after their free trial and demonstrate its performance.

Data summary

- Dataset contains a list of 950 customers.
- Each customer row contains information about:
 - company id
 - customer company type
 - flag indicating whether the user bought the subscription
 - number of weekly entries throughout the first week

Workflow

1. Extracting the number of weekly entries from the .log file
2. Analysis of order and correctness of the dataset, categorical data encoding
3. Data examination and finding important relations
4. Building a model and demonstrating its performance

Most important findings

- All users who visited the website more than 6 times throughout the week, bought the subscription
- 72% of user who visited the website exactly 6 times, bought the subscription (16/22)
- Only 2% of users who visited the website less than 6 times bought the subscription (20/803)

Conclusions drawn

- The company could try to encourage their clients to visit the system 7 times during the first week of the free trial

Model - Predicting if a customer will subscribe after their free trial

- Highest accuracy achieved: 98.42% - Logistic Regression (5-fold cross validation)
- Average accuracy achieved: 97.47% - Logistic Regression (5-fold cross validation average)
- Other classifiers used:
 - Linear classifier optimized by the Stochastic Gradient Descent
 - Linear regression
 - Multinomial Naïve Bayes algorithm