



Продуктивизация ML- пайплайнов

вопросы и ответы на кейсах М.Видео-Эльдорадо

Карев Виталий

Разработчик Центра компетенций по
техническим решениям для
машинаного обучения, Группа
М.Видео-Эльдорадо

Топ-10



Крупнейшие ритейлеры
электроники в мире

~100м ежемесячный трафик²

#1



Онлайн-ритейлер
электроники в России³

33% доля рынка онлайн³

Ведущий



Ритейлер электроники в
России³

27% доля рынка в 2020³

19м



Активных клиентов⁴

72м база лояльных
клиентов⁵

~60%



Доля онлайн-продаж в
обороте

~1м м²



Площадь фулфилмент-
центров⁶

100%



Узнавание брендов⁷

RUB 0.5trln



GMV⁹

+15% год-к-году в 2020

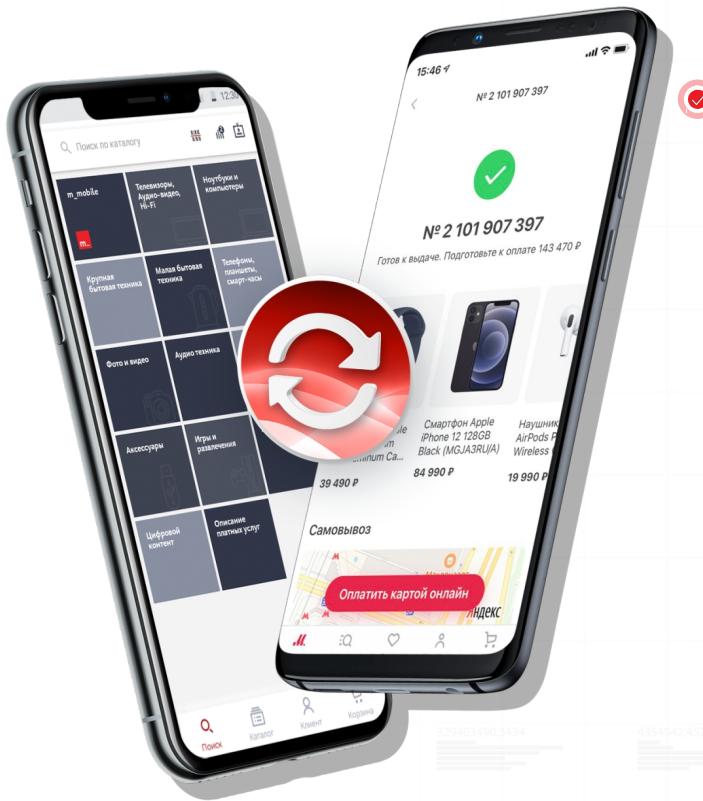
~5-7%



Показатель
маржинальности по EBITDA

(1) Across global CE retailer specialists by revenue in 2019, data from Thomson Reuters; (2) Average monthly traffic for website, app and stores in FY 2020; (3) Data for FY 2020 market share, GFK; (4) Identified active customer denotes identified customers with at least one purchase for the last 12 months, data as of 31 Dec 2020; (5) Loyal customers represent a group of authorized customers with active loyalty cards; (6) Warehouse area in stores of ~540k м², distribution centers of ~418k м², and regional distribution centers of ~83k м² as of 31 Dec 2020; (7) Brand Health Tracking, 260 cities of presence M.Video and Eldorado stores (100k+ population), base 2020, customers CE during the last 6 months (9,599 respondents, Male / Female, 16-65 years), Millward Brown A/R M/I-Marketing; (8) Q1 data for 2020, base 39/35 M.Video / Eldorado stores, survey at the exit from the store "M&P ANALYTICS LLC"; (9) Gross Merchandise Value (GMV) includes purchases in retail stores (including pick up orders), paid and delivered online orders, paid shipments to legal entities from warehouses. Purchases made in stores and online orders may be made by individuals and legal entities. GMV includes VAT, excludes returns and discounts offered to customers for the reporting period. GMV is not Revenue of the Company; (10) IAS 17 EBITDA margin in 2017-2019;

Source: Company data, GFK "Consumer Electronics industry report" (Jan 2021), M&P ANALYTICS LLC, Millward Brown A/R M/I-Marketing



Взаимодействие с клиентами

Машинное обучение

Алгоритмы умного поиска

Pick Up using chat bot and machine learning

Цифровизация бизнес-процессов

Проблемы продуктивизации

- Нетипичное описание пайплайнов для DevOps-инженеров
- Организация хранения и версионирования артефактов
- Необходимость типизации потоков данных

Нетипичное описание пайплайнов для DevOps-инженеров

AirFlow DAGs Tools▼ Browse▼ Admin▼ Docs▼

DAG: example1

Tree View Graph View Task Duration Landing Times Gantt Code

example_dags/example1.py

```
from airflow.operators import BashOperator, DummyOperator
from airflow.models import DAG
from datetime import datetime

args = {
    'owner': 'airflow',
    'start_date': datetime(2015, 1, 1),
}

dag = DAG(dag_id='example1')

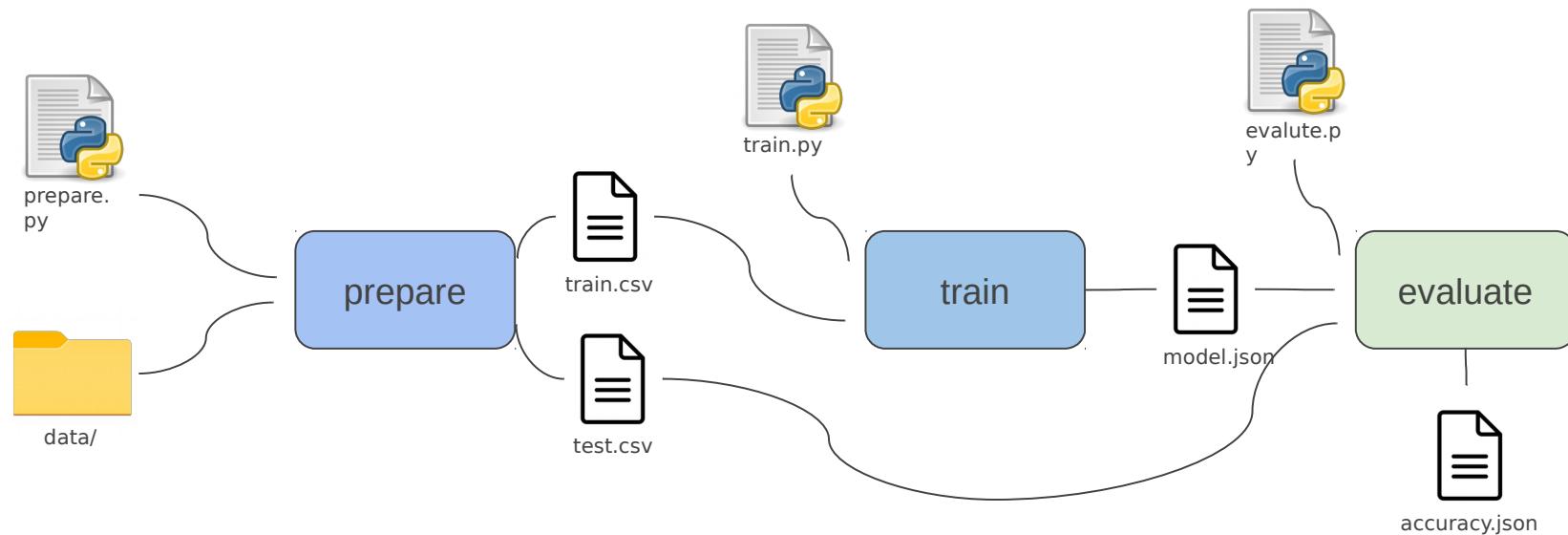
cmd = 'ls -l'
run_this_last = DummyOperator(
    task_id='run_this_last',
    default_args=args)
dag.add_task(run_this_last)

run_this = BashOperator(
    task_id='run_after_loop', bash_command='echo 1',
    default_args=args)
dag.add_task(run_this)
run_this.set_downstream(run_this_last)
for i in range(9):
    i = str(i)
    task = BashOperator(
```

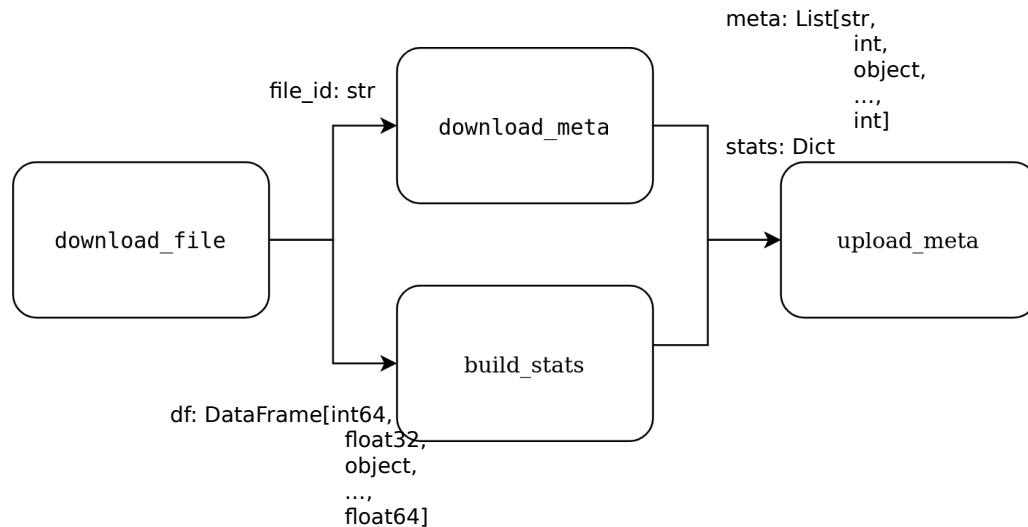
VS

```
1 version: 1.2
2
3 PROJECT_NAME: W2V_ON_BROWSING
4 PIPELINE_NAME: W2V_WINDOW_4
5
6
7 context:
8   run_params:
9     week_start: $week_start
10    week_end: $week_end
11    window_size: $window_size
12    categories_path: settings/categories/
13    context_generator:
14      - "cubes/generate_context.py:gen_db_configs"
15
16 steps:
17   getMeta:
18     exec: cubes/common.py:GetMeta
19     context:
20       cities_path: settings/cities.json
21     RESOURCES:
22       MEMORY_LIMIT: 261
23
24 checkTable:
25   exec: cubes/common.py:CheckTable
26   named_inputs:
27     categories: getMeta[categories]
28
29 category_predictors:
30   generate_with: cubes/gen_w2v_learn.py:generate_w2v_learn
```

Хранение и версионирование артефактов



Типизация потоков данных



В итоге

Реализовано и активно используется в production решение, которое покрывает все перечисленный проблемы:

- Поддержку ML-пайплайнов осуществляют DevOps-инженеры
- Хранение и версионирование артефактов реализуются незаметно для ML-разработчиков
- В продакшне используются типизация потоков данных, в том числе DataFrame`ов (для дева типизация опциональна)

Спасибо за внимание!

По возникшим вопросам можно обратиться по адресам:

Yuliya.Lavrushkina@mvideo.ru

Evgeny.Vorobey@mvideo.ru