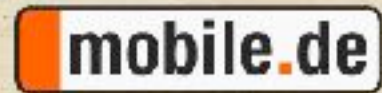


Building prediction pipelines that rocks in the real world

Albert Gorski



WELCHER IST DEIN NÄCHSTER?

About me



Albert Gorski

Lead Backend Engineer at mobile.de GmbH

 albgorski



Germany's largest online vehicle marketplace

13.5M
Unique Users per
Month

43K
Dealers

2M
Items on Page



ebay[™]
classifieds
group



Bilbasen

dba



kijiji

2 Marktplaats



vivanuncios
Bonds compras y vendas rápido



Collecting data

H2O

Checking candidate



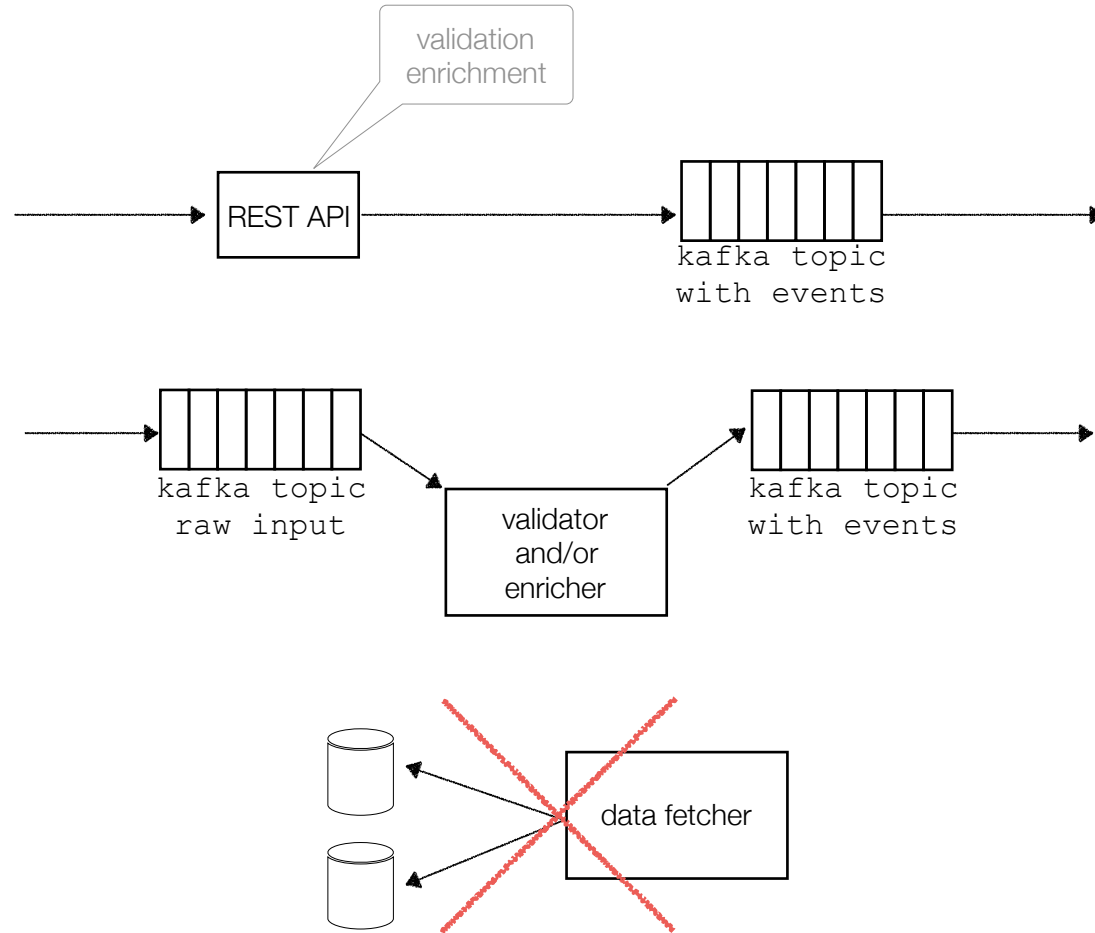
Building the model

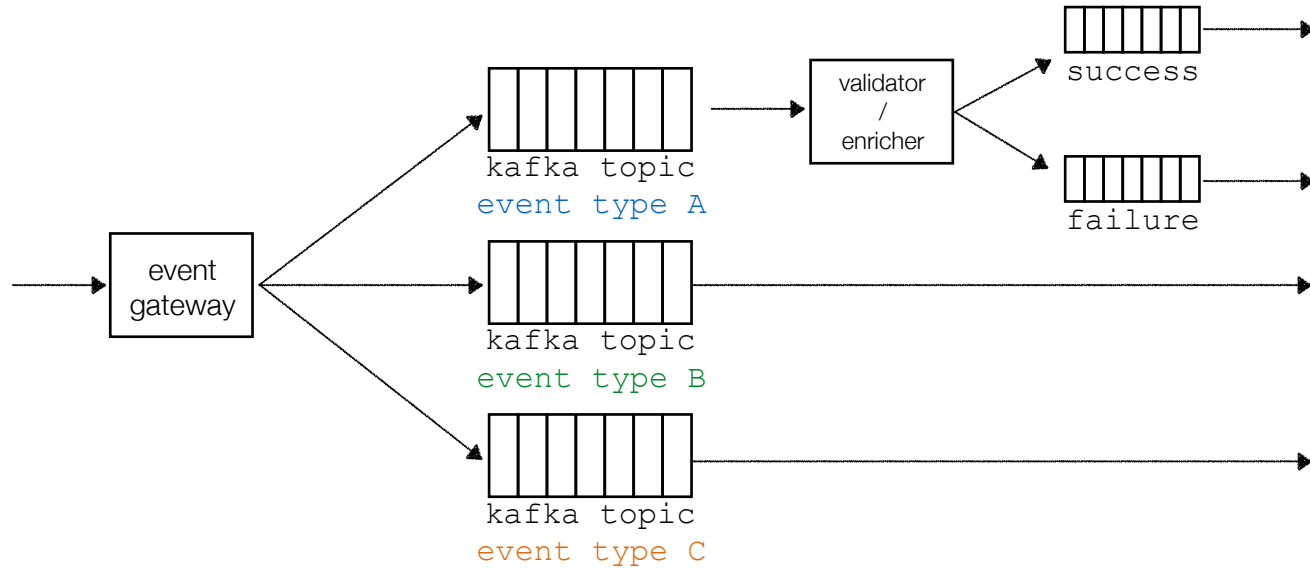
Serving model

Collecting data

mobile.de

events





Collecting data

events

context

schema

schema

```
{
  "info": {
    "title": "General title",
    "version": "1",
    "contact": {
      "email": "xyz@mobile.de"
    }
  },
  "definitions": {
    "MyEvent": {
      "description": "MyEvent description.",
      "type": "object",
      "required": [
        "head", "msg"
      ],
      "properties": {
        "head": {
          "$ref": "../types/Common.json#/definitions/Head"
        },
        "msg": {
          "$ref": "#/definitions/Message"
        }
      }
    },
    "Message": {
      "description": "Describing the message part",
      "type": "object",
      "required": [
        "dataOne"
      ],
      "properties": {
        "dataOne": {
          "description": "data one description",
          "type": "string",
          "example": "one value"
        },
        "dataTwoMaybe": {
          "description": "data two description",
          "type": "string",
          "example": "two value"
        }
      }
    }
  }
}
```

context

context

Collecting data

events

context

schema

external data providers

Building the model

filtering and pre-processing

feature engineering

sampling

experiment with algorithms

check, tune and repeat!

H2O

Python, R, Java

split models

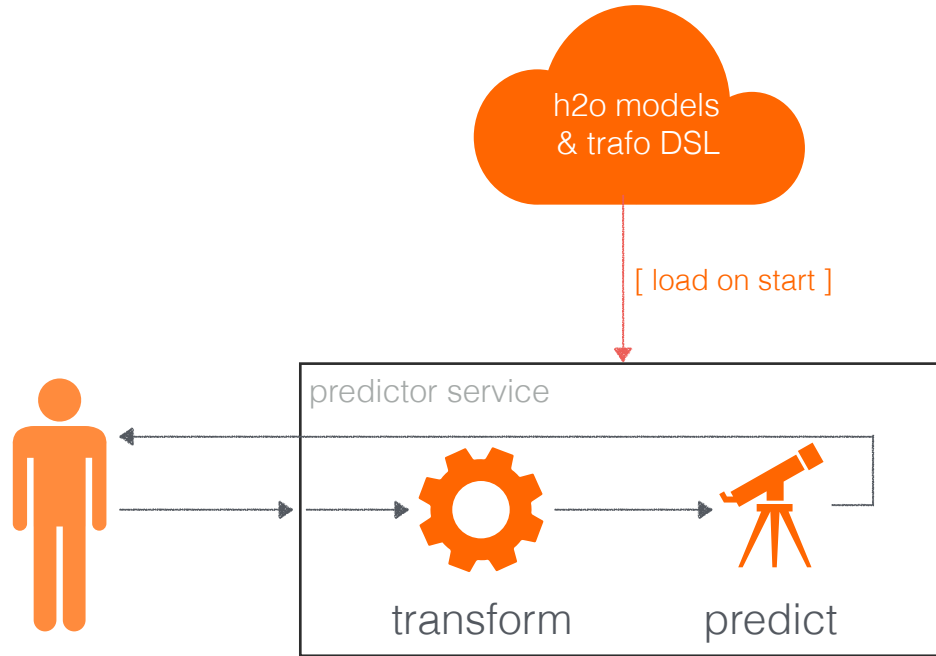
export as MOJO

Serving model

load model(s) on start

Anti-Corruption Layer

scala, akka-streams, akka-http



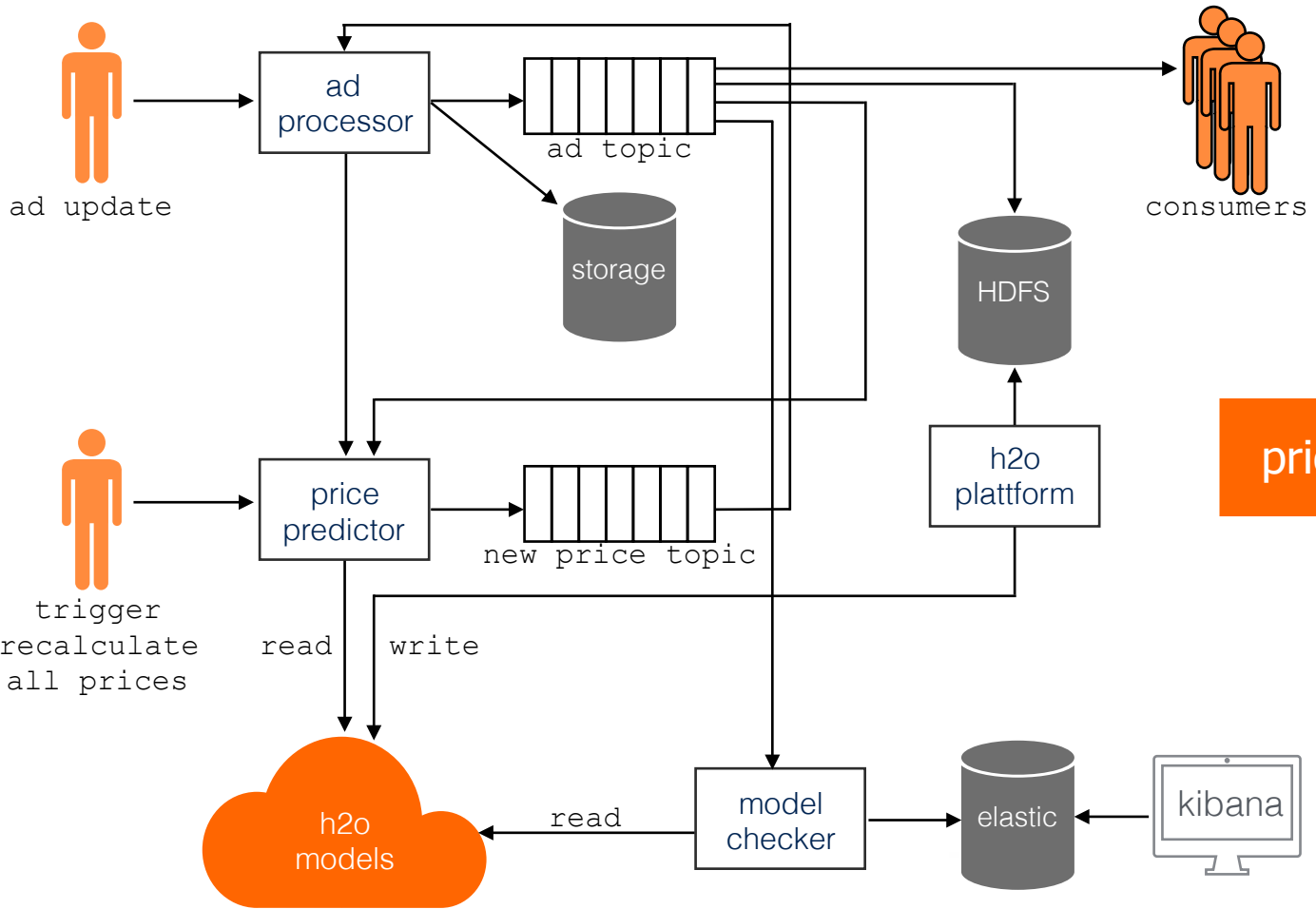
Checking a candidate

mobile.de

dry run

test with live data

consistency check



price prediction example

Conclusion

serving a model is just a part of the pipeline

data quality matters

Used photos are released under **Attribution 2.0 Generic (CC BY 2.0)** <https://creativecommons.org/licenses/by/2.0>

https://www.flickr.com/photos/minnesota_social_marketing/4518138579/in/photolist-7TfDc2-8knxwW-cbXR2E-23ELY8v-75CV57-hhp7BT-qU2EER-21gcL11-5htGyH-jEzVsj-4RDYdm-hkycCh-buw8Et-o9HJWG-nivRxo-3ZK1A-4zz8EV-VimqyW-kxgi-9mKWS2-b4ngKp-ftaB5f-8DFzum-pDFnbc-a2V1WS-mS1Twa-ai4N75-23mxYta-22Siwhs-H3Sy6q-oCmbqb-KhraoD-ezivT-nQ9aPP-RQfnsf-r4QKXB-RHYZ6o-oyizat-GDidRn-W3hces-YqoNx9-aD58ec-9fCjeR-8jrrfh-8BrZjy-aP9hbK-4kJF8H-4pXMCi-nCkRzb-G7dDTe

<https://www.flickr.com/photos/patzs/9592640975/in/photolist-fBEMXD-6eDUK8-efETNM-efLoU5-8bpCma-pihJcs-pzKPGH-bd7GuE-piiqCV-T7eyfL-TKGpLU-CmtQM-dQg1U5-9yiN9N-efEDv6-bS1Y8n-cuyYlW-VNGcar-bmVWJZ-TiEkn-dXttYa-73vN6Q-4XjLwM-82qxGH-gXGRL-6vdhWK-Go7W6m-efEU5a-59bmTv-4ZsLo-bS1Ydn-HMu1Bk-UqT4Jh-eeLvAy-aVgMSH-FdFbX-fH5A9P-o4Dz6A-jcFzFo-9bcc6M-c3aFed-gQab6E-qyTU2-5LeW6t-6ScB4-Dicxzo-fe3aCZ-7h8Y4H-sekzY-S9LxRu>

<https://www.flickr.com/photos/fdecomite/3872685816/in/photolist-6Udwz7-9Atm4D-yiNUq-9qaig-LTDFp-4aUHXV-nKDY9i-qTKkBX-qBtDUZ-qBjVrj-4vuRQZ-iZKep-6Jeggk-aVHJ5-cJFs3y-9jtab3-4dcSgh-78FNjs-4WVBSj-ecXXVU-dxDMgL-dLFy6m-4Cw2wD-345oHy-qRevwn-cM8xMh-qwbD22-ehYeem-22na7m6-25xqkBM-U8Gavy-3eAyDh-eUfEDx-22x2rdj-a8WGwG-U-M3DFw3-oNQmey-85cyir-7DuBzB-anzmZN-Ap4ru-dLFTxJ-pLCnWD-24nPLAQ-iRj9N1-9JcfoA-9JbyG7-393qZ6-3KyJpy-ixqYab>

<https://www.flickr.com/photos/wwward0/16205435108/in/photolist-qG26gA-Jhbty-52YwK-dKGcK-Scn5EY-4APMcU-9yBfUp-BTqiX-aqRVF-B4DmnB-4Jor87-CMGA7s-4HK52P-9t4YVf-bn6kyr-4Do9qL-8nPkWZ-5ebMTj-6mwPoo-e5qhAc-75igDp-5F1Aj-BCBmP-mN7w82-9kLded-8ribZJ-BiJCf-ahjXjd-3mx9u-tyvZD-9cAKvF-8rmwWC-627nF5-BYnq8B-nmPDeC-26o89W8-7ehcur-6HLEJ-XgXff-y426Rc-9ArBX-4AJRhu-Y9xXD9-7uukXa-3UiQHx-2jCSMy-D1GYJh-prpqE-9qch1e-6d5obH>

<https://www.flickr.com/photos/ryanready/4686650997/in/photolist-899j4x-pHGYdc-bdWS68-d96Q8q-r64qZf-94k3W8-nDLZ2B-8GF3pG-gFRnch-22feXoU-U5rYhm-bELrSU-9o4WeD-6V6Kka-jdh3vC-6FD5JN-8xSPAw-3ML83-57XsnL-ed5xWX-91nxV1-8bwhRz-bEpMFN-9cVNeY-cW4fmW-8oauX3-4QLoTD-9ijfZf-81Fivw-Fm7dYD-4m2bge-aWS7ti-q9WBY9-8VLucd-9Cv2bA-9Jmcr2-4qWTYp-6FD5VL-6EgmTu-7zieR3-99PEPE-69ahdi-rcYkw1-Q1Ji-9yP4xd-nJh9JV-7pfN6t-5g21fh-21gNwid-bMzGM4>

<https://www.flickr.com/photos/yeowatzup/5079168819/in/photolist-8JQ4Wc-e6EKVZ-U2Tgu8-5t6HAQ-aFBRBZ-w9uei-aFBRX8-aFBStT-25hYwSm-4wjoJi-byyH3R-25hYE3U-dfgR22-66uVRz-f65ZX2-6SkirC-aFBR2v-atUg3z-TMYgRb-a9tqAg-62p32U-5XH92-9kYZZF-4ksHdM-9ZUuLE-atfqm-9QfN9V-22P53Gn-9TL4Wt-22ft2XY-dWh5Vt-8Dh4Pp-8JT9xS-4gyZCu-a9ztT-TMYgk1-cRe6m-7UdnSJ-a9ztR-e6Zw3B-8xzctu-f69bVK-WKwAuG-dMG7Yj-5hQsSK-5T7zuD-dDhRQZ-4YCYhd-7uieFm-daCFCV>

<https://www.flickr.com/photos/juhoholmi/3535289559/in/photolist-6ophqg-VCLKxJ-4YVEa7-6ovkj7-8quFW9-5Wr6Hj-2t7nsn-aqc5Yx-5hFVqb-59SFXL-5hBN42-2t7nt2-2uWkoX-6oraAZ-4HvMJ6-7Cgk29-4YRkXr-ZJUY5W-84hFi5-6ovghy-4LVEyZ-4vpmZE-9Ui2A8-4YVA6o-24YZ5sz-SBhXiq-HK2Fj-Gf1rPX-HS79KY-Xjo1Hs-QiUVuw-9qVcwG-9qVcd3-9Uhy3X-4YRkbn-bAuENF-TooZgf-pwEmR8-audNkd-9qSdNB-9UkMxs-aAgD1m-augE4L-pPeh93-pwEmZz-RB5cgr-9Ui17D-9wfb9Y-qQ865E-9UkKMj>

