

R User Meetup

NOTAM classifier in R

João Banha,

Swiss International Airlines, Lta
S/IBC
S/ZC

13.11.2019



Topics we'll touch on

- Me: I'll be quick (promise)
- The Introduction: What is a NOTAM?
- The Problem: What is the difficulty?
- The Approach: What methods did we use?
- The Results: What can we conclude?

Me!
Very short summary



The boring part (Me)



Setúbal 1991

Lisboa
(2008-2015)

Technical University of Lisbon – MSc.
Aerospace Engineering

Lisboa
(June 2016-May 2018)
TAP Maintenance & Engineering
Avionics Specialist



Zürich
(July 2018–December 2018)
SWISS International Airlines -IT
Intern

Zürich
(January 2019 - Present)
SWISS International Airlines
Operations Steering Business Analyst

The Introduction

What is a NOTAM



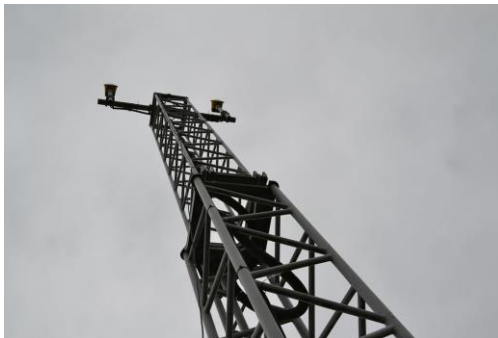
The Introduction

What if?

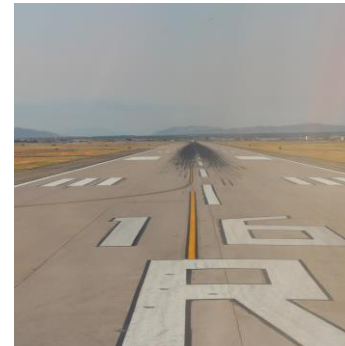
Military area



New structure near airport



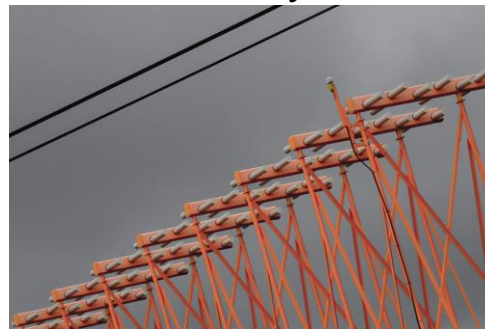
Closed airport/runway



Fireworks



Unavailability of services/systems



The Introduction

What if?

Military area



New structure near airport



Closed airport/runway



Need to inform pilots

Fireworks



Unavailability of services/systems



The Introduction

What if?

Military area



New structure near airport

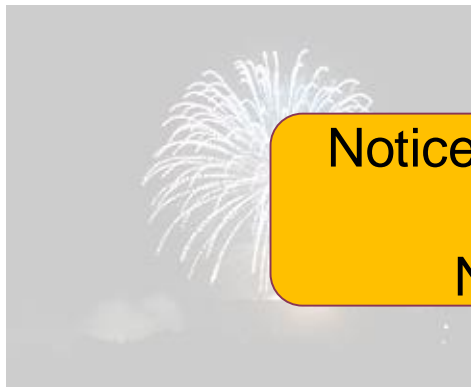


Closed airport/runway

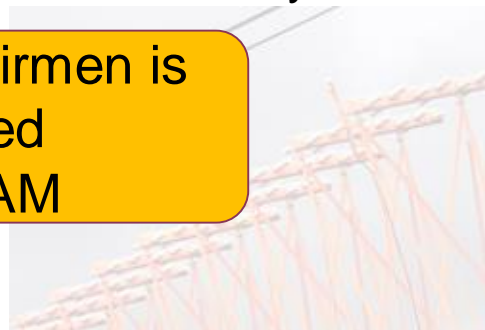


Need to inform pilots

Fireworks



Unavailability of services/systems



Notice to airmen is issued
NOTAM

The Introduction

Format of a NOTAM

H7501/18 NOTAMN Q)EGPX/QWGLW/IV/M/W/100/195/5716N00443W124 A)EGPX B)1810050700 C)1810051742 E)GLD FLY WO SSR TRANSPONDER WI AREA 553000N 0063000W - 563000N 0063000W - 571500N 0050000W - 590000N 0050000W - 590000N 0024500W - 573000N 0013000W - 570000N 0013000W - 56700N 0023000W - 56703N 0040000W - 561400N 0040926W - 555330N 0050000W - 553000N 0050000W - 553000N 0063000W REF AIP ENR 1.1 F)FL100 G)FL195

A0701/18 NOTAMN Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD. DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT.

A1884/17 NOTAMR A1302/17 Q)DRRR/QMRAP/IV/NBO/A/000/999/2015N00059E005 A)GATS B)1712231134 C)1803241200EST D)MON WED FRI 0800-1200 E)NO LANDING AND TAKING OFF IN TESSALIT A/D DUE TO RWY CLOSED FOR UNAUTHORIZED FLIGHTS. AUTHORIZATION MUST BE OBTAINED FROM MALIAN AIR FORCE. OPERATIONS COORDINATION CENTRE E-MAIL: PLANIF.CCOA1(AT)GMAIL.COM PHONE:(00223)20221631 TESSALIT UNPAVED RWY 05R/23L IS CLOSED DUE TO MAINTENANCE WORKS EVERY MONDAY, WEDNESDAY AND FRIDAY FROM 0800Z TO 1200Z. TESSALIT RWYS 05L/23R AND 15/33 REMAIN OPENED FOR ALL AUTHORIZED TRAFFIC. FOR EMERGENCY FLIGHTS ON RWY 05R/23L DURING MAINTENANCE, AIR OPERATORS MUST COORDINATE WITH MINUSMA TESSALIT TELEPHONE : (+223)95999596 AND MINUSMA AVIATION E-MAIL: MINUSMA-AIROPERATIONS CENTRE(AT)UN.ORG TELEPHONE : (+223)94951019/94951018 GND TO FL160

The Introduction

Format of a NOTAM – Structured & Unstructured Data

A0701/18 NOTAMN

Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO

B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.

DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT.

Tags (Q, A, B, C, D, E, F, G) identifying structured data (green), and free-text (blue)

The Problem



The Problem

What (else) is wrong?

A0701/18 NOTAMN

Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO

B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.

DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT

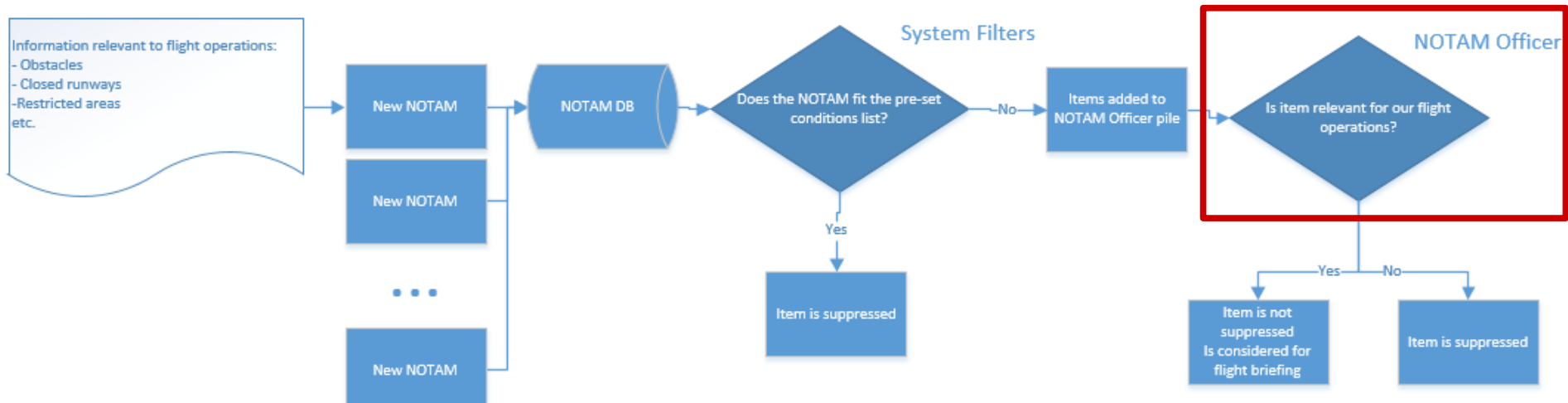
LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT.

X ~3000 / day
(~1M/year)

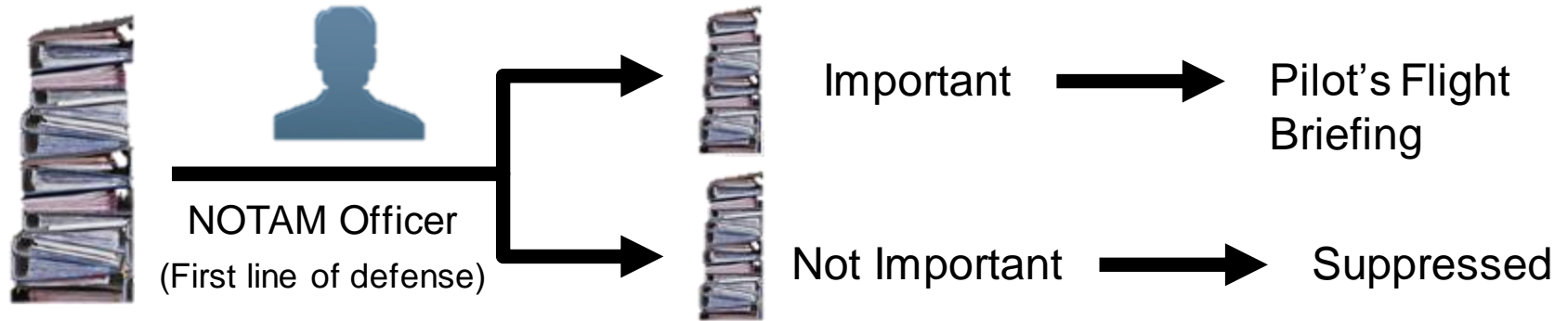
The Introduction

How does it work within SWISS?



The Introduction

What are we looking at?



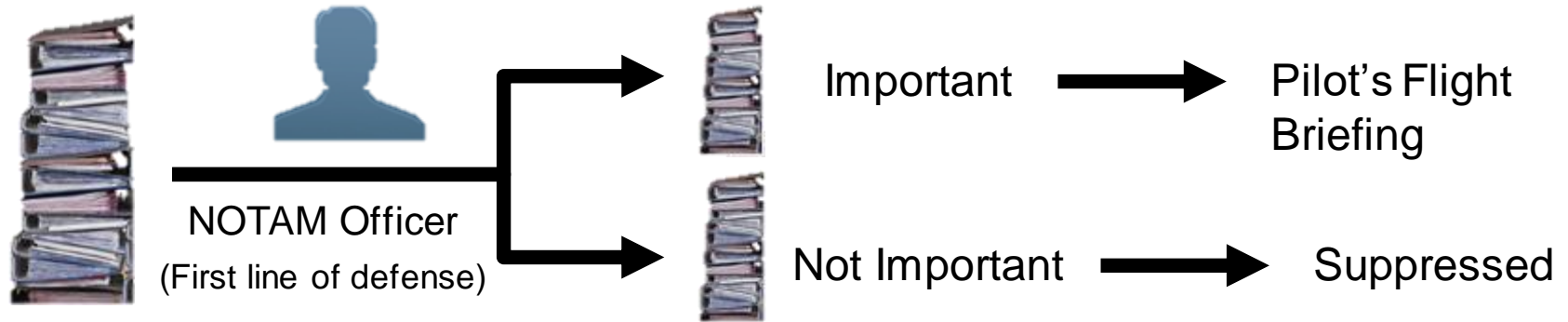
The Problem

Quantity overview

≈

The Introduction

What are we looking at?



Can we help on this step?

The Approach



The Approach

Easy part

A0701/18 NOTAMN

Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO

B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.

DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT.

The Approach

Easy part

| Subject condition | Geospatial coordinates | Radius | Flight Levels (height) |
|-------------------|--|--------|------------------------|
| Message subject | | | |
| Valid date range | A0701/18 NOTAMN Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD. DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT. | | Region |

The Approach

Easy part

Subject condition Geospatial coordinates Radius Flight Levels (height)

Message subject

A0701/18 NOTAMN

Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO
B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.

Region

Valid date range

DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT
LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT
LDA 7000FT.

Free features!

The Approach

Fun part

Whatever someone
decided to write

Too much
/ Too little

Contractions
(sometimes!)

Coordinates
/ Flight Levels
(again ?)

A0701/18 NOTAMN
Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO
B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.
DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT
LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT
LDA 7000FT.

The Approach

Fun part

Whatever someone
decided to write

Too much
/ Too little

Contractions
(sometimes!)

Coordinates
/ Flight Levels
(again ?)

A0701/18 NOTAMN

Q)PHZH/QMRLC/IV/NBO/A/000/999/1943N15503W005 A)PHTO

B)1808311700 C)1809010300 E)RWY 08/26 W 2800FT CLSD.

DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT

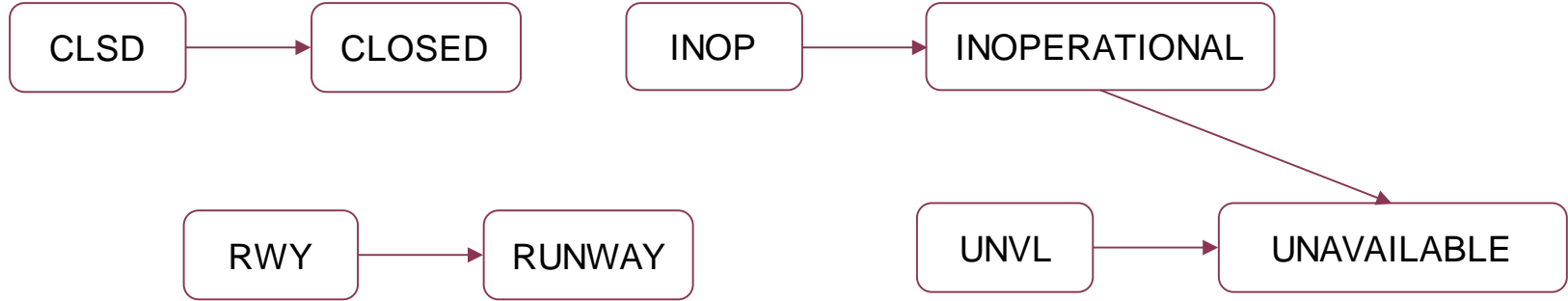
LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT

LDA 7000FT.

Not ready to use
information!

The Approach

Processing free text



E)RWY 08/26 W 2800FT CLSD. DECLARED DIST: RWY 08 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT. RWY 26 TORA 7000FT TODA 7000FT ASDA 7000FT LDA 7000FT.

The Approach

Processing free text

```
1 ABN;airport beacon
2 U/S;UNSERVICEABLE
3 ABV;above
4 ACC;area control center ARTCC
5 ACCUM;accumulate
6 ACFT;aircraft
7 ACR;air carrier
8 ACT;active
9 AD;AERODROME
10 ADJ;adjacent
11 ADZD;advised
12 AFD;airport facility directory
13 AFSS;automated flight service station
14 AGL;above ground level
15 ALS;approach light system
16 ALT;altitude
17 ALTM;altimeter
18 ALTN;alternate
19 ALTPLY;alternately
20 ALSTG;altimeter setting
21 AMDT;amendment
22 AMGR;airport manager
23 AMOS;automatic meteorological observing system
24 AP;airport
25 APCH;approach
```

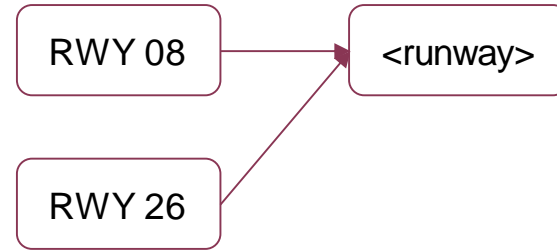
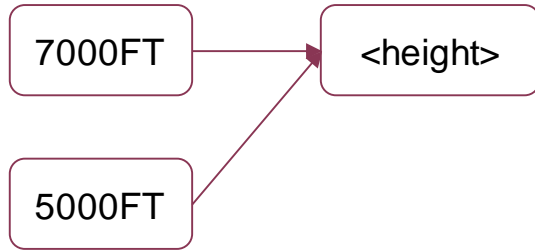


```
349 UNMKD;unmarked
350 UNMNT;unmonitored
351 UNREL;unreliable
352 UNUSBL;unusable
353 VASI;visual approach slope indicator
354 VDP;visual descent point
355 VIA;by way of
356 VICE;instead versa
357 VIS;visibility
358 VMC;visual meteorological conditions
359 VOL;volume
360 VOR;vhf Omni-directional radio range
361 VORTAC;VOR and TACAN
362 W;west
363 WB;westbound
364 WED;Wednesday
365 WEF;with effect from or effective from
366 WI;within
367 WKDAYS;Monday through Friday
368 WKEND;Saturday and Sunday
369 WND;wind
370 WPT;waypoint
371 WSR;wet snow on runways
372 WTR;water on runways
373 WX;weather
```

Mapped contractions to full words

The Approach

Processing free text



E)RWY 08/26 W 2800FT CLSD. DECLARED DIST: RWY 08 TORA
7000FT TODA 7000FT ASDA 7000FT LDA 7000FT. RWY 26 TORA
7000FT TODA 7000FT ASDA 7000FT LDA 7000FT.

The Approach

Processing free text

```
1 \b\d+ ([, .]\d{1,2})? ?(K|M)HZ\b;<frequency>
2 \bCH[0-9]{2} [XY]\b;<channel>
3 \bFL[0-9]{3}\b;<flightlevel>
4 \b[0-9]+[.]?[0-9]? ?FT( ?(AMSL|AGL))?\b;<height>
5 \b[0-9]+[NS] ?[0-9]+[EW]\b;<coordinate>
6 \b\d+ ([, .]\d{1,2})? ?(M|KM|NM)\b;<distance>
7 \b(RUNWAY|RWY) ?[0-9]{2} (R|L|C)? (/ [0-9]{2} (R|L|C)?)? (?!\w);<runway>
```

Special numbers mapped with regular expressions

The Approach

Processing free text

| | |
|----|---------|
| 1 | a |
| 2 | about |
| 3 | above |
| 4 | after |
| 5 | again |
| 6 | against |
| 7 | all |
| 8 | am |
| 9 | an |
| 10 | and |
| 11 | any |
| 12 | are |
| 13 | as |
| 14 | at |
| 15 | be |
| 16 | because |
| 17 | been |
| 18 | before |
| 19 | being |
| 20 | below |
| 21 | between |
| 22 | both |
| 23 | but |
| 24 | by |
| 25 | cannot |

| | |
|----|---------|
| 26 | could |
| 27 | did |
| 28 | do |
| 29 | does |
| 30 | doing |
| 31 | down |
| 32 | during |
| 33 | each |
| 34 | few |
| 35 | for |
| 36 | from |
| 37 | further |
| 38 | had |
| 39 | has |
| 40 | have |
| 41 | having |
| 42 | he |
| 43 | her |
| 44 | here |
| 45 | hers |
| 46 | herself |
| 47 | him |
| 48 | himself |
| 49 | his |
| 50 | how |

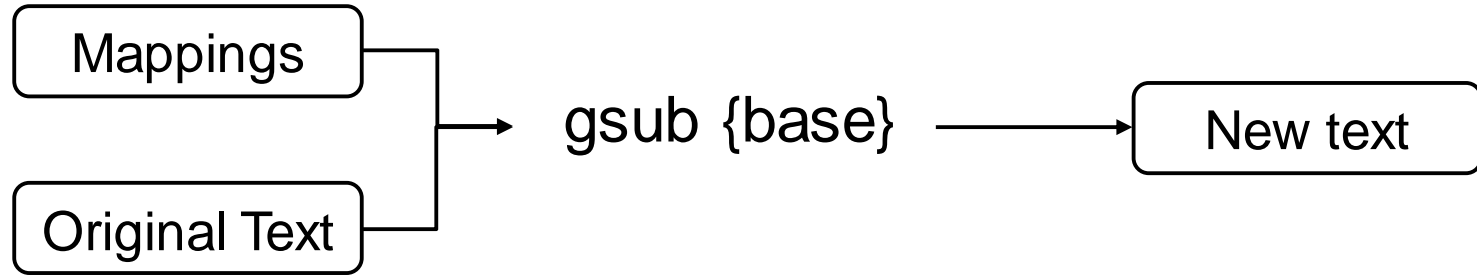


| | |
|-----|------------|
| 97 | through |
| 98 | to |
| 99 | too |
| 100 | under |
| 101 | until |
| 102 | up |
| 103 | very |
| 104 | was |
| 105 | we |
| 106 | were |
| 107 | what |
| 108 | when |
| 109 | where |
| 110 | which |
| 111 | while |
| 112 | who |
| 113 | whom |
| 114 | why |
| 115 | with |
| 116 | would |
| 117 | you |
| 118 | your |
| 119 | yours |
| 120 | yourself |
| 121 | yourselves |

Listed stop words

The Approach

Processing free text



```
> message <- "All the good stuff"  
> gsub("stuff", "things", message)  
[1] "All the good things"
```

The Approach

Extracting features from free-text

{ngram} + TF-IDF

What are the features?

What are the values of the features?

The Approach

{ngram} + TF-IDF

{ngram}

1-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM E)VDF 120.100 MHZ WITHDRAWN REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 AND 4-15, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.



The Approach

{ngram} + TF-IDF

{ngram}

1-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM E)VDF 120.100 MHZ WITHDRAWN REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 AND 4-15, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.



The Approach

{ngram} + TF-IDF

{ngram}

1-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM F)VDF 120.100 MHZ **WITHDRAWN** REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 **AND 4-15**, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.

The Approach

{ngram} + TF-IDF

{ngram}

2-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM E)VDF 120.100 MHZ WITHDRAWN REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 AND 4-15, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.



The Approach

{ngram} + TF-IDF

{ngram}

2-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM E)VDF 120.100 MHZ WITHDRAWN REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 AND 4-15, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.

The Approach

{ngram} + TF-IDF

{ngram}

3-gram

A3936/17 NOTAMR A0537/16 Q)ENOR/QNXAW/V/B/A/000/999/6829N01641E005
A)ENEV B)1710301454 C)PERM E)VDF 120.100 MHZ WITHDRAWN REF AIP NORWAY
AD 2 ENEV 4-3, 4-7, 4-11, 4-13 AND 4-15, SID AND STAR CHARTS DATED 29 MAY
2014, AD 2 ENEV 5-7, 5-9 AND 5-10, INSTRUMENT APPROACH CHARTS DATED 30
MAR 2017.



The Approach

{ngram} + TF-IDF

$$f_{t,d} = \frac{\text{number of occurrences in message}}{\text{number of terms in message}} : \text{Term Frequency}$$

$$idf_t = \log\left(\frac{\text{number of messages}}{\text{number of messages where term appears}}\right) : \text{Inverse Document Frequency}$$

$$w_{t,d} = f_{t,d} * idf_t : \text{Overall Importance}$$

The Approach

Feature Engineering – Expert Intuition

What information can we include that is accounted for by experts?



The Approach

Feature Engineering – Expert Intuition

What information can we include that is accounted for by experts?

- How close is the NOTAM center to an airport we operate?
Feature: `closestDestAirportDistance`
- How close is the NOTAM center to an alternate airport?
Feature: `closestAltAirportDistance`
- How long is the free text portion of the NOTAM?
Feature: `freeTextSize`

The Approach

Feature Collection

Feature matrix:

- 16 features from structured section
- 3 engineered features
- 200 ngram features

Label from historical data of messages that were/ weren't suppressed (0/1)

| | | | | |
|-----------------------------------|---|--------------------------------|-------------------------------|----------------------------|
| [1] "flmin" | "flmax" | "lat_rad" | "long_rad" | "rad" |
| [6] "nat_N" | "traffic_ifr" | "traffic_vfr" | "traffic_checklist" | "purp_N" |
| [11] "purp_B" | "purp_O" | "purp_M" | "scope_A" | "scope_E" |
| [16] "scope_W" | "closestDestAirportDistance" | "closestAltAirportDistance" | "freeTextSize" | "<coordinate> " |
| [21] "<coordinate> <coordinate> " | "<coordinate> <coordinate> <coordinate> " | "<distance> " | "<distance> ABOVE " | "<distance> ABOVE GROUND " |
| [26] "<distance> RADIUS " | "<flightlevel> " | "<flightlevel> <flightlevel> " | "<frequency> " | "<height> " |
| [31] "<number> " | "<number> <number> " | "<number> <number> <number> " | "<number> <number> PCT " | "<number> PCT " |
| [36] "<number> PCT WET " | "<runway> " | "<runway> <number> " | "<runway> <number> <number> " | "<runway> CLOSED " |
| [41] "ABOVE " | "ABOVE GROUND " | "ABOVE GROUND LEVEL " | "ACTIVATED " | "ACTIVATED AIRSPACE " |
| [46] "ACTIVATED AIRSPACE CLOSED " | "ACTIVATED TEMPORARY " | "ACTIVE " | "AERODROME " | "AIP " |



| | | | | |
|----------------------------|-----------------------------|------------------------------|-------------------------------------|-----------------------------|
| [166] "RAIM " | "RANGE " | "REF " | "REF AIP " | "REMARK " |
| [171] "RESTRICTED " | "RESTRICTED AREA " | "RESTRICTED AREA ACTIVATED " | "ROUTE " | "RUNWAY " |
| [176] "RUNWAY <runway> " | "RUNWAY <runway> <number> " | "RUNWAY CLOSED " | "RUNWAY CLOSED <runway> " | "SERVICE " |
| [181] "SERVICE NOW " | "SIDE " | "SOUTH " | "SURFACE " | "SURFACE <flightlevel> " |
| [186] "SURFACE <height> " | "SYSTEM " | "SYSTEM <runway> " | "SYSTEM UNSERVICEABLE " | "TAKE " |
| [191] "TAKE PLACE " | "TAXIWAY " | "TAXIWAY CLOSED " | "TAXIWAY CLOSED TAXIWAY " | "TAXIWAY TAXIWAY " |
| [196] "TEMPORARY " | "TEMPORARY RESTRICTED " | "TEMPORARY RESTRICTED AREA " | "TOWER " | "TRAFFIC " |
| [201] "UNSERVICEABLE " | "UNSERVICEABLE <runway> " | "UNSERVICEABLE INSTRUMENT " | "UNSERVICEABLE INSTRUMENT LANDING " | "VHF " |
| [206] "VHF OMNI " | "VHF OMNI DIRECTIONAL " | "WET " | "WET OBSERVATION " | "WET OBSERVATION <number> " |
| [211] "WILL " | "WILL TAKE " | "WILL TAKE PLACE " | "WIP " | "WITHIN " |
| [216] "WITHIN <distance> " | "WITHIN AREA " | "WORK " | "WORK PROGRESS " | |

The Approach

Model Training

Model training:

```
xgbparams <- list(booster = "gbtree",  
                  silent = 0,  
                  eta = 0.3,  
                  max_depth = 10,  
                  subsample = 0.5,  
                  objective = "binary:logistic",  
                  nthread = 4)
```

```
bst_model <- xgb.train(params = xgbparams,  
                      data = dtrain,  
                      nround = 150,  
                      early_stopping_rounds = 5)
```

{xgboost}

~ 10min training time (laptop)

The Results



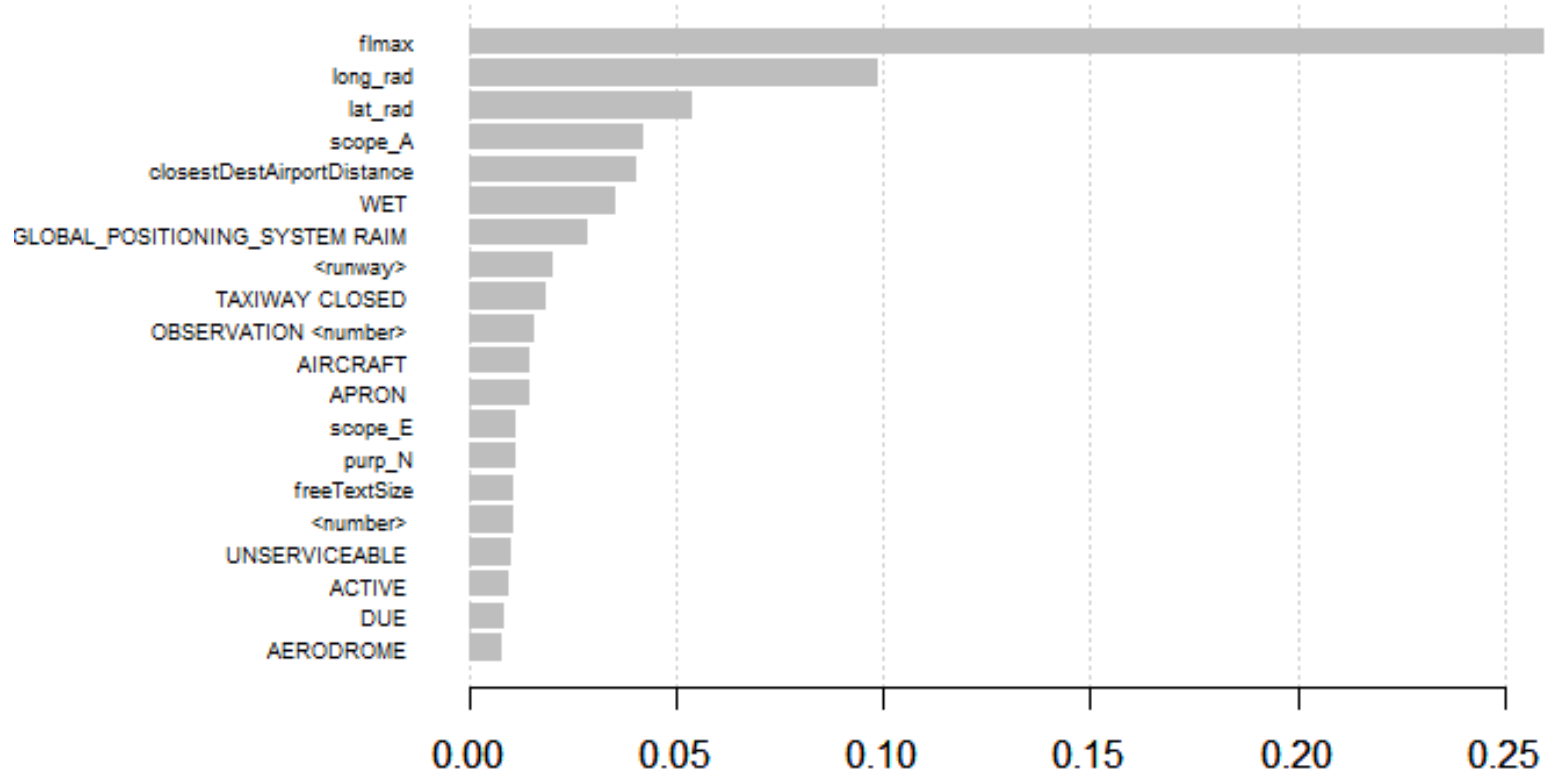
The Results

Model – Feature Importance

```
model <- xgb.dump(bst_model, with_stats = TRUE)
feature_names <- dimnames(dtrain)[[2]]
importance_matrix <- xgb.importance(feature_names, model = bst_model)
xgb.plot.importance(importance_matrix[0:20])
```

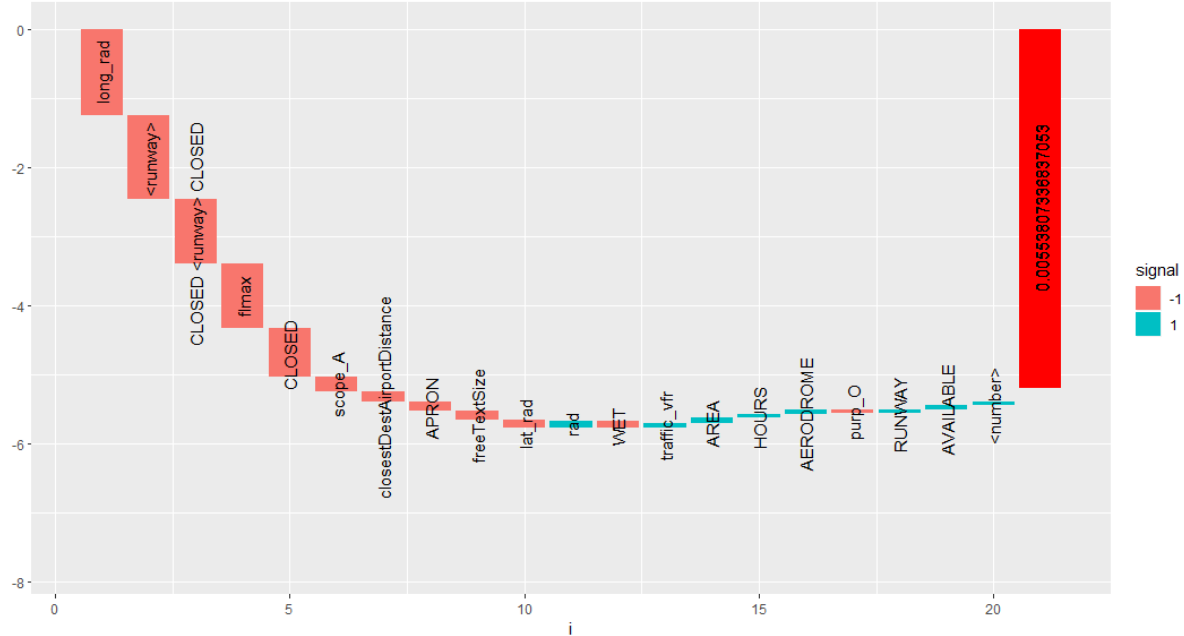
The Results

Model – Feature Importance



The Results

Model – Feature Importance on one item



The Results

Model – Prediction

Prediction on Test Dataset:

```
pred <- predict(bst_model, dtest)
pred.resp <- ifelse(pred > 0.5, 1, 0)
```

{stats}

The Results

Model – Accuracy

```
Accuracy : 0.9413  
95% CI : (0.9403, 0.9422)  
No Information Rate : 0.5501  
P-Value [Acc > NIR] : < 2.2e-16
```

The Results

Model – Confusion Matrix

Accuracy : 0.9413

95% CI : (0.9403, 0.9422)

No Information Rate : 0.5501

P-Value [Acc > NIR] : < 2.2e-16

How good is this really?

The Results

Model – Confusion Matrix

Accuracy : 0.9413
95% CI : (0.9403, 0.9422)
No Information Rate : 0.5501
P-Value [Acc > NIR] : < 2.2e-16

Confusion Matrix and Statistics

| Prediction | | Reference | |
|------------|--------|-----------|---|
| | | 0 | 1 |
| 0 | 120477 | 7781 | |
| 1 | 5690 | 95385 | |

The Results

Model – Confusion Matrix

Accuracy : 0.9413
95% CI : (0.9403, 0.9422)
No Information Rate : 0.5501
P-Value [Acc > NIR] : < 2.2e-16

Confusion Matrix and Statistics

| Prediction | | Reference | |
|------------|--------|-----------|---|
| | | 0 | 1 |
| 0 | 120477 | 7781 | |
| 1 | 5690 | 95385 | |

Potentially not important information being shown

The Results

Model – Confusion Matrix

Accuracy : 0.9413

95% CI : (0.9403, 0.9422)

No Information Rate : 0.5501

P-Value [Acc > NIR] : < 2.2e-16

Confusion Matrix and Statistics

| | | Reference | |
|------------|---|-----------|-------|
| | | 0 | 1 |
| Prediction | 0 | 120477 | 7781 |
| | 1 | 5690 | 95385 |

Potentially important
information being hidden

Potentially not important
information being shown

The Results

Model – Deeper Analysis

The Results

Model – Deeper Analysis

What to discuss with demo:

- How probabilities are distributed
- How, dividing by bins of probabilities, we can achieve greater accuracies
- How, choosing the bins of greater accuracy, we can reduce rather than eliminate the workload (3 piles suggestion)

The Results

Model – Demo

What to discuss with demo:

- How accuracy within clusters differs (pure clusters – candidates for automization)
- Point out how types of clusters may be characteristic of certain geographical areas
- Point out how the clustering can bring insight to how people group messages instinctively (bring the user's learning to the paper)

Gradient Boosted Trees - Classification Distribution

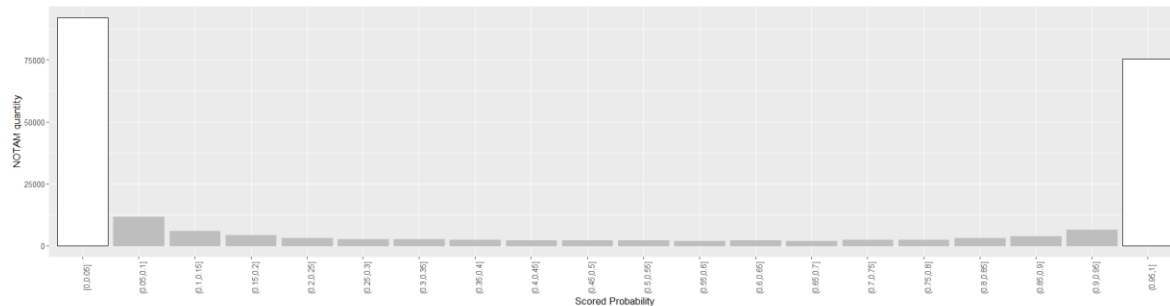
Left Hand
 Right Hand
 Width of bins:
 Number of left bins:

Left hand NOTAM removal:
 Percentage of NOTAMs removed: 40.11
 Accuracy of removal of NOTAMs: 98.92

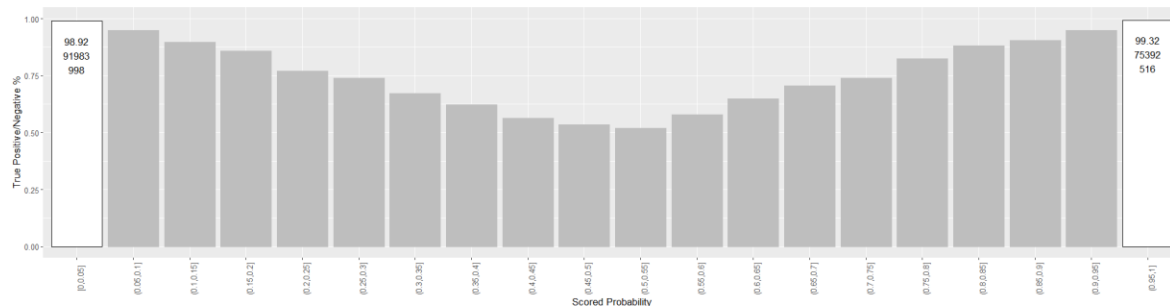
Right hand NOTAM removal:
 Percentage of NOTAMs removed: 32.87
 Accuracy of removal of NOTAMs: 99.32

Overall NOTAM removal:
 72.98

Probability distribution of test data



Accuracy distribution of test data

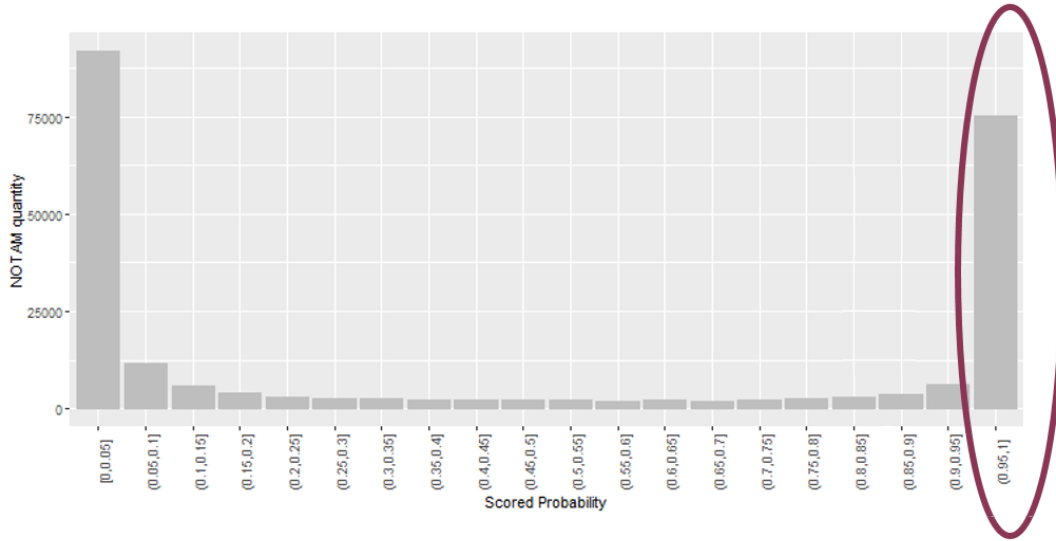


Conclusions:

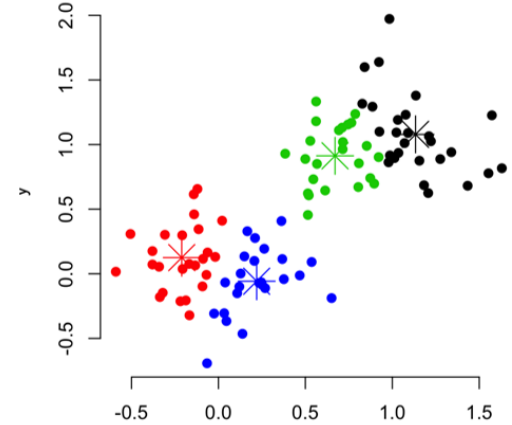
- Maybe there isn't "sufficient" accuracy that justifies automatically throwing out messages.
- We can instead suggest if a message should be suppressed by pre-classifying.
- We have a large number of messages either automated or "pre-classified", leaving more time to evaluate other, more ambiguous messages

The Results

Model – Deeper Analysis



+

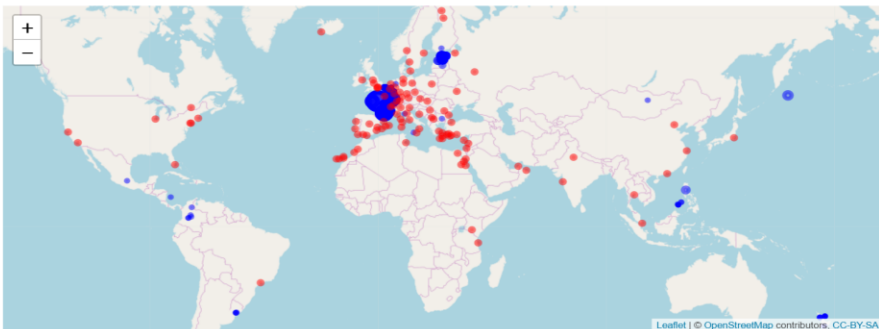


kmeans {stats}

Precision of selected cluster: 99.45 %

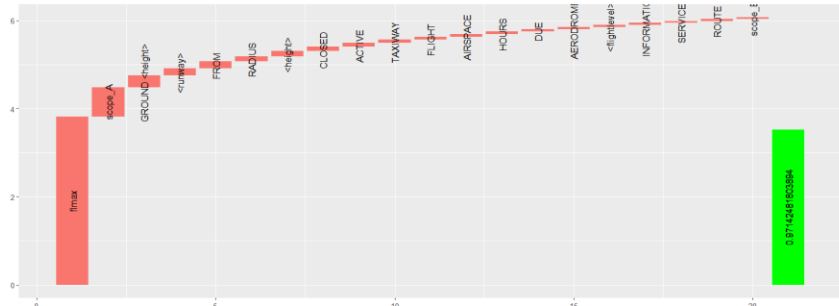
Cluster Selection

Show only mismatches



Show 10 entries

Item ID for explanation



Search:

| enrichedE | supress | item_id | header |
|---|---------|---------|--------------------|
| 1 AIRSPACE RESERVATION ACTIVATED AIRSPACE RESTRICTED ACTIVE RADIUS <distance> NXT COORDINATE <number> <number>N0733446 <number>WEST GROUND <height> | 1 | 4085635 | A0225/18 NOTAMN |
| 2 RESTRICTED AREA ACTIVATED LOW ALTITUDE AREA ACTIVATION SR <number> SS <number> AREA R191A HERAULT EST <number> <number> ACTIVE AREA R191C HERAULT NORD <number> <number> ACTIVE AREA R193A TARN EST <number> <number> ACTIVE AREA R589A LOT <number> <number> ACTIVE AREA R589B LOT <number> <number> ACTIVE AREA R590A MENDE SUD <number> <number> ACTIVE AREA R590B MENDE NORD <number> <number> ACTIVE AREA R591 ALLIER <number> <number> ACTIVE AREA R592 CANTAL <number> <number> ACTIVE SURFACE <height> | 1 | 4080209 | X0016/18 NOTAMN |
| 3 OBSTACLE ERECTED REF AIP ENROUTE <number> <number> NEW OBSTRUCTION OBSTACLE ERECTED AEOIC PROPELLER REGION SICILIA DISTRICT CALTANISSETTA TOWN CALTANISSETTA SITE INCAUSO SCORSONE MIMIANI SANTALENA COMUNELLO POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED POSITION COORDINATE WGS84 <coordinate> ELEVATION AGL125M <height> ELEVATION AMSL <distance> <height> ICAO SIGNAL UNPROVIDED | 1 | 4079643 | A0630/18 NOTAMN |
| 4 RESTRICTED AREA ACTIVATED LOW ALTITUDE AREAS ACTIVE SR <number> SS <number> AREA R45A BOURGOGNE <number> <number> ACTIVE AREA R45B AUTUNOIS <number> <number> ACTIVE AREA R45C ARBOIS <number> <number> ACTIVE AREA R45D DOUBS <number> <number> ACTIVE AREA R45S1 FRANCHE COMTE <number> <number> ACTIVE AREA R45S2 LANGRES <number> <number> ACTIVE AREA R45S3 YONNE <number> <number> ACTIVE AREA R45S4 MACONNAIS OUEST <number> <number> ACTIVE AREA R45S5 MACONNAIS CENTRE <number> <number> ACTIVE AREA R45S<number> MACONNAIS NORD EST <number> <number> ACTIVE AREA R45S<number> MACONNAIS SUD EST <number> <number> ACTIVE AREA R45S7 JURA <number> <number> ACTIVE AREA R45NS <number> <number> ACTIVE SURFACE <height> | 1 | 4065135 | Z0014/18 NOTAMN |
| 5 ACTIVATED HTA10 COASTAL HELICOPTER TRAINING AREA ACTIVE GROUND <height> | 1 | 4051269 | B0275/18 NOTAMN |
| 6 ACTIVATED HTA02 ARDENNES <number> AREA ACTIVE GROUND <height> | 1 | 4051239 | B0273/18 NOTAMN |
| 7 RESTRICTED AREA ACTIVATED LOW ALTITUDE AREAS ACTIVITY SR <number> SS <number> ZONE R57 BRETAGNE <number> <number> ACTIVE ZONE R139 CHER SS <number> ACTIVE ZONE R142 NIEVRE SS <number> ACTIVE ZONE R143 AUVERGNE SS <number> ACTIVE ZONE R144 LOIRE SS <number> ACTIVE ZONE R145 CREUSE SS <number> ACTIVE ZONE R147 CHARENTE <number> <number> ACTIVE ZONE R149D VENDEE <number> <number> ACTIVE ZONE R149E MAINE ANJOU <number> <number> ACTIVE ZONE R165 VIENNE SS <number> ACTIVE SURFACE <height> | 1 | 4045819 | Y0013/18 NOTAMN |

The Summary

- Current state: the challenge of **how to implement**
- Within this context **full automation is not advisable**
- However, **supervised** and **unsupervised ML** can both provide valuable **insights** as well as **assistance in decision making**

Thank you



Questions?