Setting up ML Projects: forecasting customers lifetime value

Yandex

Outline

- Intro to unit economics
- Discussing parts of ML project
  - Setting the goal
  - Getting data
  - Modelling
  - Deployment

Intro to unit economics

Background

- Yandex has a lot of customers both private person and businesses.

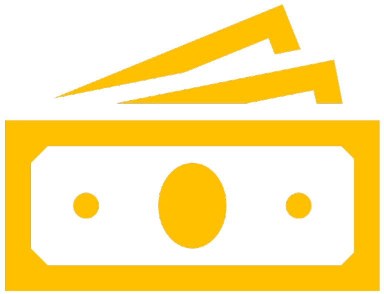
- LTV pipeline is implemented for both types of customers.

Unit economics – a calculation of profit and loss for a particular business model on a per-unit basis. Basically, it tells you how much value each item or unit creates for the business.

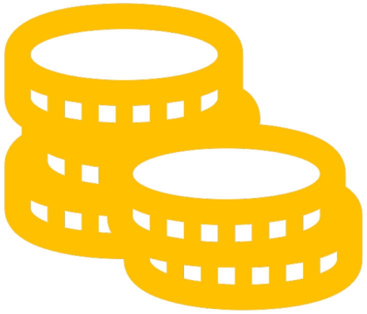
Unit = 1 client / user

The goal

- Estimate the effectiveness of sales and distribution channels.



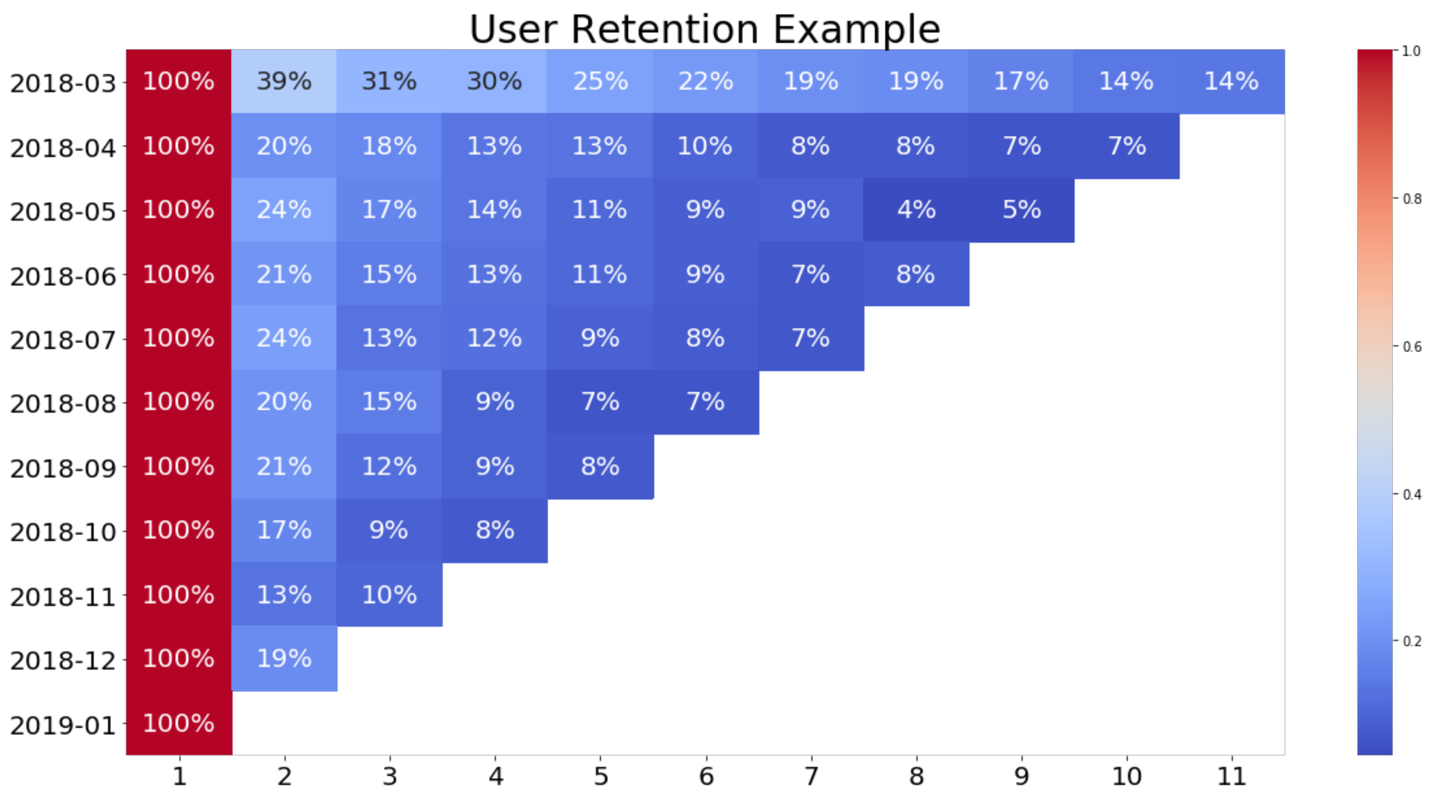
Assess the financial value of each customer.



In other words: How much do we earn from 1 customer

Definitions

- **Cohort –** a group of users, who made target action at the same time.



Definitions

- **LTV** (Life Time Value) – all the value that customer bring to the company for the whole period of his/her life
- When we talk about customer acquisition we keep in mind costs and payback time of invested money. Usually this period is 1 year.
- **CAC** (Customer Acquisition Costs) – total money spent on the customer acquisition divided by number of customers
- **ROI** = LTV / CAC – return on invested capital

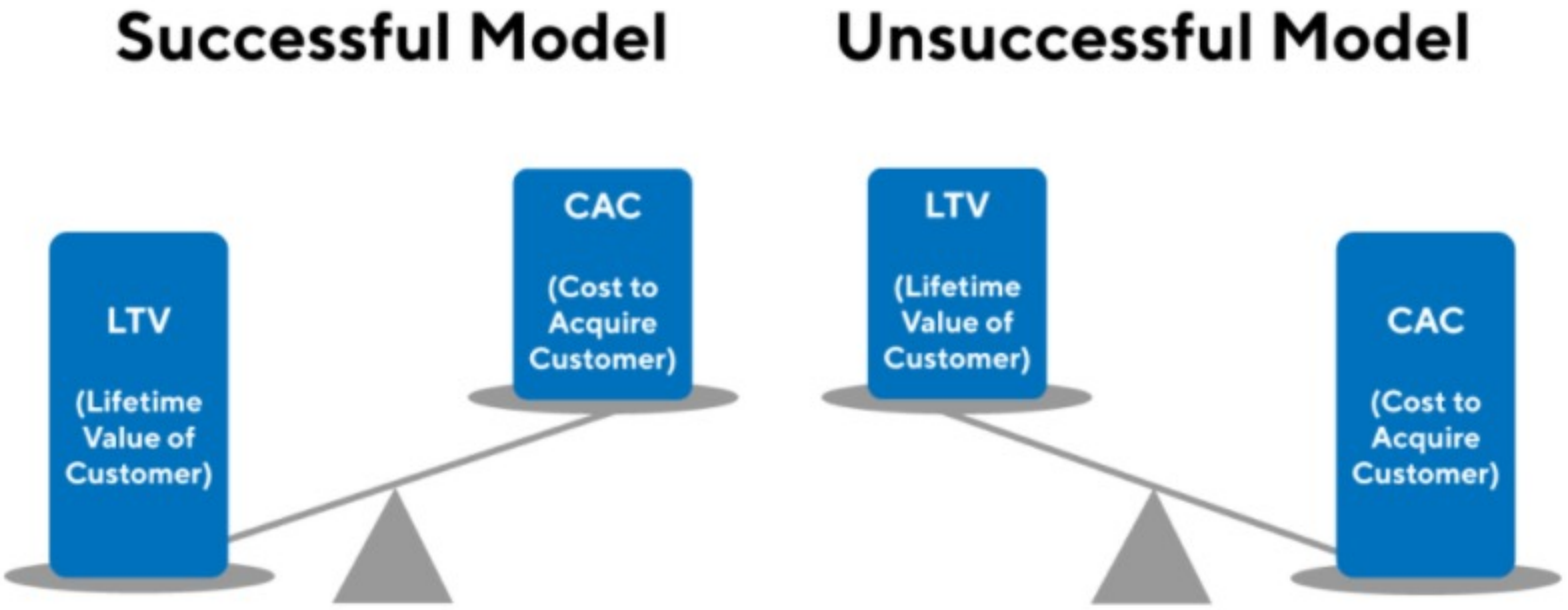
- **ROI** = LTV / CAC – return on invested capital
- CAC – is visible at the moment of acquisition
- LTV – longer indicator (have to wait some time or forecast)

Why

- Transparent argumentation for investments we do: explainable to CFO & CMO
- Comparison of customer acquisition channels
- Comparison of different products
- Comparison user cohorts in time (old vs new)

channel	activations	costs	CAC	LTV 365	ROI 365	LTV 60	ROI 60
cold sales	934	100000	107.06638	110000	1.1	15000	0.15
emails	459	45781	99.740741	38748	0.84637732	4649.76	0.10156528
pre-installs	781	8543	10.93854	7549	0.88364743	1509.8	0.17672949
total	2174	154324	70.986201	156297	1.01278479	21159.56	0.13711127

Importance of business model



https://www.productplan.com/glossary/customer-acquisition-cost/

Machine learning project: life cycle

Planning project

- Goals and aims
- Requirements
- Resource allocation
- Assessing feasibility of ML project
- Setting up the codebase

Data Collection

- Define objects
- Collecting trainings dataset
- Verification & correction data
- Dataset regularization

Training & validating

- Implementation several approaches
- Realizing of model limitations
- Sufficient accuracy
- Business features and risks

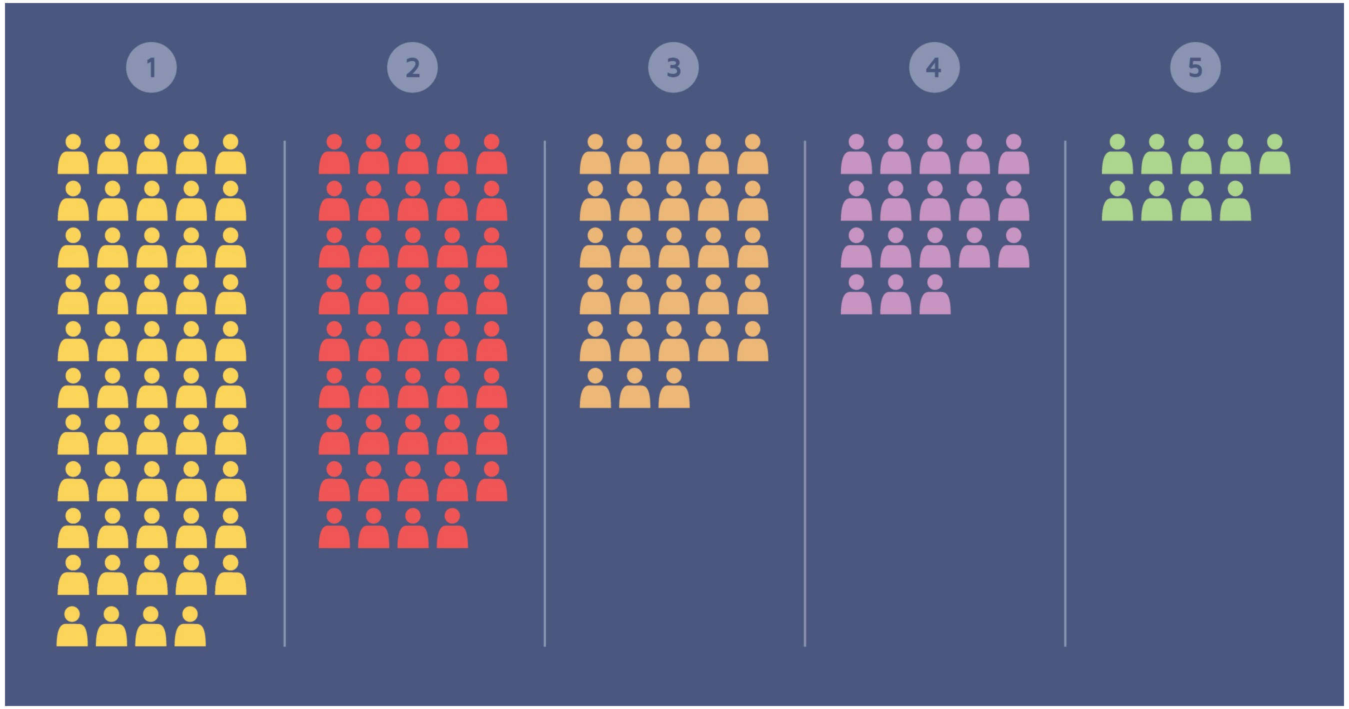
Deploying & monitoring

- Model monitoring systems
- Rolling model out

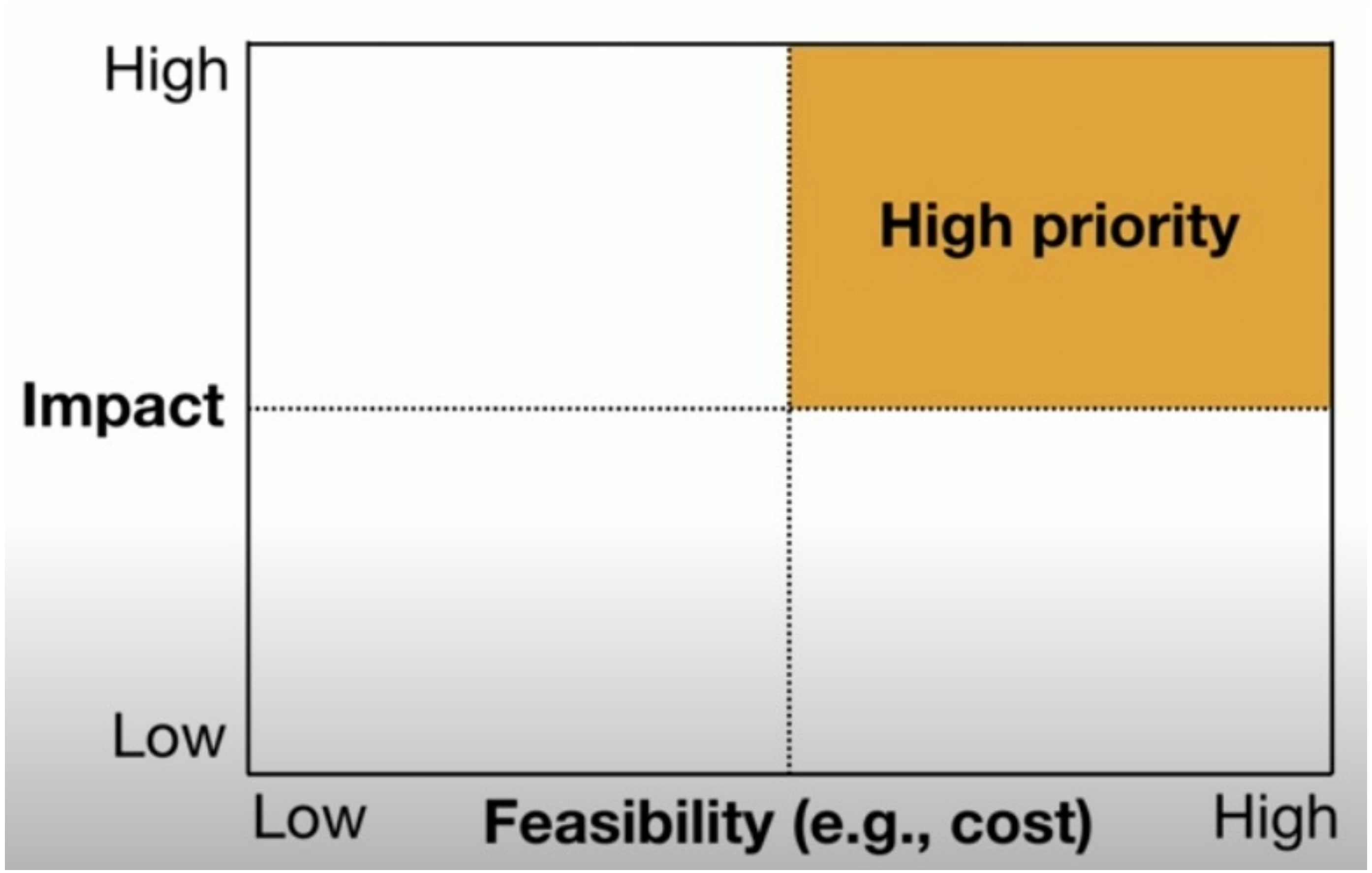
Planning project

Requirements

- Forecast of LTV at 365 and 60 days from event (may be activation or current date)
- Interested in cohort forecast (average of user-forecasts)
- Accuracy requirements (better than baseline)
- SLA: daily calculations, delays are critical



Prioritizing projects



Feasibility of ML project

Data availability

- How much data do we have
- How much data costs
- Is it hard to renew dataset regularly

Accuracy requirement

- What is the cost of error
- What is error margin

Problem complexity & sources

- How hard (novel) is the problem statement
- What computational time is needed to run model

Feasibility of ML project: LTV prediction

Data availability

- We have a lot of data for previous periods (years)
- Data is partially collected for another processes
- We can schedule regular calculation of new portion of data

Accuracy requirement

- The single (1-user) error doesn’t cost much
- The total error (channel-wise) costs a lot
- After all estimations +-20% accuracy (channel-wise error)

Problem complexity & sources

- The task of prediction LTV is common
- Computational time depends on model we use

Why do we need accuracy

- **ROI** = LTV / CAC

New service
Low CAC

low accuracy
requirement

it may be enough
if we guess only
magnitude of LTV

Mature service
High CAC

high accuracy
requirement

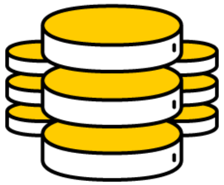
We compete for every
percent of ROI -> need
precise LTV forecast

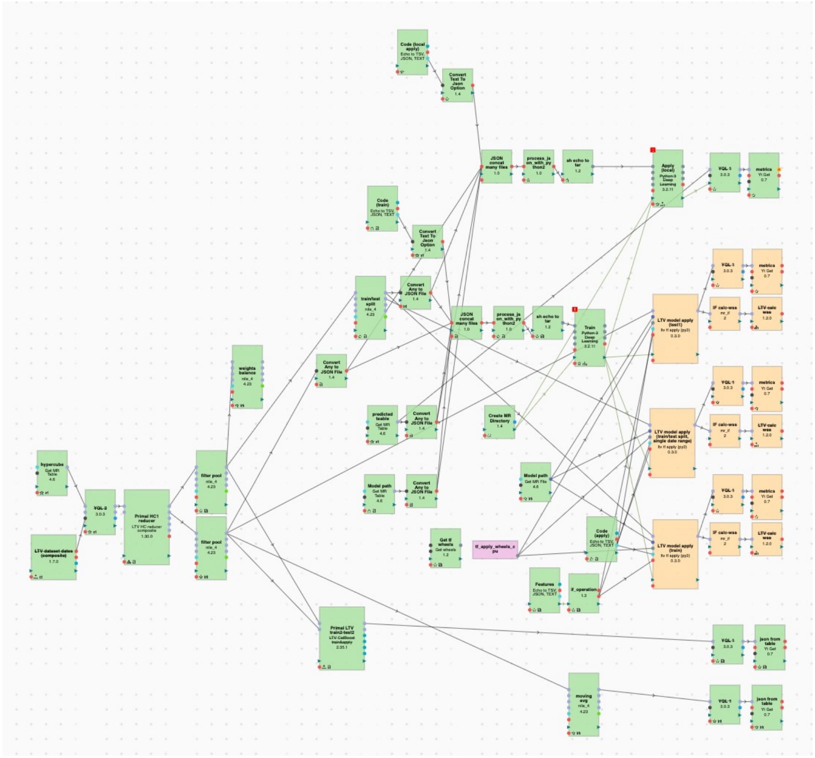
Setting up the codebase

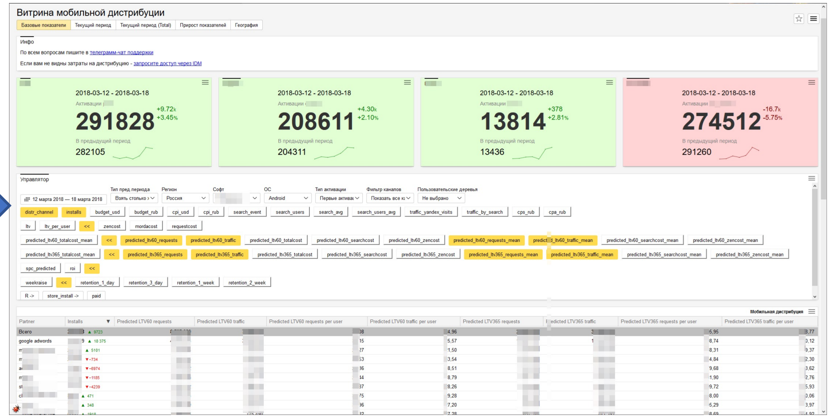
- MapReduce (Yandex cluster, inhouse solution)
- Python (libraries: CatBoost, TensorFlow)
- Cloud workflow platform (inhouse solution)
- Scheduler (inhouse solution)

Workflow

New dataset arrives







Scheduler

Data preparation and model

Dashboards

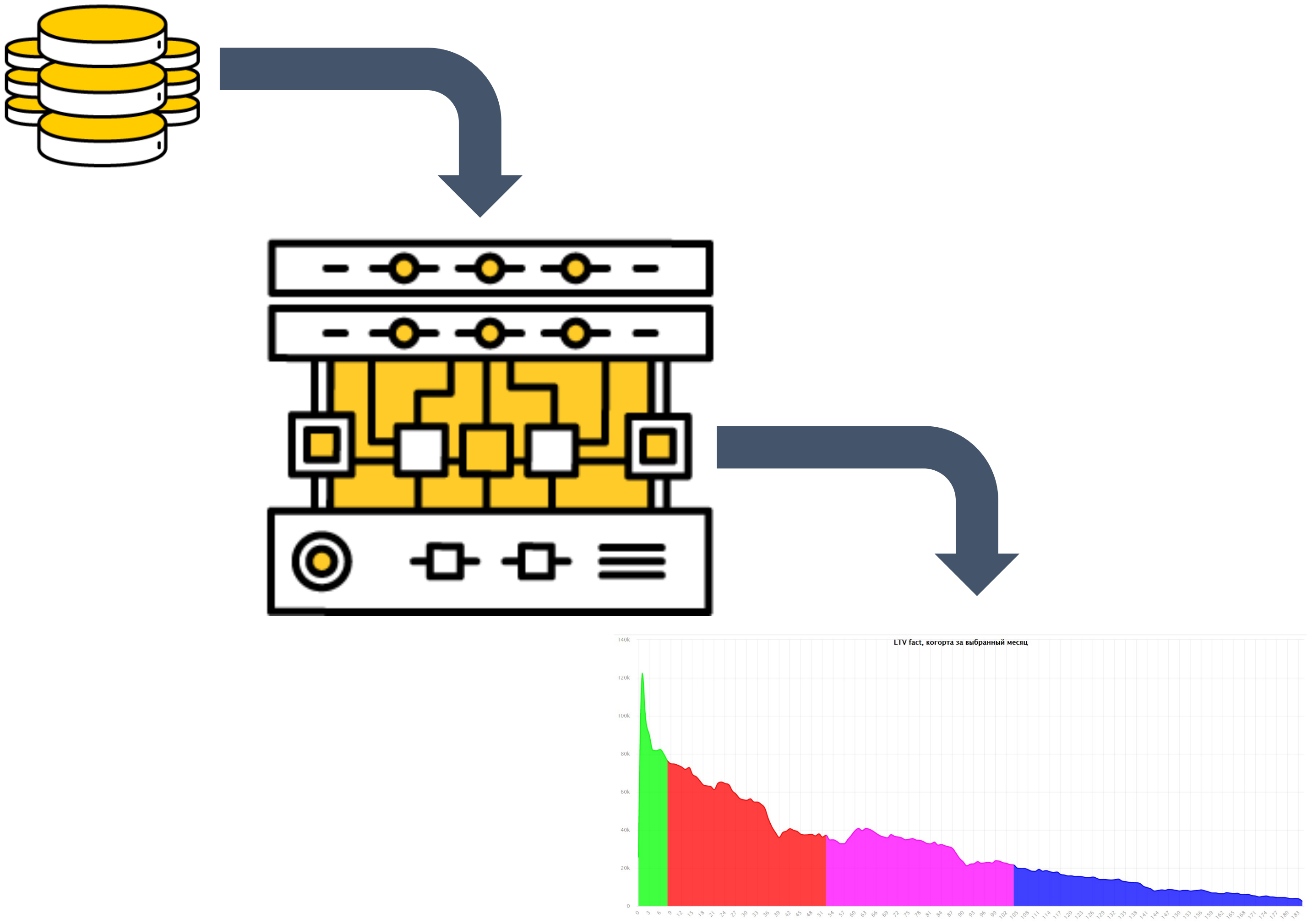
Planning summary:

- We are interested in cohort LTV forecast
- Project seems to be feasible:
  - We have necessary data
  - We have existing baseline
  - LTV is clear metrics to predict

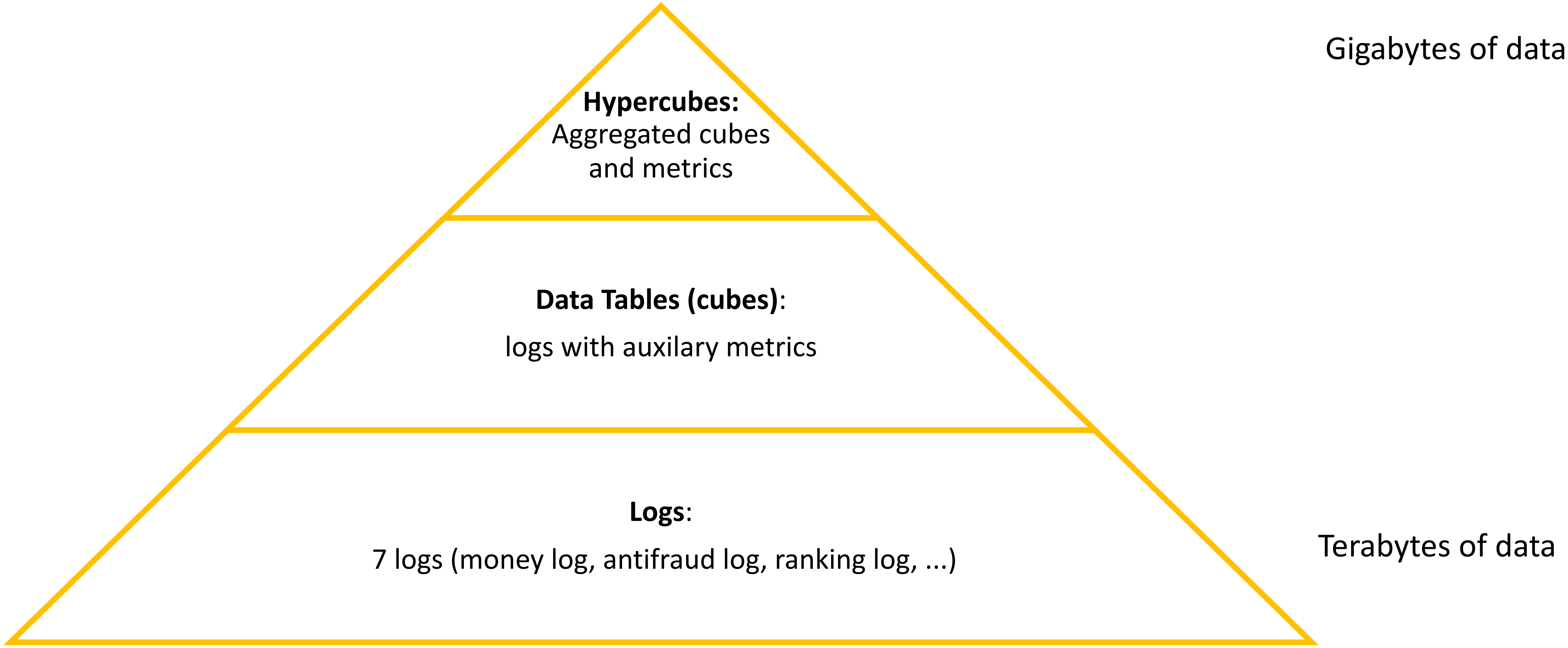
Data collection

Technical architecture

- In:
  - Aggregated data tables
- Out:
  - LTV forecasts (by user)
  - Logs (errors, metrics, …)
- After:
  - Dashboards
  - Tables with predictions



Raw model data



Raw date is subject to change

- ~90 days to arrive new data:
  - Fraud
  - Errors in collection pipeline
  - Logical errors

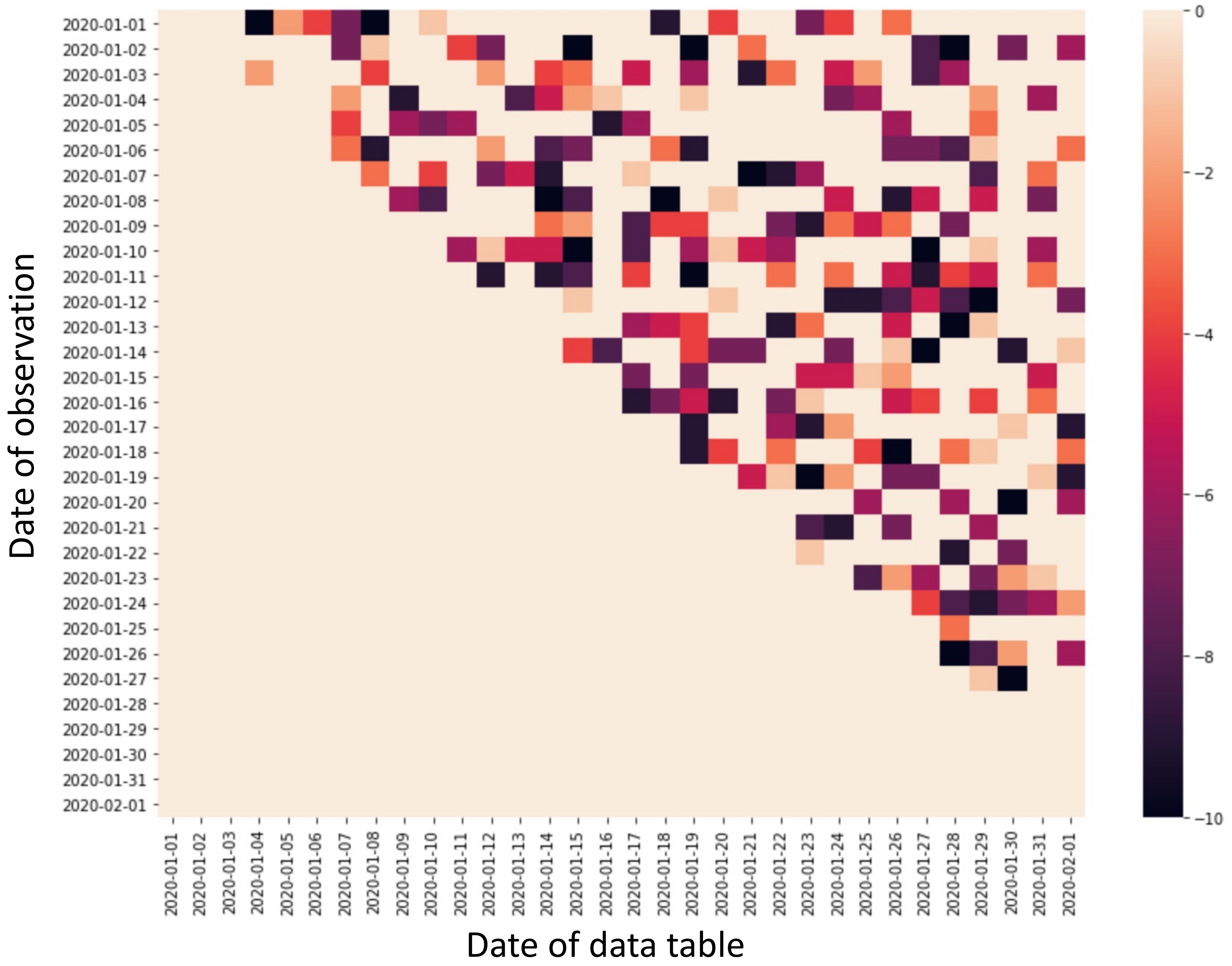
Monitoring of differences in data: 2-dimensional representation of time

- 1st Dimension: date of event
- E.g., activation cohort
- 2nd Dimension: calculation date
  - Forecast logging
  - Calculation and storing differences in fact data
  - "2D" monitorings (index data with tuple

    <date of observation, date of data table>

)

Monitoring of differences in data: 2-dimensional representation of time



Общие наблюдения за качеством ЛТВ

- 1. DWH имеет значение
- 2. Учитывайте все CAC вашего бизнеса, включая неявные (альтернативные)
- 3. Автоматически обновляйте прогнозы
- 4. Следите за показателями в динамике
- 5. Иногда больше фичей – хуже

Data collection summary

- Data pipeline matters!
- New portion of data can correct the old one -> need to live with this
- Collect as much features as it is possible -> extra features can be easily removed

Training and validation

Training and validation: baseline

Baseline – simple approach that gives lower bound for expected model performance.

Good practice: set existing approach as baseline.

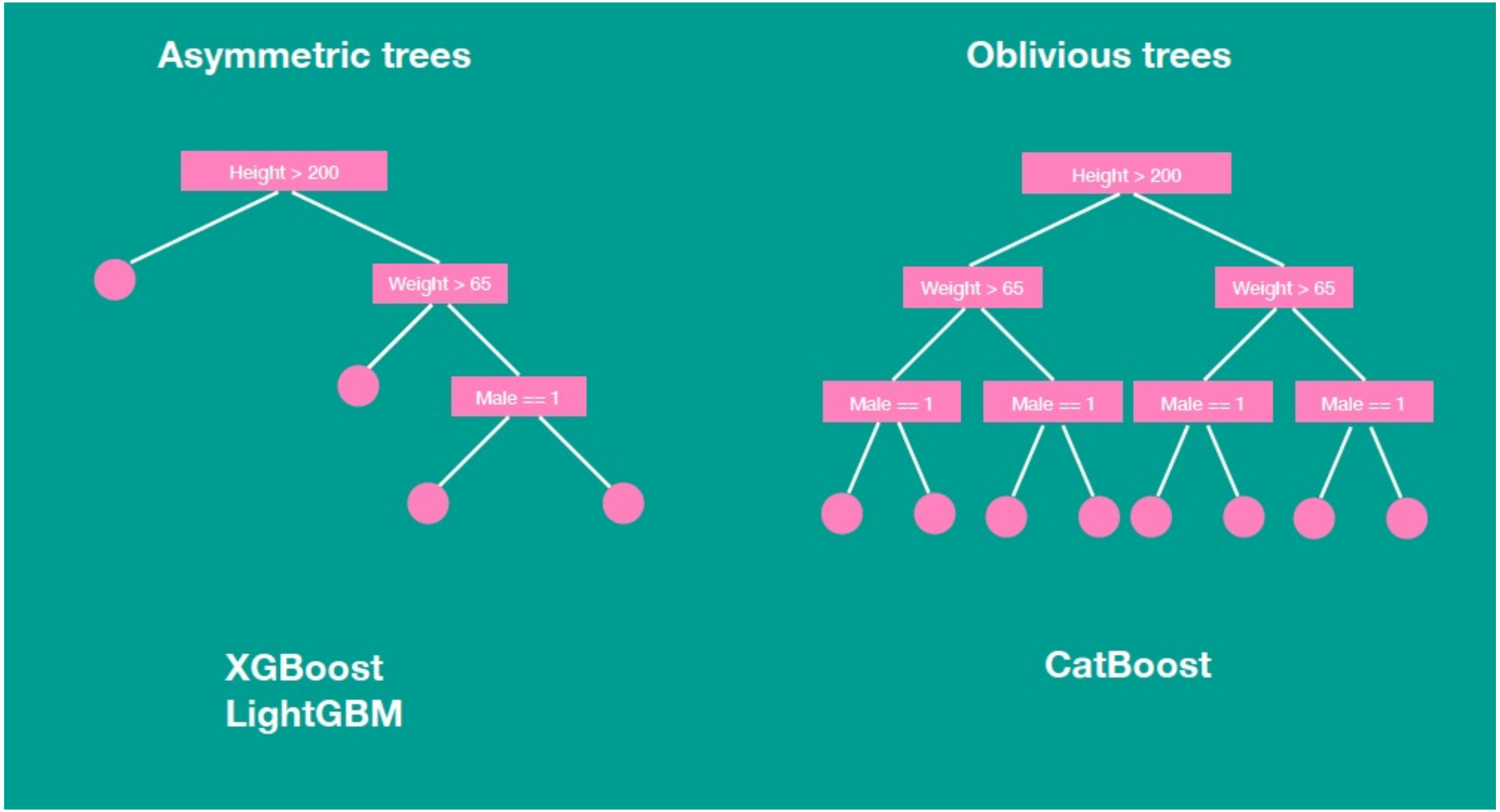
First baseline is moving average for N previous periods:

$$Money_i = \frac{\sum_{prev} Money}{N}$$

RMSE (baseline) = 3795

Training and validation: Model v.1

Yandex CatBoost



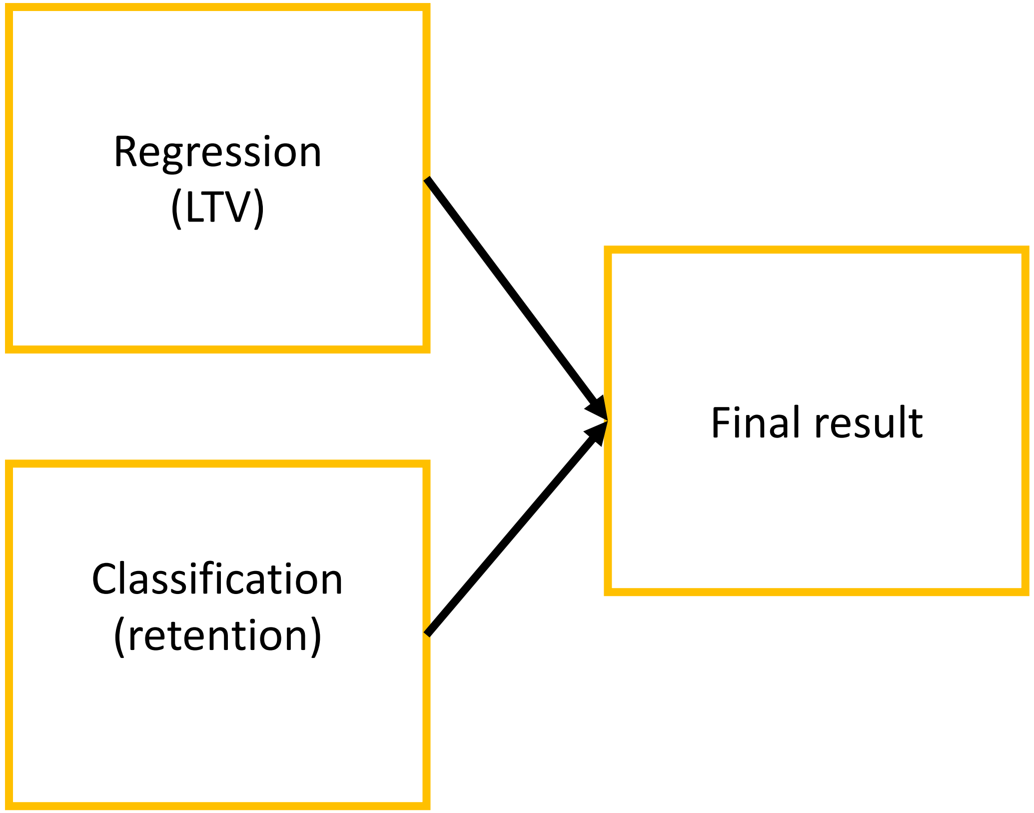
Model v.1 structure

Yandex
CatBoost

1st step: solve regression problem to estimate LTV

2nd step: solve classification problem to estimate if customer is going to pay or not

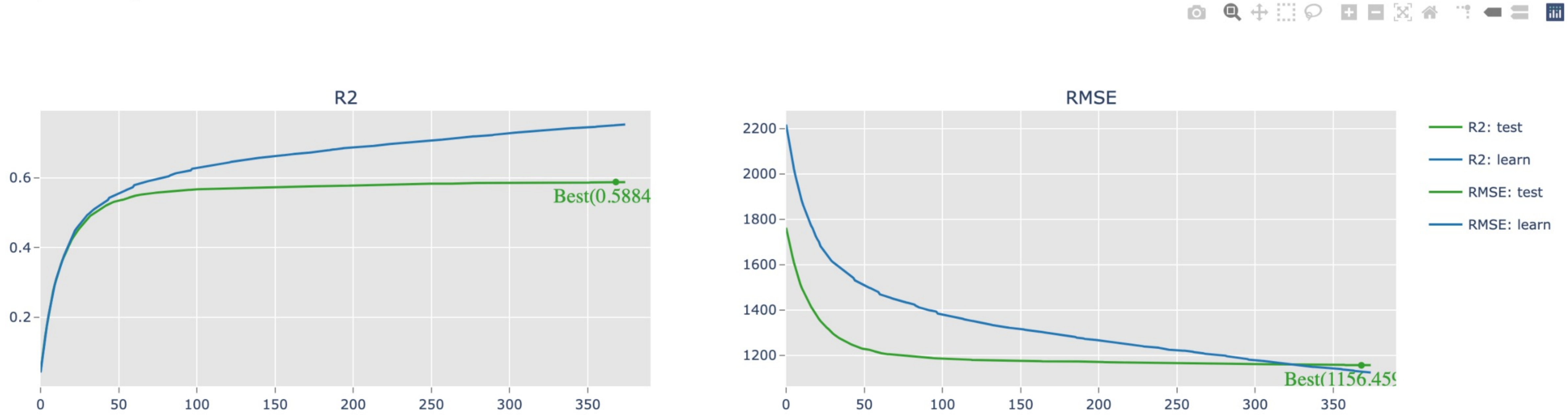
3rd step: combine these prediction in Final result



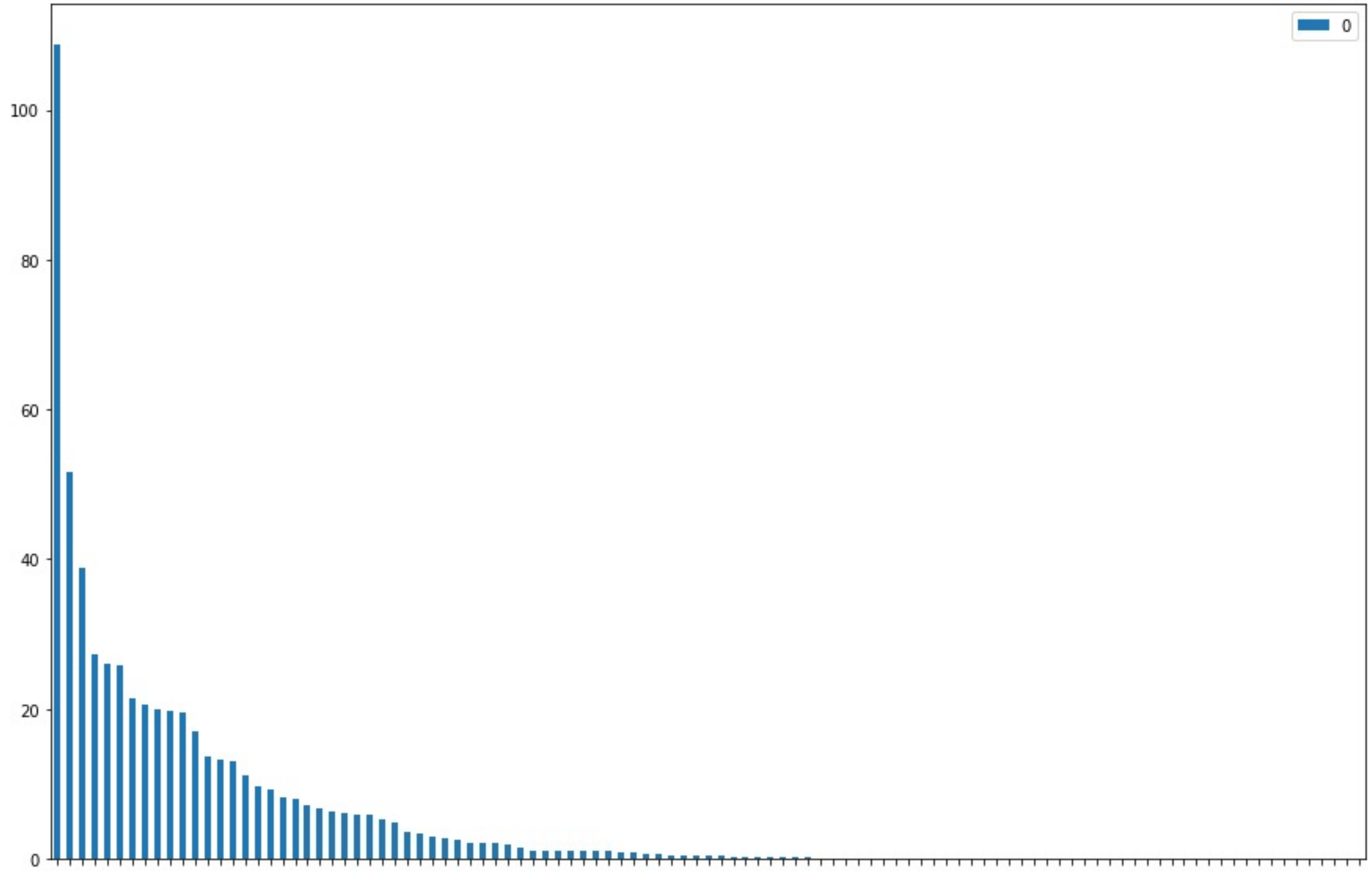
Metrics

For training we use RMSE. For reporting MAPE.

Best Values
R2: test(0.588482)
RMSE: test(1156.459549)



Feature importance regression



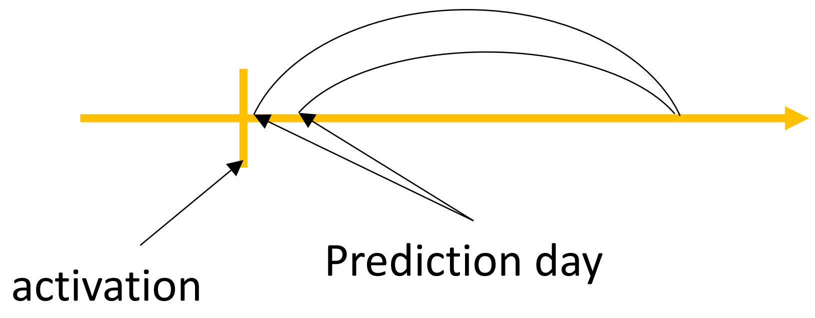
Heavily relies on features that correspond volume of previous activities

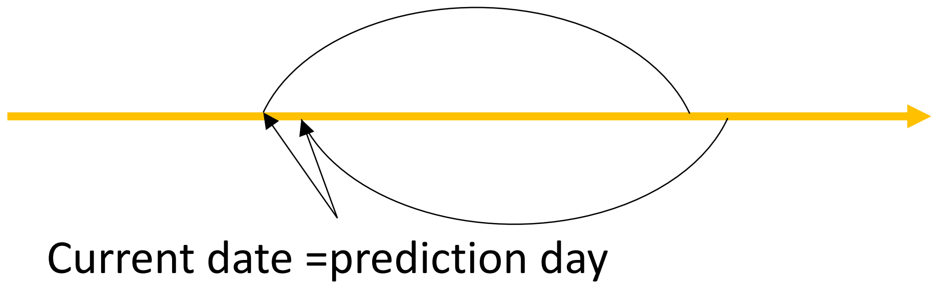
Feature importance classification

Heavily relies on features that correspond fact of previous activity

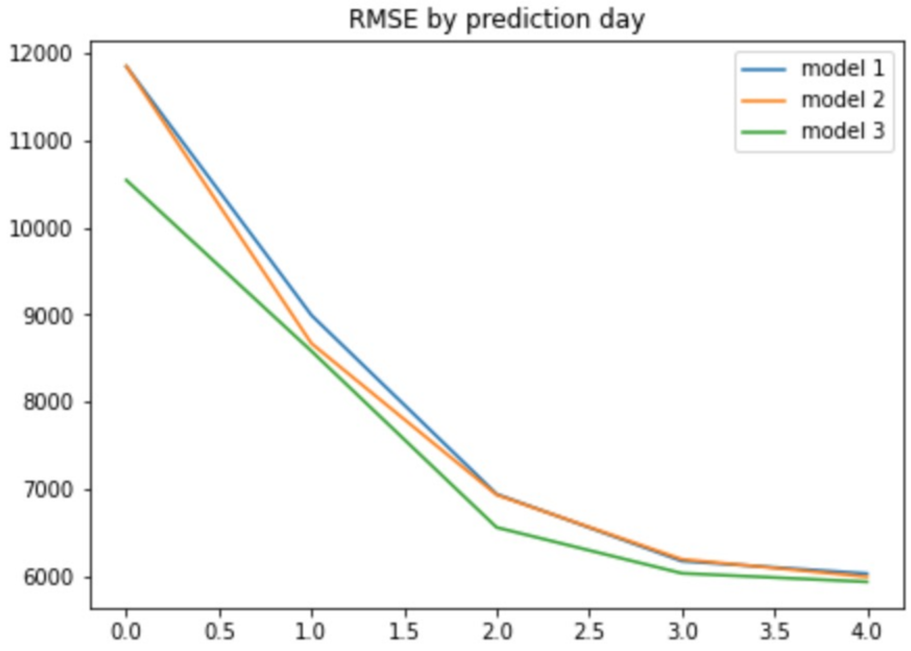
Parameter optimization

2 scenarios:

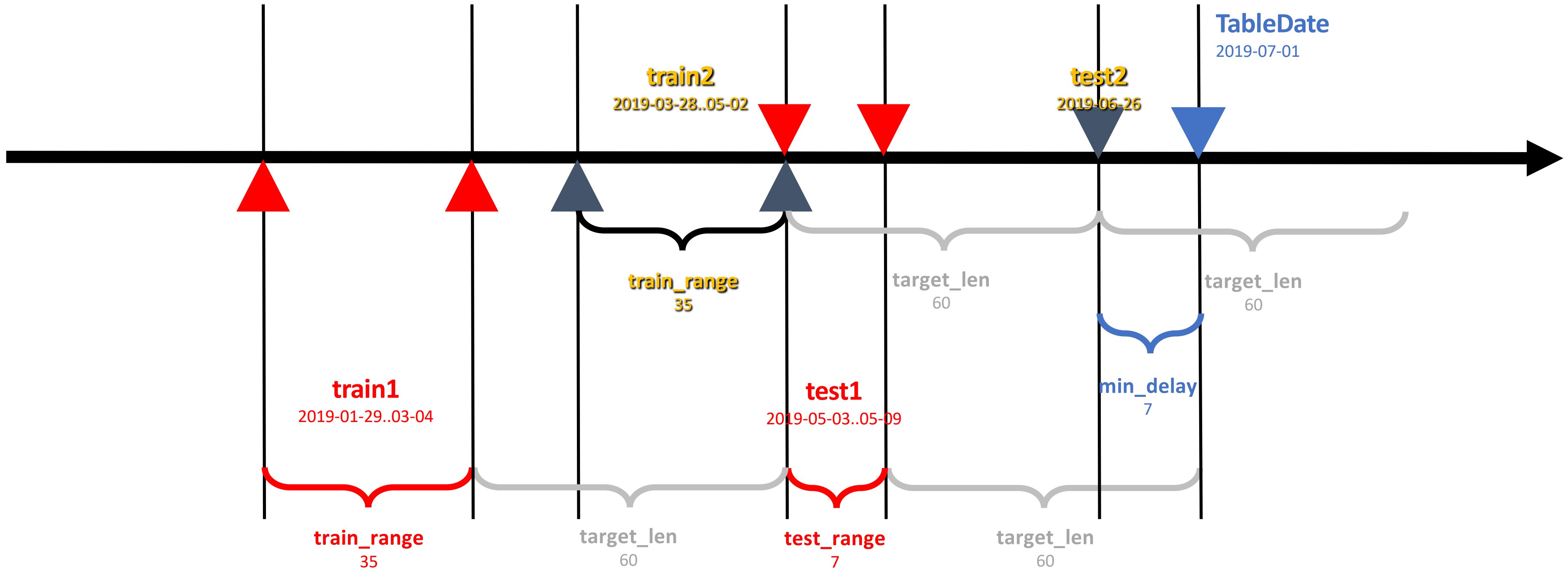




- Predict only the rest of window
- We get more accurate predictions for latest dates
- The value of predictions in first days is greater than in last days



Training and validation



Quality improvement retrospective

- Good quality at large channels (many samples)
- Poor quality at tiny channels (few samples)
- Solutions:
  - Add more features (geo, channel-attributes, social features) – FAILED
  - Smart sampling procedure (pick diverse samples – from different channels) – SUCCESS
- Reduce number of features: from 320+ to 80

Training and validation: Model v.2

- Baseline: CatBoost model
- RMSE improved: 30% less than CatBoost

These are metrics for old clients:

RMSE CatBoost	RMSE TF model
3795	2651

Quality of forecast in time

- The data can be modified (due to various things) -> we have to recalculate forecasts
- Log previous forecasts “for the record”

Training and validation summary

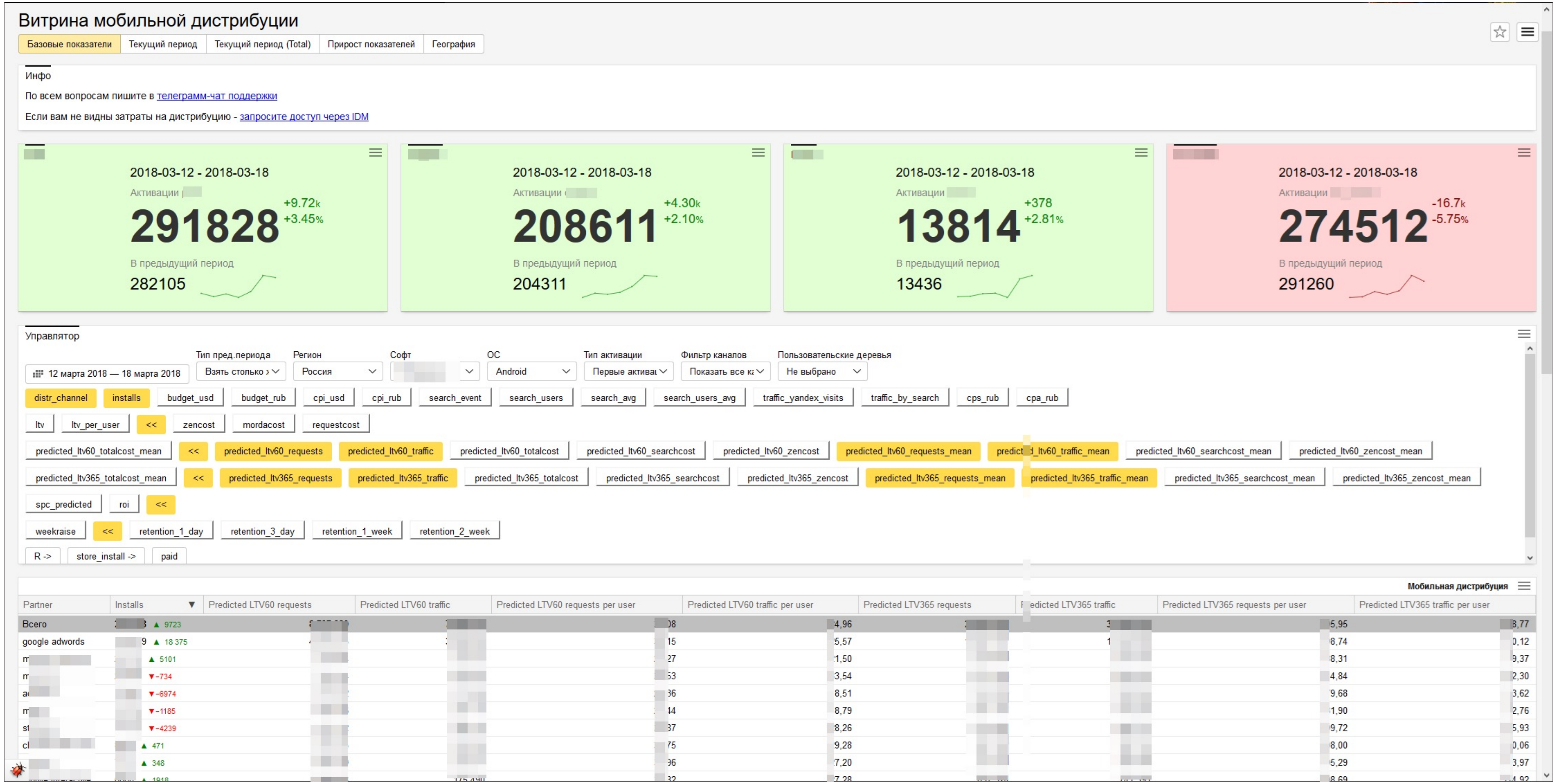
- Baseline matters!
- Not only train and validate model but go through scenarios of pipeline (data modifications in history, SLA, …)

Deploying and monitoring

Monitoring

- Cases:
  - Anomalies in data
    - Create metrics linked to data (e.g. number of rows in dataset, average value in columns, …)
    - Monitor this metrics behavior (deviation from usual values, length of this deviation, …)
  - Data structure changes
    - Changes are significant and model calculation fails – good scenario, we immediately check this problem;
    - Changes don’t lead to computational failure -> we don’t see this
  - Not enough computing power
    - restart

Dashboards


Витрина мобильной дистрибуции
Базовые показатели
Текущий период
Текущий период (Total)
Прирост показателей
География
Инфо
По всем вопросам пишите в телеграмм-чат поддержки
Если вам не видны затраты на дистрибуцию - запросите доступ через IDM
2018-03-12 - 2018-03-18
Активации
291828
+9.72k
+3.45%
В предыдущий период
282105
2018-03-12 - 2018-03-18
Активации
208611
+4.30k
+2.10%
В предыдущий период
204311
2018-03-12 - 2018-03-18
Активации
13814
+378
+2.81%
В предыдущий период
13436
2018-03-12 - 2018-03-18
Активации
274512
-16.7k
-5.75%
В предыдущий период
291260
Управлятор
12 марта 2018 — 18 марта 2018
Тип пред.периода
Взять столько ж
Регион
Россия
Софт
ОС
Android
Тип активации
Первые активац
Фильтр каналов
Показать все ка
Пользовательские деревья
Не выбрано
distr_channel
installs
budget_usd
budget_rub
cpi_usd
cpi_rub
search_event
search_users
search_avg
search_users_avg
traffic_yandex_visits
traffic_by_search
cps_rub
cpa_rub
ltv
ltv_per_user
zencost
mordacost
requestcost
predicted_ltv60_totalcost_mean
predicted_ltv60_requests
predicted_ltv60_traffic
predicted_ltv60_totalcost
predicted_ltv60_searchcost
predicted_ltv60_zencost
predicted_ltv60_requests_mean
predicted_ltv60_traffic_mean
predicted_ltv60_searchcost_mean
predicted_ltv60_zencost_mean
predicted_ltv365_totalcost_mean
predicted_ltv365_requests
predicted_ltv365_traffic
predicted_ltv365_totalcost
predicted_ltv365_searchcost
predicted_ltv365_zencost
predicted_ltv365_requests_mean
predicted_ltv365_traffic_mean
predicted_ltv365_searchcost_mean
predicted_ltv365_zencost_mean
spc_predicted
roi
weekraise
retention_1_day
retention_3_day
retention_1_week
retention_2_week
R ->
store_install ->
paid
Мобильная дистрибуция
Partner
Installs
Predicted LTV60 requests
Predicted LTV60 traffic
Predicted LTV60 requests per user
Predicted LTV60 traffic per user
Predicted LTV365 requests
Predicted LTV365 traffic
Predicted LTV365 requests per user
Predicted LTV365 traffic per user
Всего
google adwords

Other scenarios

- We provide our user based forecasts for a lot of purposes:
  - Validating hypothesis (A/B without control)
  - Planning headcount
  - Calculating dynamics

Deploying model

- Monitor your solution!
- Provide results in people-friendly format :)