ML in MT

Markus Lippus
What we do?

Machine learning and AI development agency

We help companies make sense of their business data and turn complex data into insights and profit.

www.mindtitan.com
Why are we doing it?

We believe that data science isn’t just a nice-to have innovation.

It will be a must have for any company looking to stay competitive in the next 10 years.
Work process

**STEP 1:** Identifying AI use case

**STEP 2:** Data exploration and proof of concept

**STEP 3:** Proposal for deliverables

**STEP 4:** Model development

**STEP 5:** Application deployment

[www.mindtitan.com](http://www.mindtitan.com)
MindTitan’s Services

- Self-learning systems
- Image recognition
- Recommender engines
- Text analysis
- Process optimization
- Time-series predictions

www.mindtitan.com
The idea

Automated age verification

Problem: Underaged people are not asked ID verification for illegal activities
Use case:

Technical solution:

Video stream → Face localization (NeuralNet) → Age Verification (DeepLearning) → Identity Check (DeepLearning)

>30? Y N

ID? Y N

SELL BEER

www.mindtitan.com
Why we continued w/ MindTitan?

+ Great team
+ Vision to build a growing ML community
+ Finding interested clients from all over the world
+ Interesting problems we didn’t even know existed
+ Work that inspires and grows on a daily basis
Machine learning in an agency

- Differs from a product company - many different problems
- Differs from academia - ROI oriented
STEP 1: Identifying AI use case
<table>
<thead>
<tr>
<th>Decisions</th>
<th>ML task</th>
<th>Value Propositions</th>
<th>Data Sources</th>
<th>Collecting Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>How are predictions used to</td>
<td>Input,</td>
<td>What are we trying</td>
<td>Which raw</td>
<td>How do we get</td>
</tr>
<tr>
<td>make decisions that provide</td>
<td>output</td>
<td>to do for the</td>
<td>data sources</td>
<td>new data to</td>
</tr>
<tr>
<td>the proposed value to the end-user?</td>
<td>to predict,</td>
<td>end-user(s) of the</td>
<td>can we use</td>
<td>learn from</td>
</tr>
<tr>
<td></td>
<td>type of</td>
<td>predictive system?</td>
<td>(internal and</td>
<td>(Inputs and</td>
</tr>
<tr>
<td></td>
<td>problem.</td>
<td>What objectives are</td>
<td>external)?</td>
<td>outputs)?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>we serving?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Making Predictions</th>
<th>Offline Evaluation</th>
<th>Features</th>
<th>Building Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>When do we make predictions on new inputs? How long do we have to featureize a new input and make a prediction?</td>
<td>Methods and metrics to evaluate the system before deployment.</td>
<td>Input representations extracted from raw data sources.</td>
<td>When do we create/update models with new training data? How long do we have to featureize training inputs and create a model?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Live Evaluation and Monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods and metrics to evaluate the system after deployment, and to quantify value creation.</td>
</tr>
</tbody>
</table>
STEP 2: Data exploration and proof of concept
Data is siloed

Siloed Data is Sad Data
No analytics stack
Everyone wants the cool stuff
Data collection pipelines

1. Change
2. Don’t exist
Elisa chat labelling

- Simple solution – manual labelling
- Scalable solution – clustering engine
1. UNLabeled Chats
2. Chat Clusters
3. New Model
4. Solution Offered as Suggestion
5. Bot Can Help?
   - Yes
   - No
6. NO
7. YES
Labels can be a mess
Labelling issues with network analysis

- Multiple phone numbers per account
- No idea when the issue occurred
- No idea where the issue occurred
Constrained systems

- Low powered systems - e.g. embedded systems
- No GPU available
Sample task

- Low power CPU
- 1GB RAM
- Can’t install anything on the machine
- 1080p images – webcam
- Binary classifier – is there something on the image or clear BG?
- 0.1% FPR 0.5% FNR
Data doesn’t hold information
Data doesn’t hold information
Data doesn’t hold information
Tools for managing DS workflows