The Integration of ML

BY THIVYESH AHILATHASAN
Agenda

AI for SIC at Statistics Norway

Working solution
AI for SIC at Statistics Norway
Business
Description from the enterprise

Coding the enterprise

Production of statistics etc.

Brønnøysundregistrene
Statistics Norway
NAV
Skatteetaten

Brønnøysundregistrene
Working solution
Pipeline

1. Prepare Data
2. Train Model
3. Package Model
4. Validate Model
5. Deploy Model
6. Monitor Model
What do we need?

• Providing an API
• Deploy to the cloud
• Traffic splitting and Canary release
• Log and filter end-user data
• Analyze logging data
• Not specific for ML
Google Cloud platform

- App engine
- Stackdriver
- Big query
- Data studio
Architecture sketch
App Engine

• Easy to use
• Application platform
• Infrastructure
• Deploy
• Traffic splitting
• Automated balancing
Traffic splitting and Canary release
Stackdrive for diagnostics

- Monitoring
- Logging
- Debugging
- Tracing

Google Stackdriver
BigQuery and Datastudio

- Big data
- Visualization with Datastudio
- Datastudio: Near Realtime in BIGQuery
- End-user to datastudio - couple of seconds
Datastudio

Using the confidence of the best prediction per request. 1 means the model gives a 100% probability of this prediction being correct, while 0 is a 0% probability.

### Average confidence

| Overall average confidence | 0.45 |

### Average confidence per model over time

<table>
<thead>
<tr>
<th>Model</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>fasttext-2020-04-02</td>
<td>0.72</td>
</tr>
<tr>
<td>fasttext-2020-03-06</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Source: Torgeir Mossige Grønning, Computas AS
Datastudio

API Predictions

Counting the number of prediction requests handled by the API, which is an indicator of the overall traffic the service receives.

Note that there will naturally occur predictions where the client (user) does not report how the prediction was used, so it is expected that the number of predictions produced is higher than the number of reported prediction usages.

<table>
<thead>
<tr>
<th>Predictions</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>143</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>fasttext-2020-03-06</td>
<td>43</td>
</tr>
<tr>
<td>fasttext-2020-04-02</td>
<td>2</td>
</tr>
<tr>
<td>1-2/2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NACE code</th>
<th>Times predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>82980</td>
<td>34</td>
</tr>
<tr>
<td>94991</td>
<td>21</td>
</tr>
<tr>
<td>73110</td>
<td>15</td>
</tr>
<tr>
<td>97000</td>
<td>9</td>
</tr>
<tr>
<td>1+22/22</td>
<td></td>
</tr>
</tbody>
</table>

Source: Torgeir Mossige Ønning, Computas AS
Ending notes

• Simple architecture
  ◦ More complicated architecture?

• Not a ML specific workflow
  ◦ Best practices

• Not a Model specific workflow

• Git repo: https://github.com/datautvikling/nace-predictor
Thank you!