

# Ticketland ML Contest

Предсказание вероятности клика на шоу в поисковой выдаче сайта

Келаскин Максим  
1-е место

# План презентации

---

- **Постановка задачи**
- **Описание решения**
- **Используемые признаки**
- **Подведение итогов**

# Постановка задачи

---

## Задача


Предсказать вероятность клика на шоу, которое появилось в поисковой выдаче сайта Ticketland.ru

## Метрика качества

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

# Постановка задачи

## Пример поисковой выдачи



СПЕКТАКЛИ   КОНЦЕРТЫ   МЮЗИКЛЫ   КЛАССИКА


Москва ▾   цирк с конями   🔍

МЕРОПРИЯТИЯ — 0   ПЛОЩАДКИ — 2   ОРГАНИЗАТОРЫ — 0   ПЕРСОНЫ — 0

### Поиск по площадкам

Найдено 2 площадки

Цирки




Проспект Вернадского, 7

[Цирк на Проспекте Вернадского](#)

Ближайшие мероприятия:  
20 января / [Царевна-Несмеяна](#)

АФИША

Театры



Триумфальная площадь, 2

[Театр Сатиры](#)

Ближайшие мероприятия:  
20 января / [Роковое влечение](#)  
20 января / [Кошмар на улице Лурсин](#)  
20 января / [Невидимые миру слёзы](#)

АФИША

# Постановка задачи

---

## Исходные данные

**Impressions** – клики по поисковой строке (train/test)

**Clicks\_no\_impressions** – история кликов

**Client\_data** – данные пользователей

**Show\_data** – данные шоу

**Show\_images** – обложки шоу

**Show\_rating** – рейтинг шоу

# Описание решения

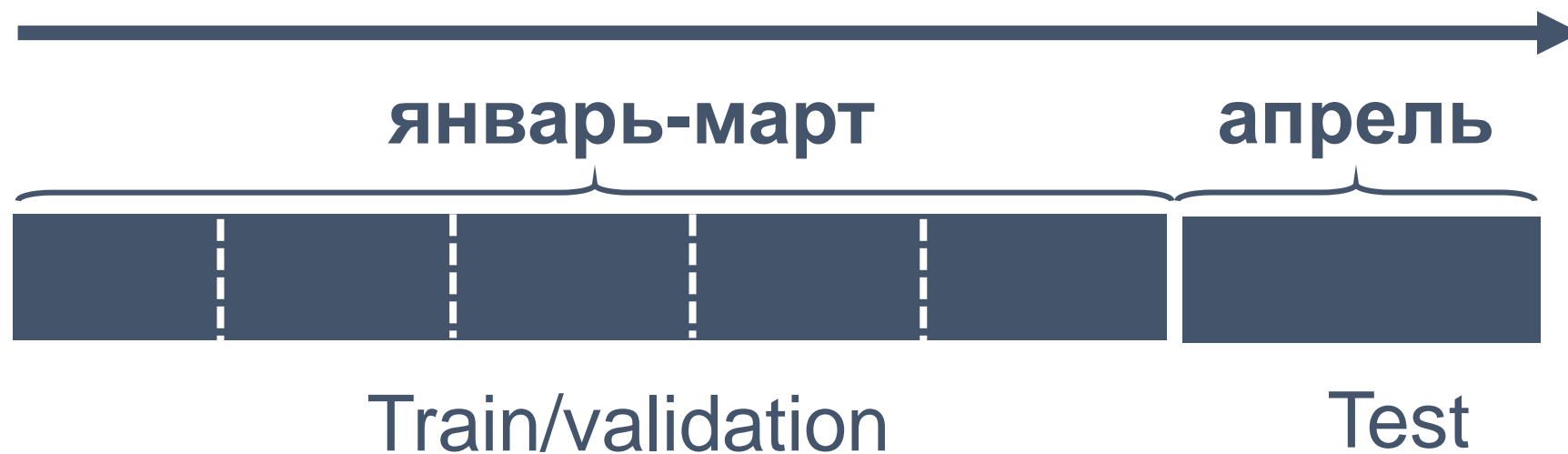
---

**Модель:** binary classification, 5 Folds CV blended (logfunc) XGBoost

---

**Использованные данные:** impressions + client\_data + show\_data

---



# Признаки

---

Всего было использовано 47 признаков

## Признаки “как есть”

rank  
event\_day\_of\_week  
event\_hour  
organizer\_id  
ldBuilding  
show\_maxprice  
show\_minprice  
show\_meanprice  
show\_stdprice  
is\_male  
is\_female  
13 dummy features из parent\_genre\_id

## Сгенерированные признаки

no\_actions\_before  
previuse\_action\_delta\_minutes  
past\_actions\_show\_count  
is\_in\_clients  
show\_day\_of\_week  
days\_to\_show

# Признаки

---

Всего было использовано 47 признаков

## Признаки “как есть”

rank  
event\_day\_of\_week  
event\_hour  
organizer\_id  
ldBuilding  
show\_maxprice  
show\_minprice  
show\_meanprice  
show\_stdprice  
is\_male  
is\_female  
13 dummy features из parent\_genre\_id

## Сгенерированные признаки

no\_actions\_before  
previuse\_action\_delta\_minutes  
past\_actions\_show\_count  
is\_in\_clients

**show\_day\_of\_week**  
**days\_to\_show**



# Признаки

---

## Информация о дате проведения шоу



# Признаки

---

## Статистика по минуте/сессии/дню

### Counters

session\_id  
session\_duration  
past\_actions\_session\_count  
past\_actions\_session\_show\_count  
show\_count\_session  
show\_minute\_count  
unique\_minute\_count  
total\_session\_counts

### Ranks

is\_top\_show\_dayly  
top\_rank  
show\_minute\_count\_%  
show\_%\_session  
show\_count\_day\_%  
rank\_show\_count\_%\_session  
rank\_show\_count\_day\_%\_session  
is\_top\_show\_%\_session  
is\_top\_show\_dayly\_session

# Признаки

---

## Статистика по минуте/сессии/дню

### Counters

session\_id  
session\_duration  
past\_actions\_session\_count  
past\_actions\_session\_show\_count  
show\_count\_session  
**show\_minute\_count**  
unique\_minute\_count  
total\_session\_counts



### Ranks

is\_top\_show\_dayly  
top\_rank  
show\_minute\_count\_%  
show\_%\_session  
show\_count\_day\_%  
rank\_show\_count\_%\_session  
rank\_show\_count\_day\_%\_session  
is\_top\_show\_%\_session  
is\_top\_show\_dayly\_session

# Итоги

## Итоговая модель

Public-logloss: 0.243776  
Private-logloss: 0.24425  
Total submits: 53

## Private лидерборд

#	Участник	Решений	Результат	Лучшее
1	Максим Келаскин	53	0.24425	2017-12-15 16:53:19
2	Василий Рязанов	17	0.252409	2017-12-17 22:55:02
3	Андрей Костин	27	0.252619	2017-12-17 17:30:11
4	Эмиль Каюмов	27	0.252923	2017-12-17 23:52:58
5	Макс Щёголев	24	0.254822	2017-12-17 14:04:49
6	Мирас	40	0.255858	2017-12-17 14:19:15

Thank you!