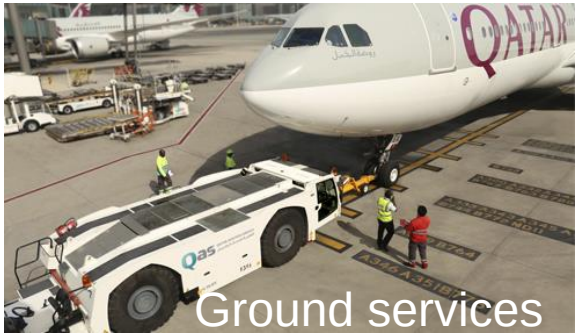
Analysing Incidents using Data Enrichment and Machine Learning

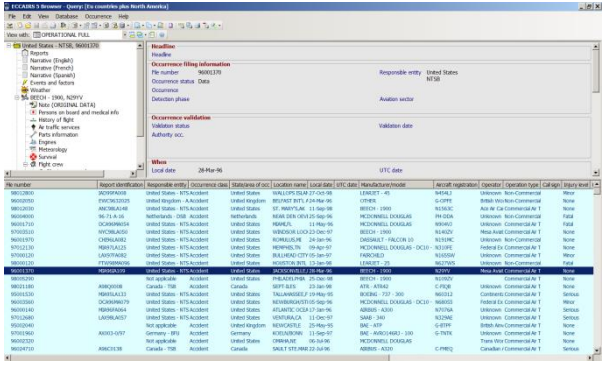
Gerard van Es

ICAO First iSTARS User Group (iUG/01)
ICAO Headquarters, Montreal – 17-19 December 2018

Background – occurrence data



Occurrence data



Background – Issues with reported occurrences

- Incomplete reports (data fields missing);
- Incorrect reports (errors in data fields).

Limits our data analyses!

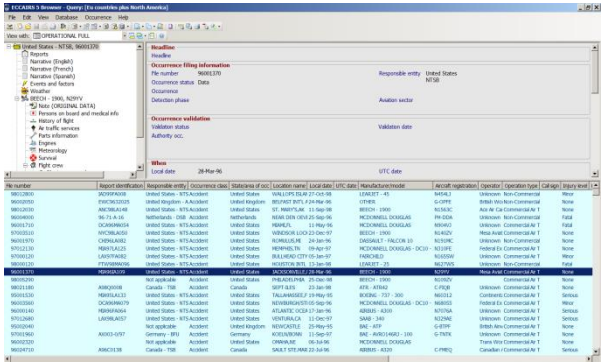
How to solve these problems?

Airport databases

METARs

```
KJFK 272251Z 18012KT 10SM FEW060 SCT250 26/18 A2988 RMK AO2 SLP119 T02610183
KJFK 272151Z 19013KT 10SM FEW060 SCT250 28/18 A2989 RMK AO2 SLP123 T02780178
KJFK 272051Z 18012KT 10SM FEW060 28/18 A2991 RMK AO2 SLP127 T02830183 56020
KJFK 271951Z 18010KT 10SM FEW060 29/18 A2992 RMK AO2 SLP132 T02890183 $
KJFK 271851Z 18009KT 10SM FEW060 29/18 A2994 RMK AO2 SLP138 T02890183
KJFK 271751Z 18012KT 10SM FEW060 29/18 A2996 RMK AO2 SLP147 T02890183 10289 20233 56025
KJFK 271651Z 20011KT 10SM CLR 28/19 A2999 RMK AO2 SLP155 T02830189
KJFK 271551Z 21008KT 10SM CLR 27/19 A3001 RMK AO2 SLP163 T02720189
KJFK 271451Z 19008KT 10SM FEW250 26/18 A3004 RMK AO2 SLP172 T02560178 58006
KJFK 271351Z 19009KT 10SM FEW250 24/18 A3005 RMK AO2 SLP176 T02440183
KJFK 271251Z 20007KT 10SM FEW250 24/19 A3006 RMK AO2 SLP178 T02440189 $
KJFK 271151Z 21004KT 10SM FEW250 23/18 A3006 RMK AO2 SLP178 T02330183 10233 20194 53006 $
KJFK 271051Z 23004KT 10SM FEW250 21/18 A3005 RMK AO2 SLP175 T02110178 $
KJFK 270951Z 24005KT 10SM FEW250 20/17 A3004 RMK AO2 SLP171 T02000167 $
KJFK 270851Z 26006KT 10SM CLR 21/16 A3004 RMK AO2 SLP171 T02060161 56010 $
KJFK 270751Z 25004KT 10SM CLR 21/15 A3004 RMK AO2 SLP173 T02060150 $
KJFK 270651Z 25006KT 10SM CLR 22/14 A3005 RMK AO2 SLP175 T02170144 $
KJFK 270551Z 25007KT 10SM CLR 21/17 A3007 RMK AO2 SLP182 T02110172 10233 20211 57016 $
```

Occurrence data



Register database

Radar/ADS-B

Example using datamining to code occurrence category

- Random Forest classifier (gave best results out of 5 techniques);
- Random Forest builds multiple decision trees and merges them together to get a more accurate prediction;
- Performance on training dataset gave accuracy of 94%;
- Performance on complete set gave accuracy of 79% (+48,000 occurrences);
- Difficulties with SCF-NP/PP and ADRM & ATC occurrences.

Occurrence data enrichment

- Add information to occurrence data normally not reported, e.g.:

  - Flown track;
  - Duration of flight;
  - Additional weather data (en-route);
  - Taxi track;
  - **Non-occurrence flights with same details as for occurrences**.

Extends analyses possibilities

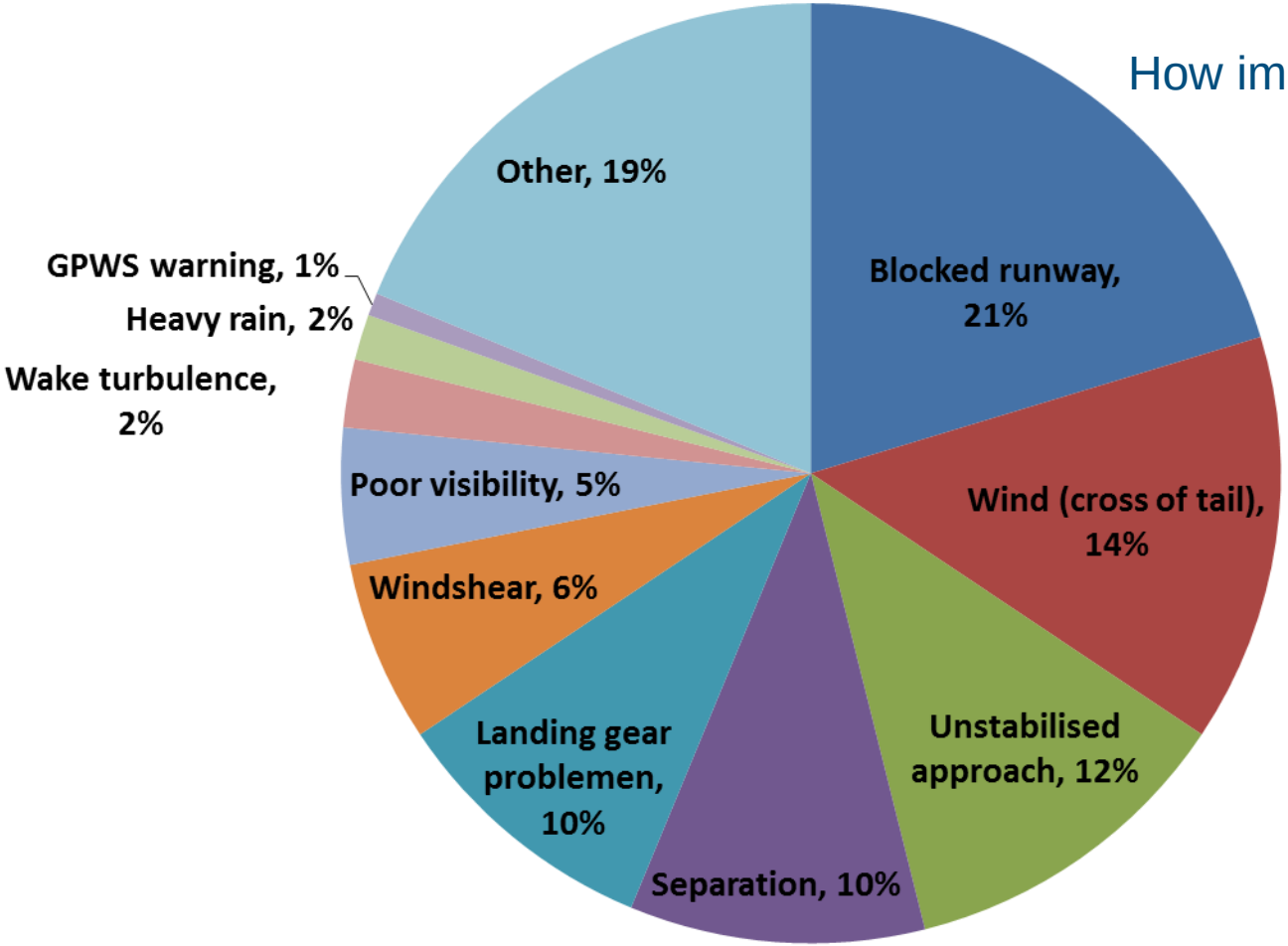
Why look at non-occurrence flights?

- Normalising of occurrences (classical approach);
- Allows to explore contribution of certain factors to overall risk increase;
- Will highlight importance of factors not visible from occurrence data only.

Example case

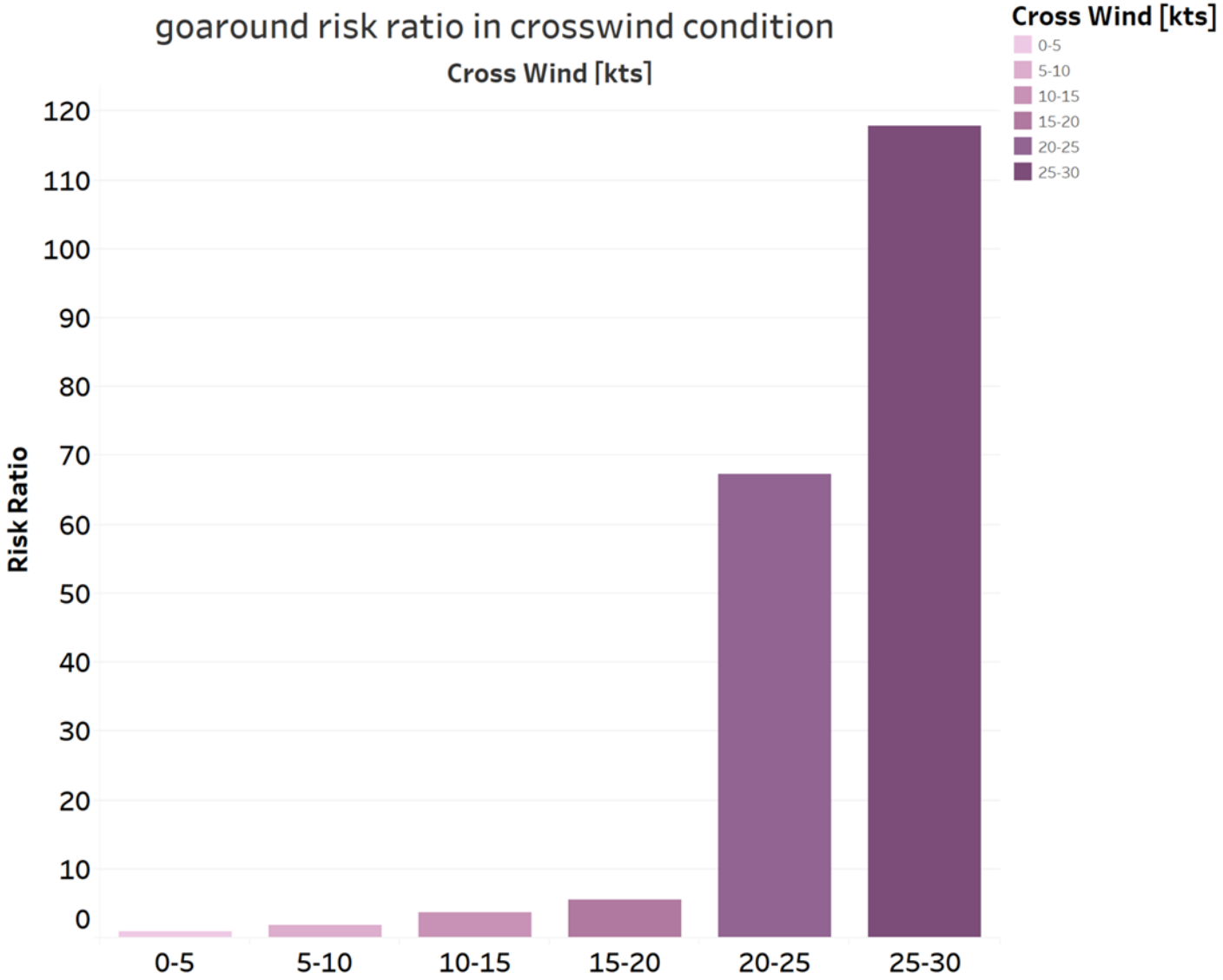
- Reported go-arounds at an airport;
- What is driving the rate of go-arounds at the airport?
- Simply plotting number of go-arounds by year etc. will not be very meaningful...

Typical reasons for a go-around

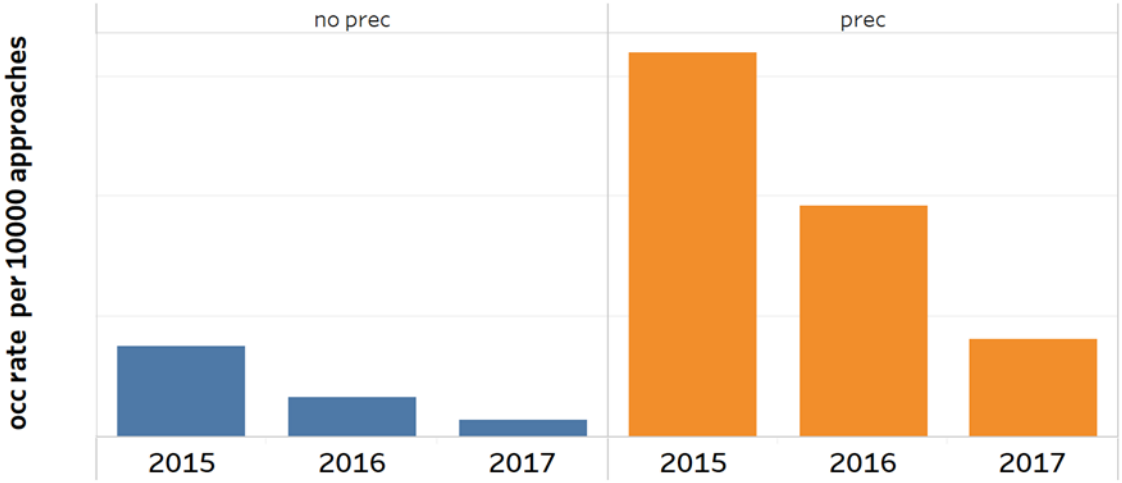


How important are these factors?

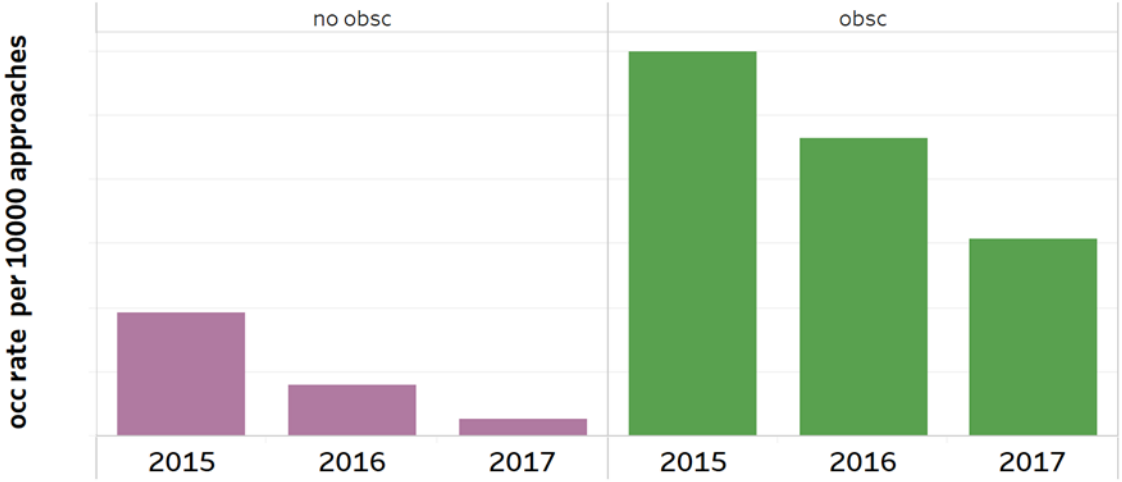
Relation between crosswind and go-around rate


goaround risk ratio in crosswind condition
Cross Wind [kts]
Risk Ratio
120
110
100
90
80
70
60
50
40
30
20
10
0
0-5
5-10
10-15
15-20
20-25
25-30
Cross Wind [kts]
0-5
5-10
10-15
15-20
20-25
25-30

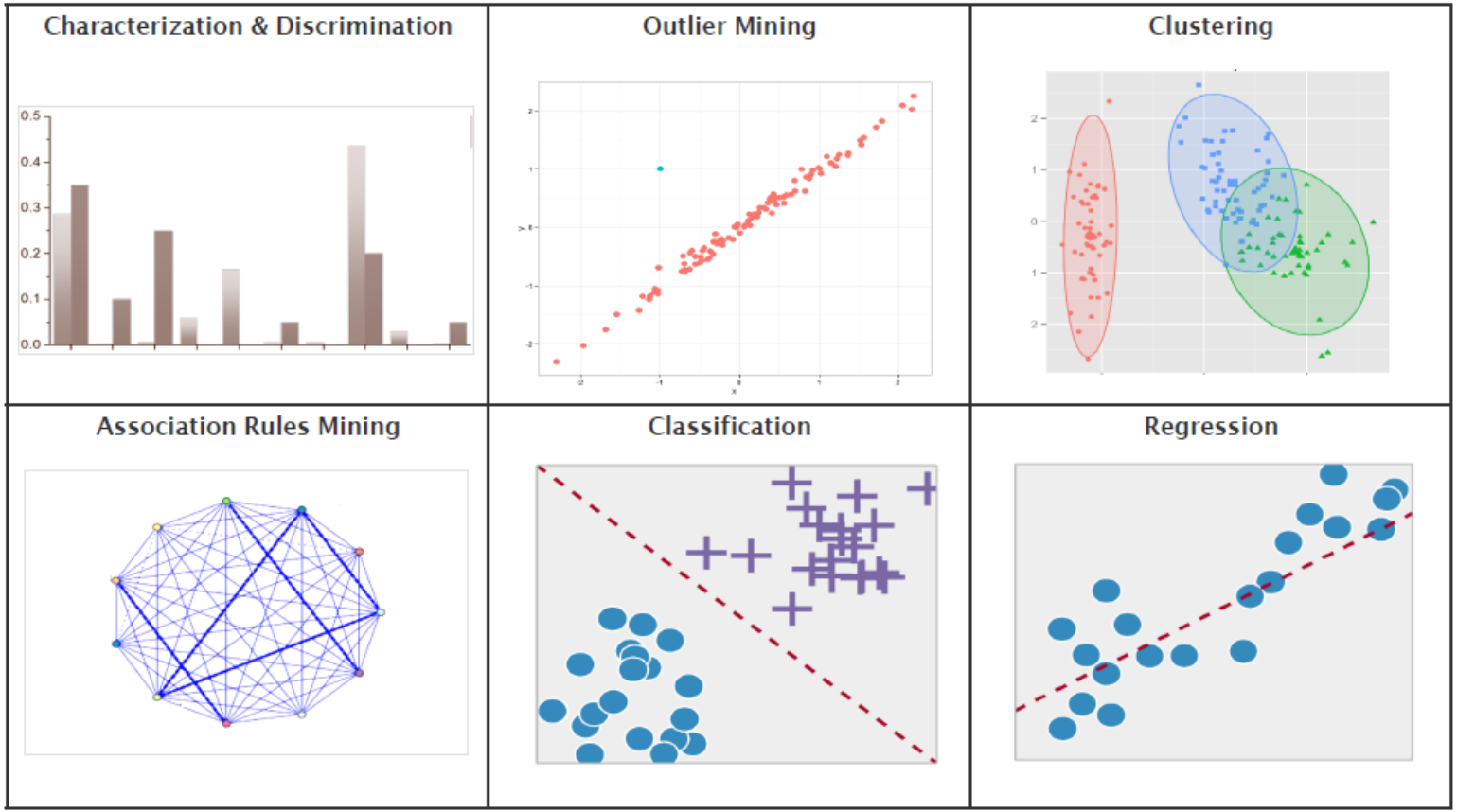
go-around occ rate per 10000 operations in precipitation condition



go-around occ rate per 10000 operations in obscuration condition

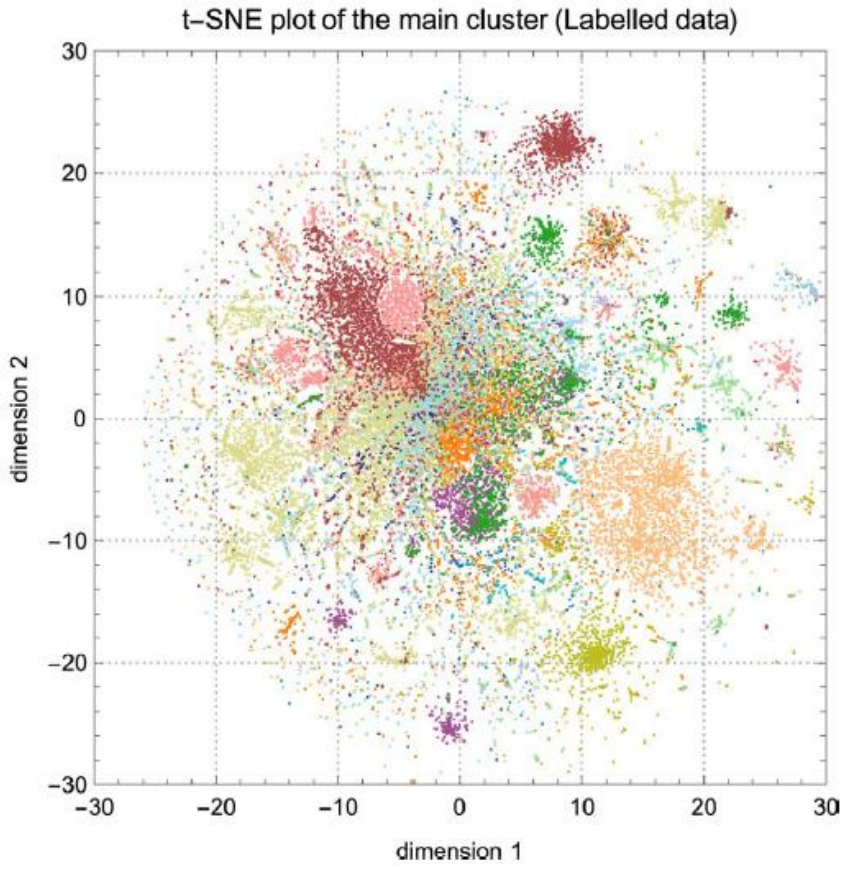


Example Machine Learning / Data Mining methods



* Credits: Innaxis Keynote, EUROCONTROL TIM/ART workshop ML and AI, 24-25 sept 2018

Text mining of occurrence data by applying t-SNE algorithm to find patterns



Occurrence categories

Applying machine learning to enriched occurrence data

- Will help to predict under which conditions there is higher risk of a particular occurrence;
- For instance, combination of wind, period in time, light conditions and runway use could increase go-around rate;
- Need sufficient amount of occurrence data to build accurate models.

Machine Learning – important rules

- ❑ You need sufficient and reliable data!
- ❑ Define your outcome!
- ❑ It is not a push button exercise – You must understand aviation!

Dedicated to innovation in aerospace

Fully engaged

Netherlands Aerospace Centre

NLR Amsterdam
Anthony Fokkerweg 2
1059 CM Amsterdam

p) +31 88 511 31 13 f) +31 88 511 32 10
e) info@nlr.nl i) www.nlr.nl

NLR Marknesse
Voorsterweg 31
8316 PR Marknesse

p) +31 88 511 44 44 f) +31 88 511 42 10
e) info@nlr.nl i) www.nlr.nl