

Data mining in advertising:
air loyalty programs and
coupon purchase prediction

Dmitry Efimov

December 15, 2015

Outline

Predicting value customers in Air Loyalty programs

Coupon purchase prediction

Data preprocessing

Features











Models for prediction

About me

- ▶ PhD in Mathematics from Moscow State University
- ▶ The main specialization: machine learning and data mining
- ▶ TOP-10 in the global ranking on kaggle.com

Kaggle users are allocated points for their performance in competitions. This page shows the current global ranking. For more information on how we calculate points, please visit the [user ranking wiki page](#).

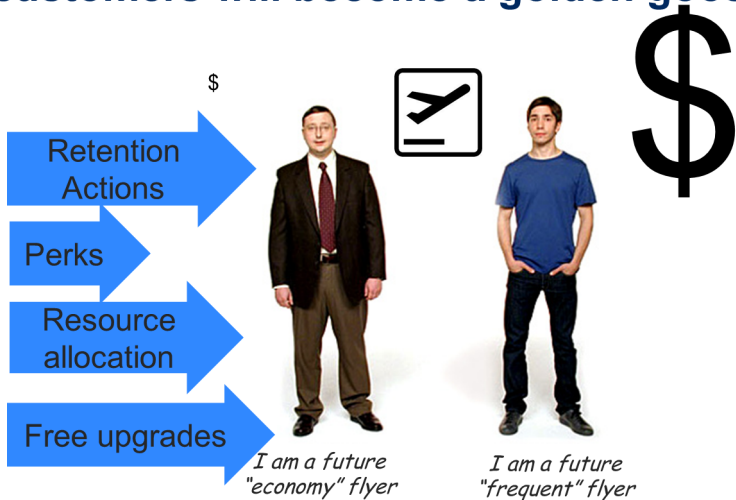
Search for users

1st 195,879 pts  Gilberto Titericz 52 competitions Sao Jose dos Campos Brazil	2nd 185,384 pts  Owen 41 competitions NYC United States	3rd 171,352 pts  Μαριος Μιχαηλιδης 61 competitions Volos Greece	4th 142,545 pts  Stanislav Semenov 25 competitions Moscow Russian Federation	5th 137,028 pts  Leustagos 41 competitions Belo Horizonte Brazil
6th 130,908 pts  Abhishek 89 competitions Berlin	7th 110,922 pts  Dmitry Efimov 33 competitions Moscow	8th 100,433 pts  José A. Guerrero 42 competitions Sevilla	9th 99,474 pts  Alexander Guschin 18 competitions Moscow	10th 97,560 pts  utility 11 competitions Moscow

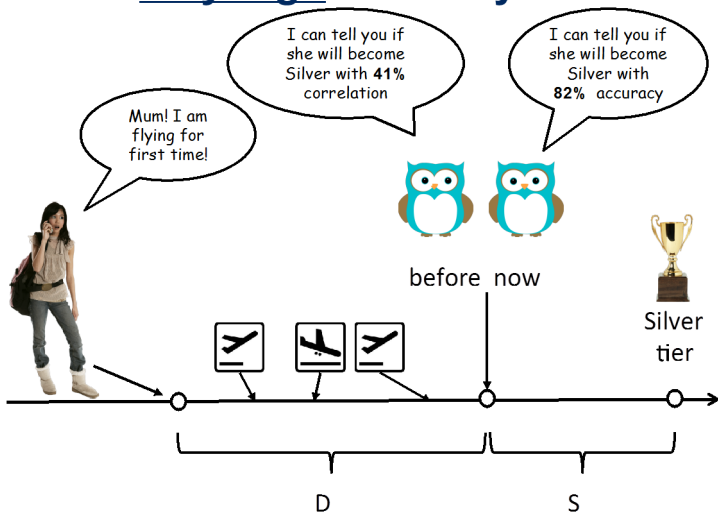
Predicting value customers in Air Loyalty programs

- ▶ Data are provided by Etihad Air Company
- ▶ The problem is to predict passengers that change tier status during the next few weeks
- ▶ The suggested methodology is patented by Dmitry Efimov and Jose Berengueres

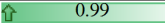
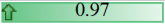
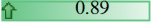
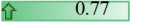
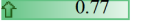
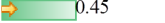
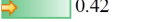
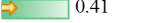
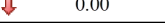
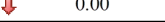
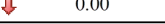
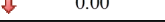
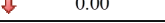
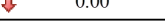
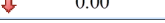
What-if your CRM could know who of your customers will become a golden goose?



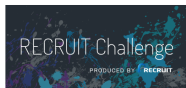
The D/S model can predict who will be high value with very high accuracy.



Example of an application that enhances the value of miles program:

Passenger Name	Tier Status	Probability they will become Silver in X months
Wolfgang Amadeus Mozart	Basic	 0.99
Ludwig Van Beethoven	Basic	 0.97
Giuseppe Verdi	Basic	 0.89
Jean-Michel Jarre	Basic	 0.77
Yasuharu Konishi	Basic	 0.77
Maki Nomiya	Basic	 0.45
Teresa Teng	Basic	 0.42
Carl Philip Emanuel Bach	Basic	 0.41
Enric Granados	Basic	 0.00
Leonard Bernstein	Basic	 0.00
Carl Loewe	Basic	 0.00
Johann Strauss	Basic	 0.00
Isaac Albeniz	Basic	 0.00
George Gerschwin	Basic	 0.00
John Williams	Basic	 0.00

Coupon purchase prediction



Completed • \$50,000 • 1,076 teams

Coupon Purchase Prediction

Thu 16 Jul 2015 – Wed 30 Sep 2015 (8 days ago)

Dashboard

Home

Data
Make a submission

Information

Description
Evaluation
Rules
Prizes
Timeline

Forum

Scripts

New Script

Leaderboard

Public
Private

My Team

Competition Details » [Get the Data](#) » [Make a submission](#)

Predict which coupons a customer will buy

Recruit [Ponpare](#) is Japan's leading joint coupon site, offering huge discounts on everything from hot yoga, to gourmet sushi, to a summer concert bonanza. Ponpare's coupons open doors for customers they've only dreamed of stepping through. They can learn difficult to acquire skills, go on unheard of adventures, and dine like (and with) the stars.

Investing in a new experience is not cheap. We fear wasting our time and money on a product or service that we may not enjoy or fully understand. Ponpare takes the high price out of this equation, making it easier for you to take the leap towards your first sky-dive or diamond engagement ring.

Using past purchase and browsing behavior, this competition asks you to predict which coupons a customer will buy in a given period of time. The resulting models will be used to improve Ponpare's recommendation system, so they can make sure their

Provided data

- ▶ Train: **user-coupon purchases** for 52 weeks
- ▶ Train: **user-coupon visits** for 52 weeks
- ▶ **User list**: gender, age, locations
 - ▶ \approx 20 000 users
- ▶ **Coupon list**: price, discount, genre name, locations
 - ▶ train: \approx 18 000 coupons
 - ▶ test: \approx 400 coupons

Predict: user-coupon purchases for the 53rd week

Cross validation

Problem: possible pairs user-coupon in train: $\approx 360\,000\,000$

- ▶ To decrease the size of train, for each week:
 - ▶ take coupons with at least one purchase
 - ▶ take users with at least one purchase
 - ▶ it gives $\approx 600\,000$ pairs for each week, or $\approx 30\,000\,000$ pairs for the whole train
 - ▶ use last few weeks to predict test week (we used last 5 weeks)
- ▶ Validation set: pairs for the last week

Feature engineering

- ▶ **Dummies:** one-hot encoding
- ▶ **Counts:** number of samples for different feature values
- ▶ **Counts unique:** number of different values of one feature for fixed value of another feature
- ▶ **Likelihoods**
- ▶ **Similarities**

Likelihood features

- ▶ **Type 1: using sliding window by weeks**

Example:

- ▶ for each week calculate the rate of purchases by each GENRE NAME based on the previous 10 weeks

- ▶ **Type 2: using multi class algorithms**

Example:

- ▶ one purchase — sample with target GENRE NAME
- ▶ for the test week: predict what will be next GENRE NAME
- ▶ ⇒ XGBOOST with 13 classes

Similarities

- ▶ the idea: for each user find the similarity between test coupons and coupons purchased before
- ▶ use cosine distance with weights
- ▶ coordinates for each coupon — coupon features

Final model and results

- ▶ XGBOOST with **rank:map** objective and **map@10** evaluation metric

Place	Team	Leaderboard score
1	Herra Huu	0.009973
2	Halla Yang	0.009848
3	threecourse	0.009484
...
20	Dmitry and Leustagos	0.007642

Thank you! Questions???

Dmitry Efimov
diefimov@gmail.com

