

ARCHITECTURES OF REAL- WORLD MACHINE LEARNING SYSTEMS IN THE CLOUD

Lukas Grolig

STRUCTURE OF THIS TALK

- WE WILL COVER SOME BASICS
- ON CASE STUDY WE WILL SHOW PROCESS OF DEVELOPING ML SYSTEM
- WE WILL DISCUSS SOME PROCESSES AND GOTCHAS

HYPE AROUND MACHINE LEARNING

- Every big company wants to apply ML
- No one wants to stay behind
- They don't know where to start
- Some have data / some will start collecting them

WHAT THEY CAN SOLVE?

- Predict sales - units / profits
- Motivate users to buy something more
- Segment clients
- Find patterns in user workflow and move him to next step
- Automatic form handling
- Chatbots
- ...



So you want to do
machine learning?

FIRST WHAT ARE NECESSARY SKILLS

- You must know technologies used in their systems (Java, .Net)
- DB knowledge is a must (MS SQL, Oracle, Postgres, Mongo, Exadata, Greenplum, Vertica)
- Python & Javascript
- Infrastructure (Kubernetes plus any cloud of your choice)



ORIFLAME
— SWEDEN —

The advertisement features a woman's face on the left side. The background is a soft-focus field of purple flowers. The central focus is a collection of Oriflame beauty products, including two tubes of lipstick, a bottle of 'POSSNESS' perfume, two round compact cases filled with golden beads, two purple eye pencils, a red lipstick, a purple mascara tube, a pink mascara wand, a small jar of cream, a pink lip balm, and a gold-colored compact case. The Oriflame logo is visible on several products.

HOW WE STARTED

- Oriflame wanted do machine learning
- There were no use cases and we were suggesting what we shall look
- As a first step we wanted to predict sales



REQUIREMENTS

- Especially in ML topics you client has only rough idea what he wants when he approaches you. **You don't receive any requirements.**
- **Communication is key.** Discuss what options there are. Where you can start and describe all steps. This way you can get rough basic plan.
- Always set right prices. In begin get flexible budget to do exploration or offer exploration as service.
- When at least part of system is clear formalise requirements and agree on budget on that part.
- **Never promise high accuracy!**

STORY CONTINUES

- We were looking for contact person who has some data for our case.
- Data were available so we started building models.
- Also we extended scope to do **automatic catalogue planning.**



COMMUNICATION IN FIRST STAGE

- This is most critical part. You have to build confidence of your client in you.
- Get direct contacts. It is better to get information that by using mediator.

PROTOTYPE FAST

- **Mock** as many things as possible. Upload CSVs to blob storage. Make HTTP calls to storage.
- Use Python to fast prototype things.
- Before you get into ML explore the data first. (you probably heard about histogram, median, standard deviation). Use some automatic visualisations - heatmap using SOM, do clustering. **You must understand data you have.**
- **Sell results to your client.** Even those visualisations can have high value. Always perfectly describe business case that you can do based on what you have.

START WITH FIRST MODELS

- **In the begin never develop custom ML model.**
- **Often linear regression is enough.** If you don't think it will be enough use **automl**.
- Important note: double check that you don't measure accuracy on training data. **High percentages** in this stage **are suspicious.**

SIDE NOTE

- **In most business cases you don't do image recognition.**
- Regression does not have so high accuracies.

SO WE STARTED BUILDING SYSTEM

- We had service that forecasted sales of catalogue.
- We started catalogue planning based on genetic algorithm.
- Note: there are algorithms like differential evolution, particle swarm optimisation but all are mainly tested on contiguous space. Often your space is discrete and values are not ordinal.

TRY DIFFERENT APPROACHES

- Forecasting
 - we tried single model for everything. Different model per country. Multiple models for country.
- Catalogue Planning
 - In genetic algorithms good approach is use coevolution algorithms. Works good for many dimensions. Some implementations like C3 work on paper (developed on continuous space) but reality is harder.



JUPYTER NOTEBOOK IS YOUR
BEST FRIEND

APPROACH TO ML FINETUNING

- Different algorithms have differences in accuracy in percentages. Rather focus on getting more data columns, model relations, or manipulate data - often applying logarithmic or exponential function will give you much more than trying different algorithms.

APPROACH TO ML FINETUNING

- In optimisation problems research papers are only starting point. You have to work with discrete space. You are looking for solutions how to make ordinal values from categoric attributes.

STORY CONTINUES

- We started replacing mocked parts with connections to real systems. This sometimes involves creating **tunnels between cloud and on-premise systems.**
- When forecasting catalogue payload for simplified version of 1.5mb. Population of 100 chromosomes takes time to forecast.



PYTHON IS SINGLE THREADED

- **Python becomes** in later stage **pain point**. You have to start doing multithreading.
- Option 1: parallel map function spamming processes.
- Option 2: use celery plus redis
- Option 3: migrate to python 3.7 and use async/await
- Option 4: you can spam many kubernetes pods.
- Does not work: lambda functions - startup is too slow, calculation slow, responses high. We easily took down Azure Functions, AWS works so so.

YOU HAVE TO START SCALING CLUSTERS AND OPTIMISE

- In Azure cluster of DS2_V2 is pricy. With 10 machines that handled running system in reasonable time (getting catalogue in 10s minutes) costs over 3000 euro per month.
- Forecasting training times were in tens of hours. Algorithms like LSTM don't work well on GPU. Use GRU or TNN or XGBOOST.
- Evaluation of bigger ML model takes 1s or more. Count with that.

START BUILDING REAL SYSTEM

- You know requirements.
- Define technologies - Python is not best choice in this stage.
- Connect to live systems = replace mocks. Your mocks probably have some rough contract already. Time to refine it.
- Learn how to do VPN tunnels. Never forget to cache or store data on your side (pool them once per day).

RESPONSES TAKE A LOT OF TIME

- Azure or AWS terminates your connection when you don't send a response within a given time.
- Build asynchronous APIs. Start task, return HTTP 201 Created with location of result. Pool the location.
- Or use real-time communication like websockets. This will lead to a custom protocol that is harder to use.

OPTIMIZE TRANSFERS

- It is easy to do 100s of MBs spikes in traffic.
- Start **gzipping requests** (not only responses!)
- Move to binary protocol if possible. Websockets or grpc handle those situations better than HTTP.

BUILD ML-OPS

- In machine learning you must also create pipelines.
- When new data arrives, start training. Evaluate new model and historical models.
- Deploy appropriate model to production system.

CREATE FRONTEND

- We suggest to start building Excel plugin first. It is written in react and can be easily converted to standalone site or widget.
- Excel will give users many function that you would have to otherwise develop in application. They will provide feedback faster.
- Notes:
 - You have to know Javascript and React
 - Debugging Excel add-ins (especially on Mac) is brutal pain.

HOW IT CONTINUED

- Because number of constraints on catalogue planning grew Oriflame wanted automated way.
- Getting patterns from data was hard. It was easier to do anomaly detection so we integrated that to fitness function.
- We started taking side projects like OCR of customer IDs or passports, sentiment analysis of messages on FB or support chat...



DON'T FORGET PRESENT RESULTS

- Present results and progress often.
- Always have business case in mind and sell it again and again! Customer will come with additional projects.
- Even 1 or 2 percent improvement will make you client happy.
- In optimisation number of constraints will grow with each presentation.
- Present results across whole company. There will be other people who will have business case for ML.

WELL, THAT'S IT FROM MY
SIDE.

ANY QUESTIONS?

THANK YOU AND HAVE A
NICE DAY