TALL SOM FORTELLER

The Integration of ML

BY THIVYESH AHILATHASAN

Statistisk sentralbyrå
Statistics Norway

Agenda

AI for SIC at Statistics Norway

Working solution

TALL
SOM FORTELLER

AI for SIC at Statistics Norway

Statistisk sentralbyrå
Statistics Norway

Brønnøysundregistrene

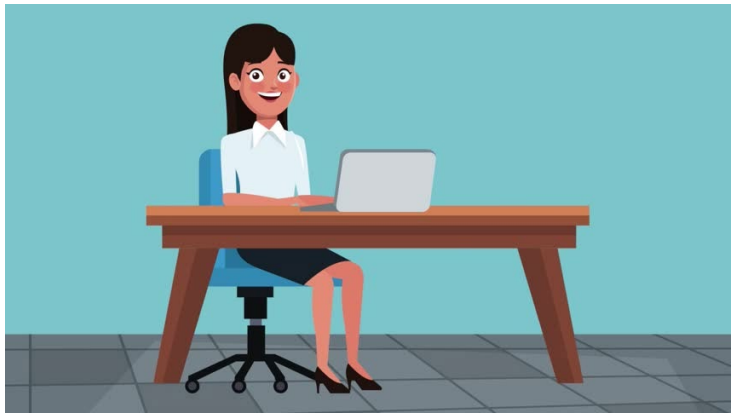
Statistisk sentralbyrå
Statistics Norway

nav

Skatteetaten

Business
Description from the enterprise

Brønnøysundregistrene



Coding the enterprise

Statistisk sentralbyrå
Statistics Norway

nav

Skatteetaten

Production of statistics etc.

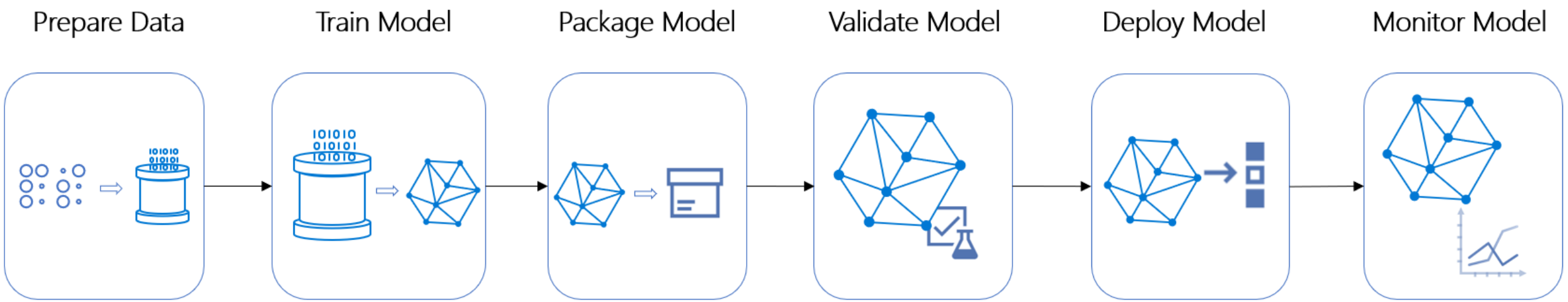
Statistisk sentralbyrå
Statistics Norway

TALL SOM FORTELLER

Working solution

Statistisk sentralbyrå
Statistics Norway

Pipeline

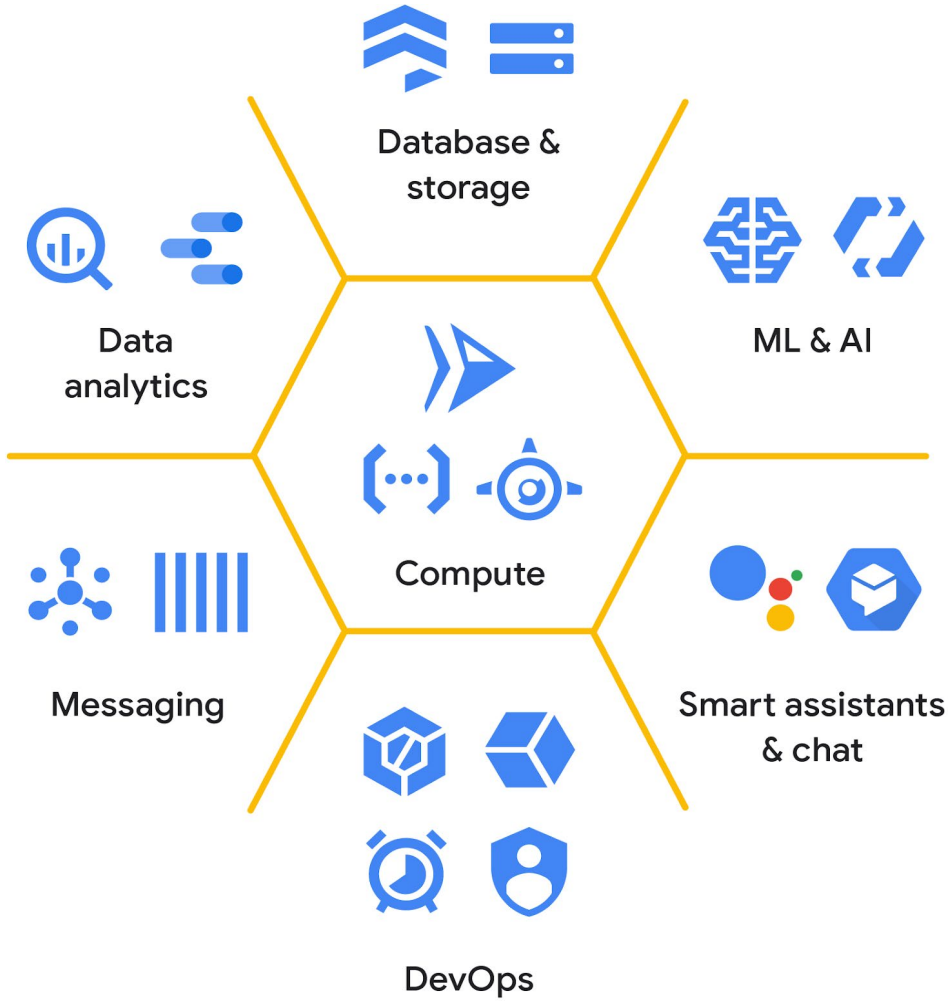

Prepare Data
Train Model
Package Model
Validate Model
Deploy Model
Monitor Model

What do we need?

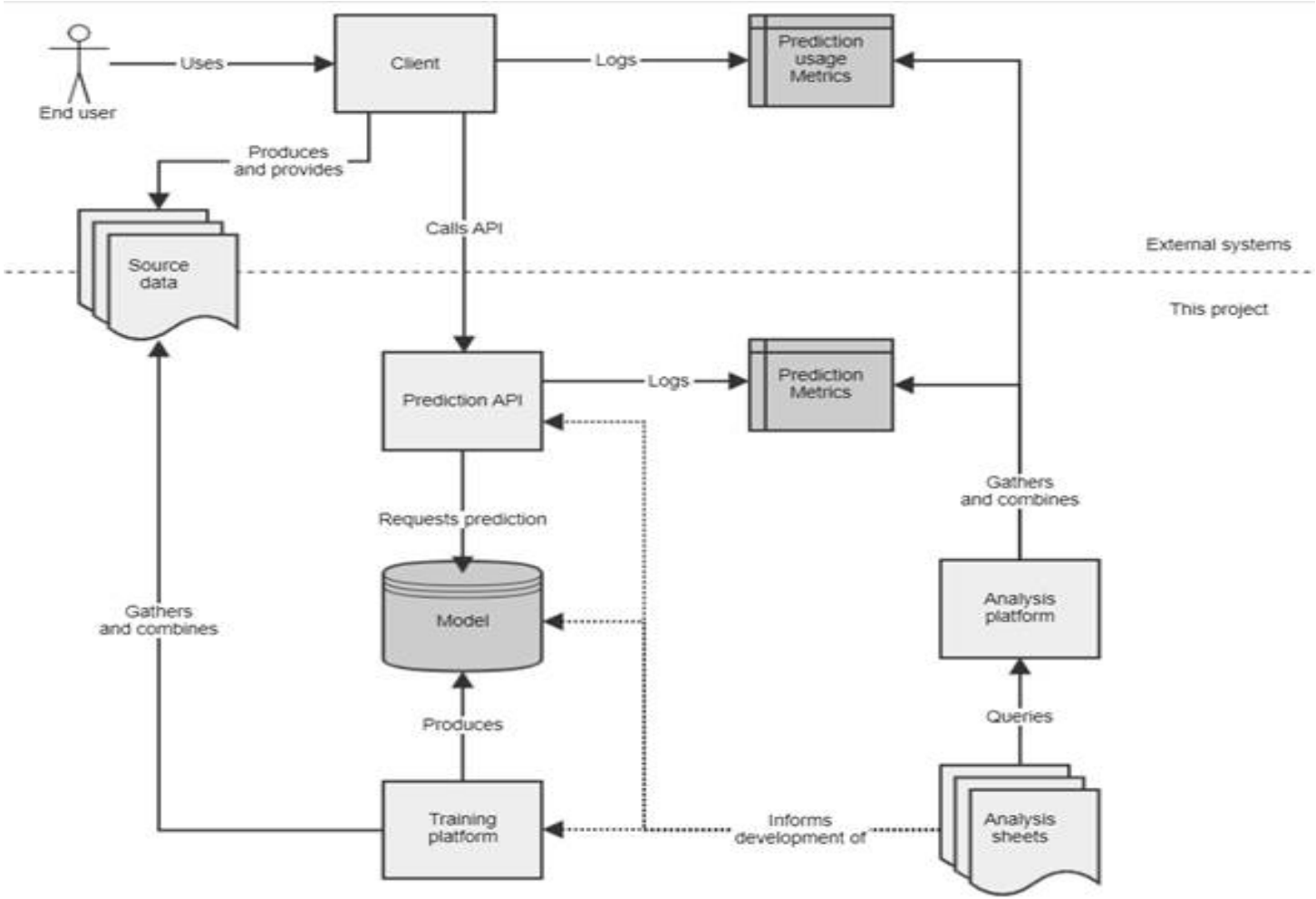
- Providing an API
- Deploy to the cloud
- Traffic splitting and Canary release
- Log and filter end -user data
- Analyze logging data
- Not specific for ML

Google Cloud platform

- App engine
- Stackdriver
- Big query
- Data studio



Architecture sketch

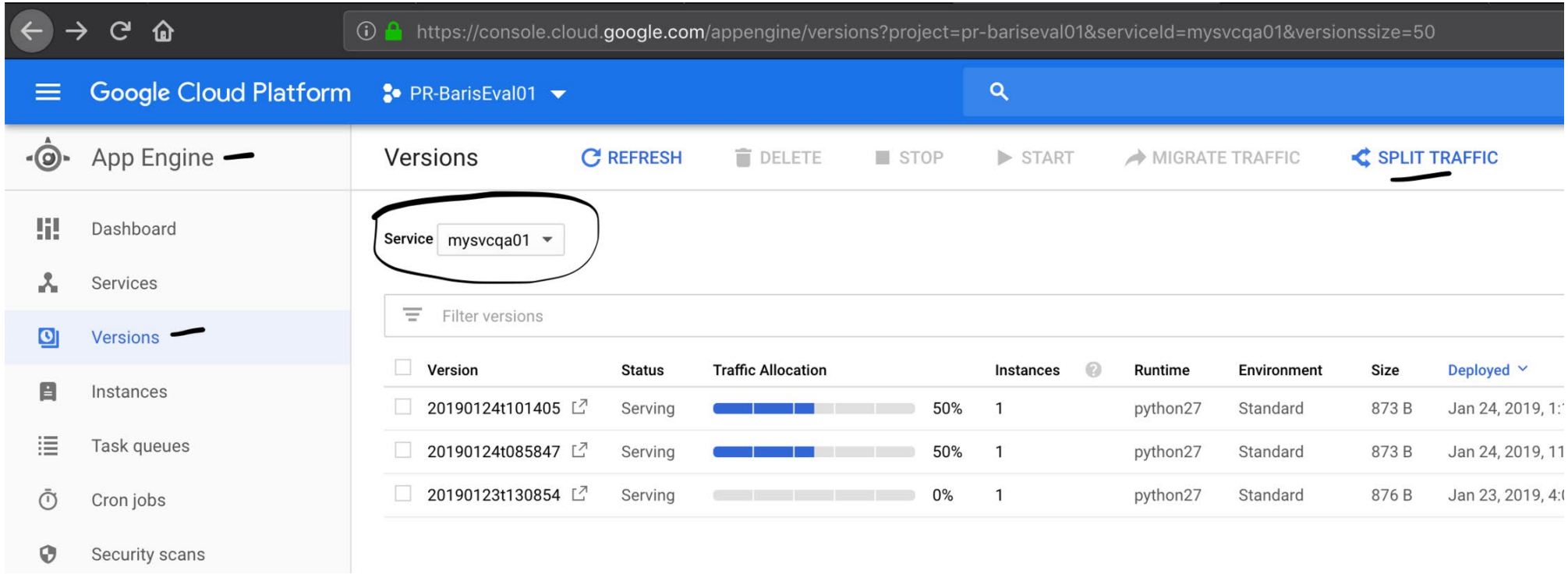


Statistisk sentralbyrå
Statistics Norway

App Engine

- Easy to use
- Application platform
- Infrastructure ?
- Deploy
- Traffic splitting
- Automated balancing

Traffic splitting and Canary release



Stackdrive for dignostics

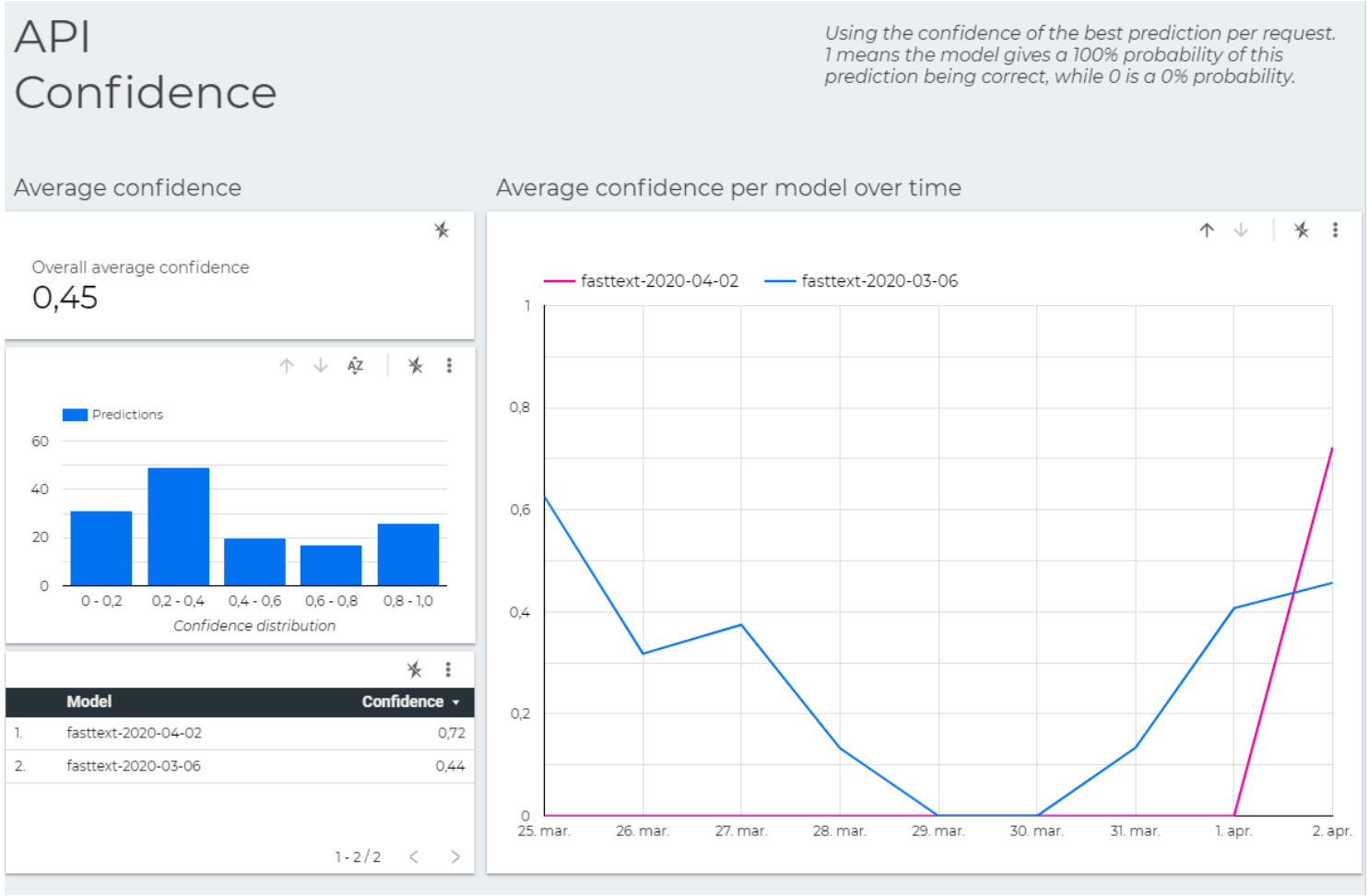
- Monitoring
- Logging
- Debugging
- Tracing

Google Stackdriver

BigQuery and Datastudio

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

Datastudio



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.

Predictions

Predictions	Models
143	2

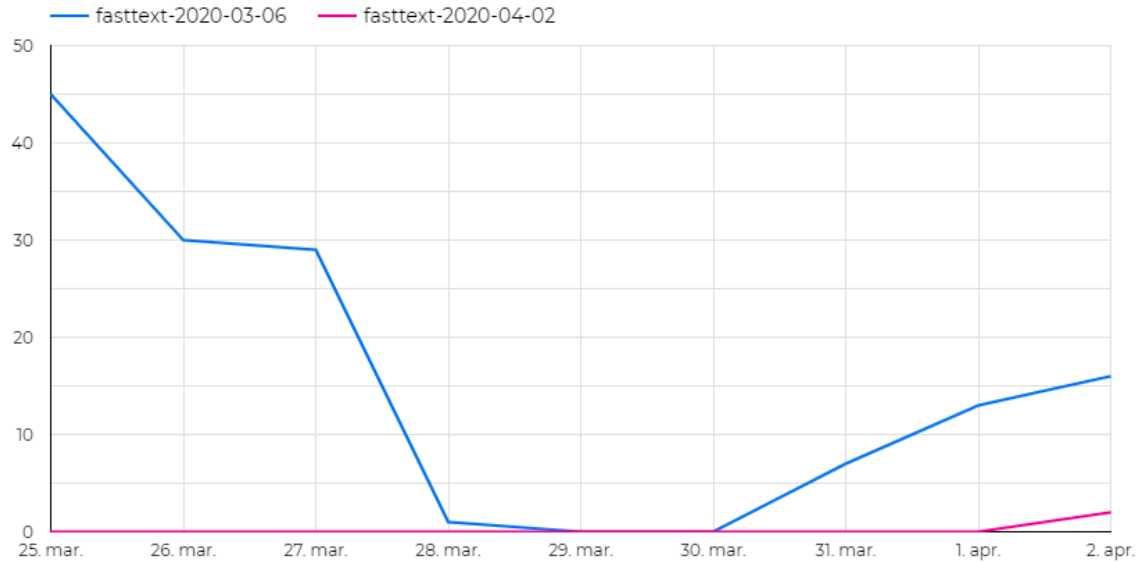
	Model	Predictions
1.	fasttext-2020-03-06	141
2.	fasttext-2020-04-02	2

1 - 2 / 2

	NACE code	Times predicted
1.	82.990	34
2.	94.991	21
3.	73.110	15
4.	97.000	9

1 - 22 / 22

Predictions per model over time



Source: Torgeir Mossige Grønning, Computas AS

Ending notes

- Simple architecture
  - More complicated architecture ?
- Not a ML specific workflow
  - Best practises
- Not a Model specific workflow
- Git repo : https://github.com/datautvikling/nace -predictor

TALL SOM FORTELLER

Thank you!

Statistisk sentralbyrå
Statistics Norway