NOTAMS
Machine Learning Analysis

S/IBC
13.08.2018

NOTAMs
Machine Learning Analysis

- NOTAM & NOTAM Workflow
  - What is a NOTAM?
  - Workflow
- Machine Learning in NOTAM Workflow
  - Data available
  - Hackathon POC
  - Approaches
    - “3 Piles”
    - Unsupervised Clustering
    - Experience categories (Japanese flight control, etc)

NOTAM & NOTAM Workflow

What is a NOTAM?

NOTAMs & NOTAM Workflow
What is a NOTAM

- NOTAM = Notice To AirMen
- Short text
- Structured + Unstructured data

B0821/18 NOTAMN
Q) LSAS/QOLAS/V/M/E/000/035/4652N00837E005
A) LSAS B) 1807161442 C) 1807232359 EST
E) OBST LGT 2.5KM W ALTDORF (ANTENNA ATTINGHAUSEN) 465157N0083644E
U/S. 75.0M / 246FT AGL, 1061M / 3481FT AMSL.
CREATED: 16 Jul 2018 14:42:00
SOURCE: LSSNYNYX

NOTAMs & NOTAM Workflow
What is a NOTAM

- NOTAM = Notice To AirMen
- Short text
- Structured + Unstructured data

B0821/18 NOTAMN
Q) LSAS/QOLAS/V/M/E/000/035/4652N00837E005
A) LSAS B) 1807161442 C) 1807232359 EST
E) OBST LGT 2.5KM W ALTDORF (ANTENNA ATTINGHAUSEN) 465157N0083644E
U/S. 75.0M / 246FT AGL, 1061M / 3481FT AMSL.
CREATED: 16 Jul 2018 14:42:00
SOURCE: LSSNYNYX

- Tags like Q, A, B, C, D, E, F, G, indicate type of data that is described

NOTAM & NOTAM Workflow

Workflow

NOTAMs & NOTAM Workflow

Workflow

NOTAM produced -> Sent to Aeronautical Authority -> Distributed
NOTAM provider -> Airline -> NOTAM officer -> Flight dispatcher

NOTAM is produced to indicate any abnormal condition that may affect a flight's operation
e.g. Obstacles, closed runways, prohibited airspace, fireworks…
On receiving the NOTAMs from the provider, hard filters are set in place that suppress those NOTAMs that are with a high degree of certainty not of interest
e.g. Notices from Flight Information Regions that are not flown over by the company
The NOTAM officer goes through the remaining notices having to suppress those that do not interest the company
The flight dispatcher goes through the NOTAMs (supressed and not) to prepare the flight briefing

Machine Learning in NOTAM Workflow

Data Available

Machine Learning in NOTAM Workflow
Data Available

- NOTAMs are provided in freetext and stored in an internal database

- Suppressions, by hard-coded filter and by human analyst, are stored in the database and are discriminated by suppressor

Machine Learning in NOTAM Workflow

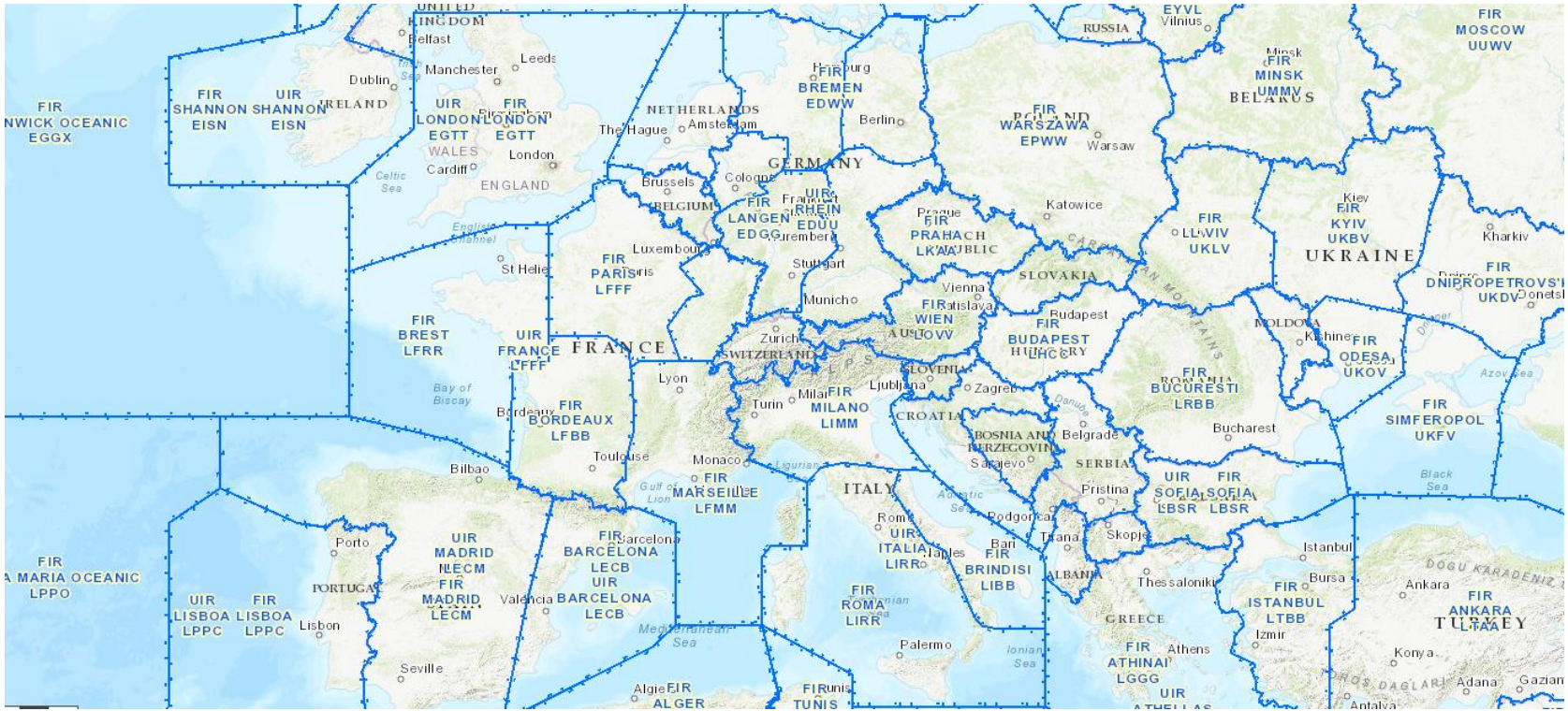
Hackathon POC

Machine Learning in NOTAM Workflow
Hackathon POC

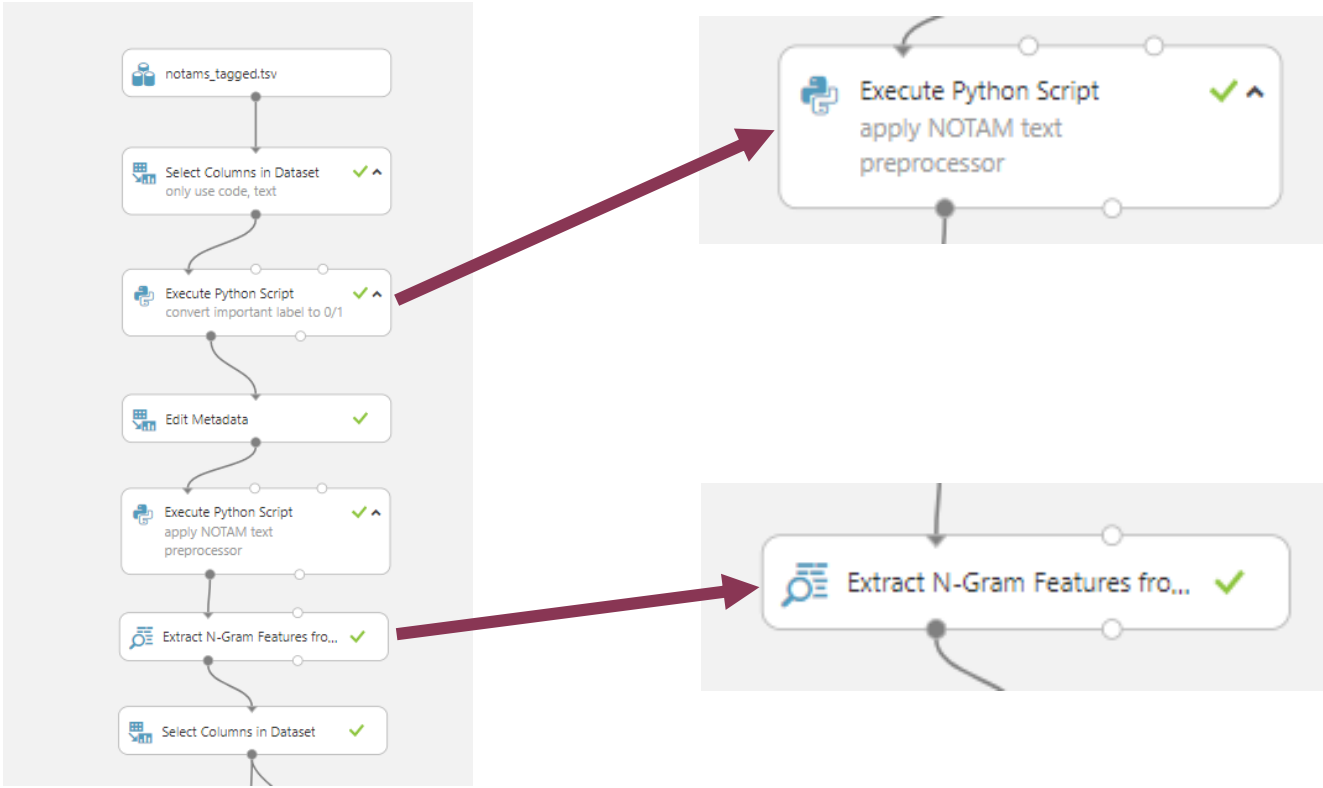
- Hackathon used 10.000 labelled NOTAMs in order to evaluate the performance of machine learning at reliably assisting the human analyst
- Data used was mainly the FIR (Flight Information Region) and the free-text

B0821/18 NOTAMN
Q) LSAS/QOLAS/V/M/E/000/035/4652N00837E005
A) LSAS B) 1807161442 C) 1807232359 EST
E) OBST LGT 2.5KM W ALTDORF (ANTENNA ATTINGHAUSEN) 465157N0083644E
U/S. 75.0M / 246FT AGL, 1061M / 3481FT AMSL.
CREATED: 16 Jul 2018 14:42:00
SOURCE: LSSNYNYX

Machine Learning in NOTAM Workflow
Hackathon POC



Machine Learning in NOTAM Workflow
Hackathon POC



Machine Learning in NOTAM Workflow
Hackathon POC

- Free-text was parsed with consideration to contractions and n-gram analysis was performed to select features on the individual NOTAMs

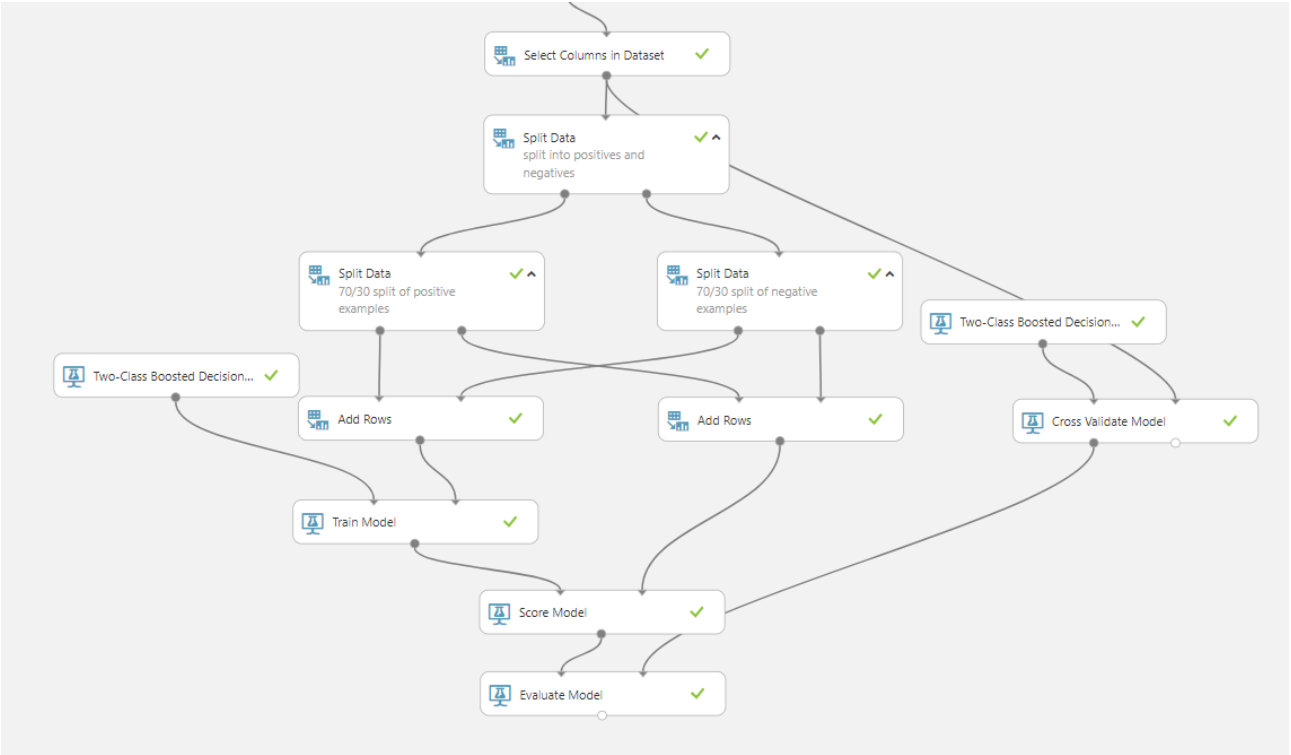
```
SubstitutionRegex(r'\bSN\b', 'SNOW'),
SubstitutionRegex(r'\bTAX\b', 'TAXI'),
SubstitutionRegex(r'\bTFC\b', 'TRAFFIC'),
SubstitutionRegex(r'\bTHR\b', 'THRESHOLD'),
SubstitutionRegex(r'\bTWR\b', 'TOWER'),
SubstitutionRegex(r'\bTWY\b', 'TAXIWAY'),
SubstitutionRegex(r'\bU/S\b', 'UNSERVICEABLE'),
SubstitutionRegex(r'\bVORTAC\b', 'VOR TACAN'),
SubstitutionRegex(r'\bWI\b', 'WITHIN'),

# Assign more meaning to numbers
# Note '<' & '>' inverts the logic of \b to \B to detect
# "non-alphanumeric" characters at the pattern boundary
SubstitutionRegex(r'\B<number[|]\d{2}> ?(K|M)HZ\b', '<freque
SubstitutionRegex(r'\bCH<number[|]02>[XY]\b', '<channel>'),
SubstitutionRegex(r'\bFL<number[|]03>\B', '<flightlevel>'),
SubstitutionRegex(r'\B<number[|]\d{2}> ?FT( ?(AMSL|AGL))?\b'
SubstitutionRegex(r'\B<number[|]\d{2}>[NS] ?<number[|]\d{2}>
SubstitutionRegex(r'\B<number[|]\d{2}> ?(M|KM|NM)\b', '<dist
SubstitutionRegex(r'\bRUNWAY ?<number[|]02>(R|L|C)?(/<number
```

NGramsString
["AND","<number>"]
["AREA","ACTIVE","<distance>","<flightlevel>"]
["AREA","AND","<height>","<flightlevel>"]
["ACTIVE","DUE","SURFACE","<flightlevel>"]
["AERODROME","CLOSED"]
["<runway>","CLOSED","DUE"]
["AND","<runway>","<distance>"]
["UNSERVICEABLE"]
["<runway>","UNSERVICEABLE"]
["<number>"]
["<runway>","CLOSED","DUE"]
["<runway>","CLOSED","DUE"]
["<number>","<flightlevel>","<number>_<number>"]
["AREA","GROUND","<height>"]
["<number>","<height>","<flightlevel>","<number>_<number>"]
["<runway>","CLOSED","DUE"]
["<runway>","UNSERVICEABLE"]

SWISS

Machine Learning in NOTAM Workflow
Hackathon POC

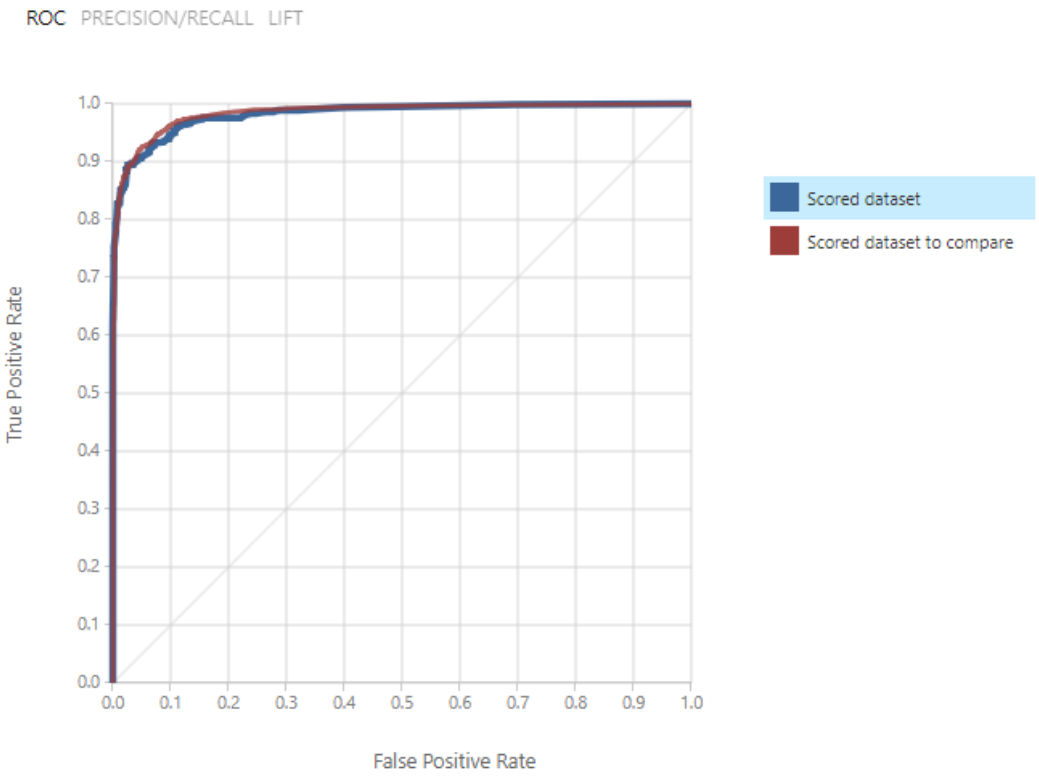


Machine Le
Hackathon

- 7000 of the NOTAM
- 3000 of the NOTAM

- The model divides ... ıportant" (1, 0)

- A Two-Class Boost



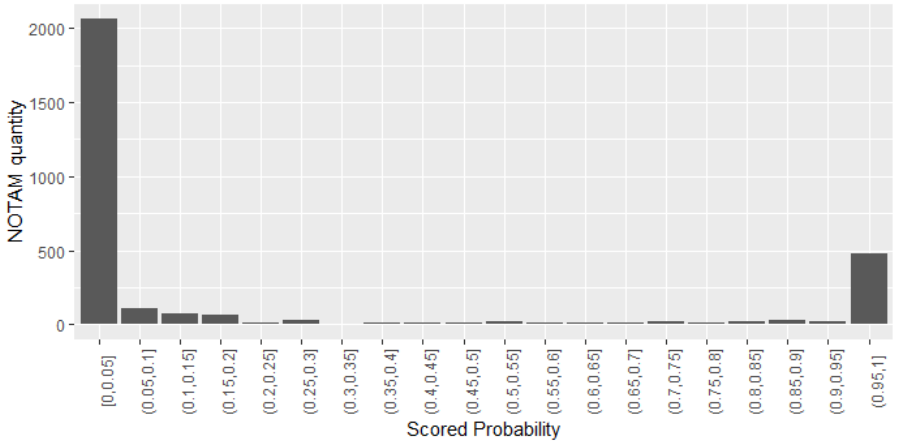
True Positive	False Negative	Accuracy	Precision	Threshold	AUC
566	84	0.954	0.914	0.5	0.982

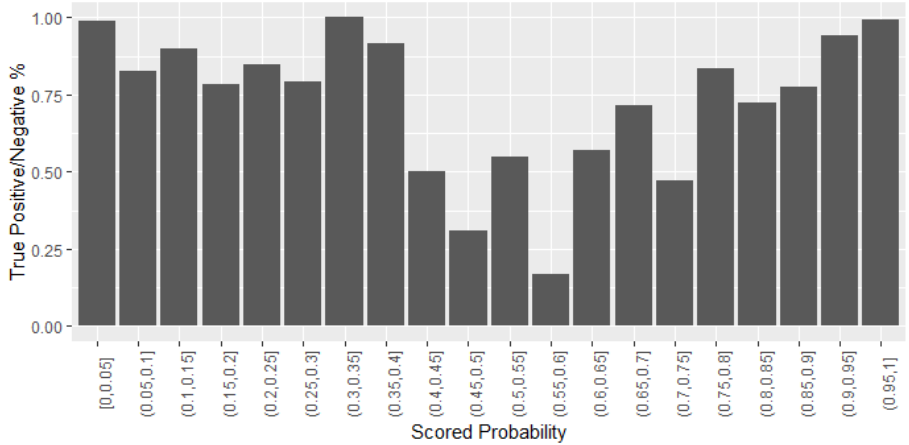
False Positive	True Negative	Recall	F1 Score
53	2297	0.871	0.892

Positive Label	Negative Label
1	0

Machine Learning in NOTAM Workflow
Hackathon POC

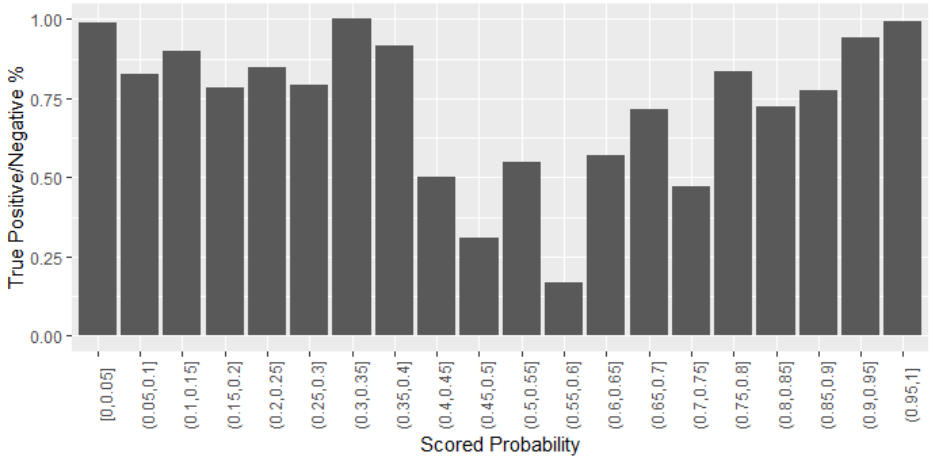
- Analysis of the hackathon model's results, more specifically, the true positives and their distribution yields that looking at the extremes of the Scored Probabilities, instead of Scored Labels, gives higher accuracy and a reasonable amount of high certainty classifications.

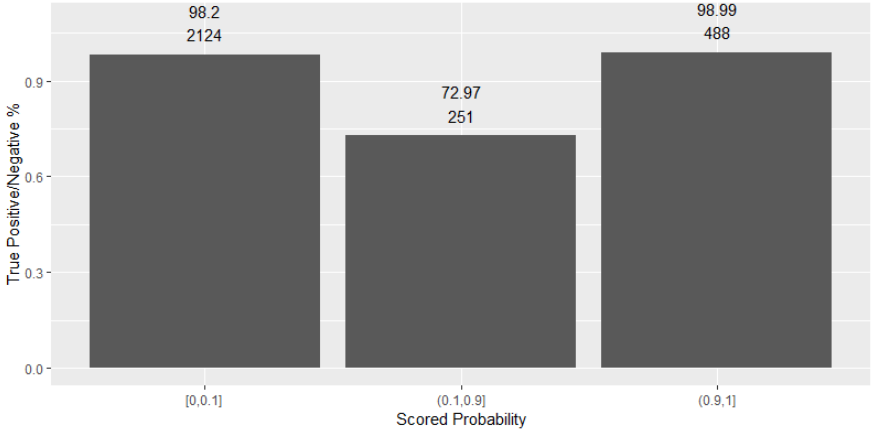




Machine Learning in NOTAM Workflow
Hackathon POC

True Positive/Negative % per Scored Probability bin





	bin	quantity	frequency
	<fct>	<int>	<S3: table>
1	[0,0.1]	2124	0.9819695
2	(0.9,1]	488	0.9898580

	bin	quantity	frequency
	<fct>	<int>	<S3: table>
1	[0,0.05]	2038	0.9898009
2	(0.95,1]	472	0.9915966

	bin	quantity	frequency
	<fct>	<int>	<S3: table>
1	[0,0.025]	1984	0.992
2	(0.975,1]	400	1.000

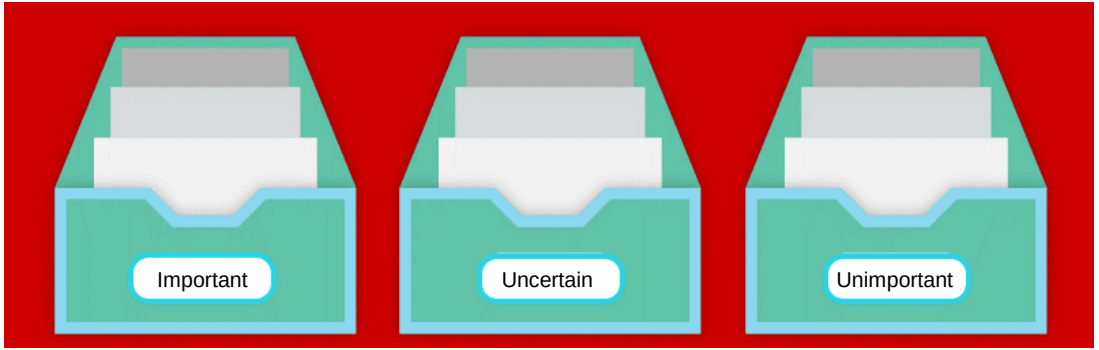
SWISS

Machine Learning in NOTAM Workflow

Approaches

Machine Learning in NOTAM Workflow
Approaches

“3 Piles”: Construct a model that divides the data between parts where the model is proficient at classifying and data where the model is not as reliable and then present the NOTAM officer with three categories:



1. **Important** NOTAMs the model with high confidence level (some errors tolerable)

3. **Uncertain** NOTAMs, where the model does not provide a confident fit into either – those should be revised by a NOTAM’s Officer or Analyst

2. **Unimportant** NOTAMs the model is confident that should be suppressed

From this analysis useful results could come up such as:

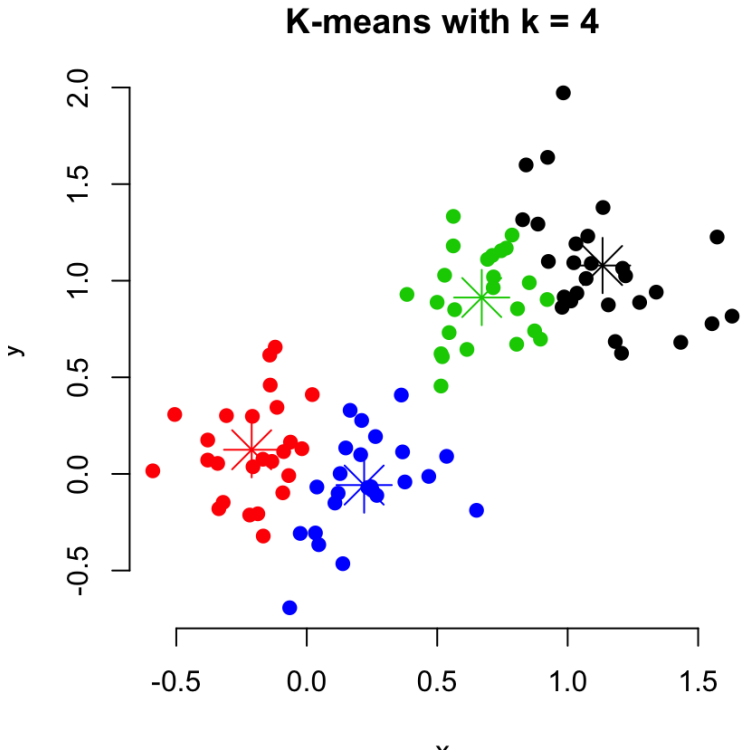
- A model which could hypothetically be inserted upstream of the NOTAM officer in the workflow
- A clear characteristics pattern for important NOTAMs which could be added to the hard-coded filters

SWISS

Machine Learning in NOTAM Workflow
Approaches

Unsupervised Clustering

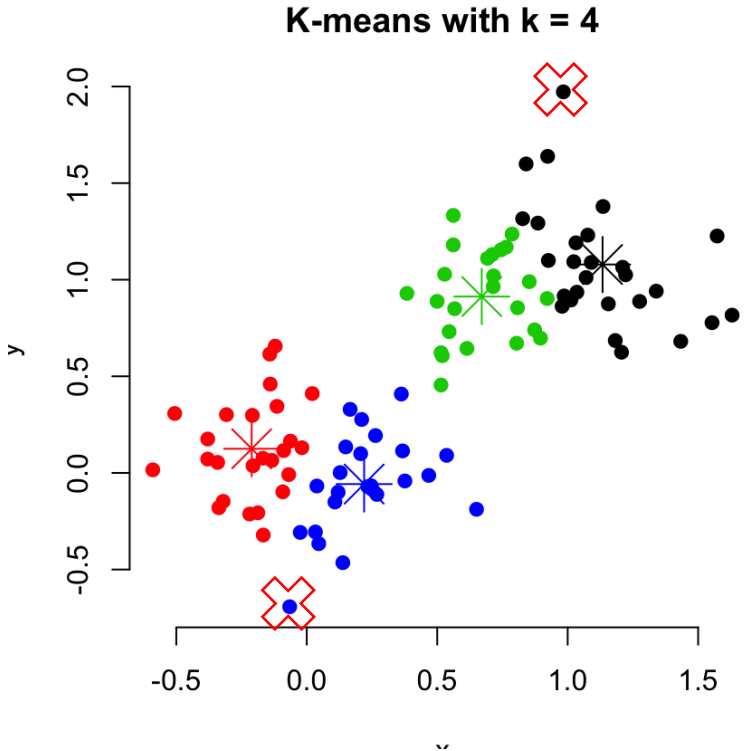
- Detect clusters of similar properties
- Verify if those could be understood in the sense of message characteristics
- Establish statistics of classifier outcomes for each cluster – are there clusters with particularly high/low classifier certainty?
- It would be interesting to see what kind of patterns show up both in the data that has gone through the hard-coded filters and that which hasn't



Machine Learning in NOTAM Workflow
Approaches

Clustering-Based Anomaly Detection

- Assumption: Data points that are similar tend to belong to similar groups or clusters, as determined by their distance from local centroids.
- K-means is a widely used clustering algorithm. It creates 'k' similar clusters of data points. Data instances that fall outside of these groups could potentially be marked as anomalies.
- It could be checked, if messages identified as anomalies end up more frequently as mismatch in the classification
- Would that be an alternative way to populate “Uncertain” bin – the one that should be checked by an expert?



Machine Learning in NOTAM Workflow
Approaches

- Geographical Interpretation
  - In obstacle NOTAMs, the closeness of the area of influence to a SWISS operated airport has influence on the analysis

- Re-evaluation of models
  - With the inclusion of data that is further refined out of the original dataset

- Iterative approach to models by interfacing with dispatch team
  - Key aspect to evaluating the models and refining the approach to the analysis
  - Source of improvement proposals by creating objectives based on the experience of users
    e.g. creation of categories of NOTAMs that are flagged for being “typical”

Thank you