A Boosted Tree framework for runway occupancy and exit prediction

SIDs 2018 - Salzburg

SafeClouds.eu

innaxis

This project is funded by
the European Union

1. Runway utilization problems

Leader misses the procedural runway exit

Follower too close when leader lands

Leader takes too much time in runway

Allocating departures in mixed operations

2. Datasets availability

Scope: Arrivals at **R34 of Vienna airport(LOWW)**

7 sources (datasets) available:

1. Radar track → **Dynamic** data
2. Airport information → **Static** data (**flight plan**)
3. METAR
4. SNOWTAM
5. Visibility
6. WMA (wind profiles)
7. SODAR

} **Meteorogical** data

2.1 Data preparation considerations

- Lack of **format** coherence
- Data **redundancy**
- **Errors** and **outliers**
- Different time **scales** and **sampling**
- Information as **message log**
  - Parsing to tabular format (dataframe)
  - Information mining

```
SPECI KDEN 200728Z 03006KT 4SM -SN BKN026 M06/M08 A2982 RMK AO2 SNB11 VIS LWR E-S P0000
SPECI KDEN 200745Z 02009KT 7SM SCT031 OVC070 M06/M08 A2983 RMK AO2 SNB11E41 P0000
METAR KDEN 200753Z 01009KT 9SM OVC065 M06/M09 A2983 RMK AO2 SNB11E41 SLP105 P0000 T10561089
METAR KDEN 200853Z 03006KT 10SM BKN065 BKN120 M06/M10 A2984 RMK AO2 SLP114 60000 T10611100 51018
METAR KDEN 200953Z 34011KT 10SM BKN070 BKN120 M06/M13 A2987 RMK AO2 SLP124 T10611128
METAR KDEN 201053Z 35012KT 10SM BKN075 M08/M14 A2988 RMK AO2 SLP148 T10781139
METAR KDEN 201153Z 00000KT 10SM FEW070 SCT120 M12/M16 A2989 RMK AO2 SLP167 60000 T11171161 11044 21117 51012
METAR KDEN 201253Z 02006KT 10SM FEW040 SCT070 SCT100 M12/M18 A2991 RMK AO2 SLP168 T11221178
METAR KDEN 201353Z 34008KT 10SM FEW030 SCT070 SCT110 M09/M18 A2993 RMK AO2 SLP170 T10941178
METAR KDEN 201453Z 30009KT 10SM FEW030 SCT070 SCT110 M08/M17 A2994 RMK AO2 SLP165 T10781172 51014
METAR KDEN 201553Z 34008KT 10SM FEW070 SCT110 SCT220 M07/M17 A2993 RMK AO2 SLP161 T10671172
METAR KDEN 201653Z 32017G26KT 10SM FEW070 SCT110 SCT220 M06/M18 A2991 RMK AO2 PK WND 32026/1650 SLP157 T10561178
METAR KDEN 201753Z 31019G27KT 10SM FEW040 SCT110 SCT220 M04/M18 A2991 RMK AO2 PK WND 31030/1739 SLP161 T10391178 11039 21122 56007
METAR KDEN 201853Z 30024G31KT 10SM FEW040 SCT220 M03/M17 A2987 RMK AO2 PK WND 31036/1819 SLP141 T10331167
METAR KDEN 201953Z 31022G31KT 10SM FEW040 SCT220 M03/M17 A2985 RMK AO2 PK WND 30035/1859 SLP138 T10331167
METAR KDEN 202053Z 30021G30KT 10SM FEW046 SCT120 SCT220 M03/M17 A2985 RMK AO2 PK WND 31032/2019 SLP138 T10331167 56017
METAR KDEN 202153Z 30022G28KT 10SM FEW040 SCT120 SCT220 M04/M16 A2987 RMK AO2 PK WND 31029/2118 SLP147 T10391161
METAR KDEN 202253Z 30017G27KT 10SM FEW040 SCT120 SCT220 M04/M16 A2989 RMK AO2 PK WND 31027/2244 SLP160 T10441156
METAR KDEN 202353Z 29016G19KT 10SM FEW040 SCT120 M05/M16 A2993 RMK AO2 PK WND 29028/2308 SLP183 T10501161 11028 21050 53021
METAR KDEN 210053Z 28011KT 10SM FEW040 SCT100 M06/M16 A2997 RMK AO2 PK WND 29027/0000 SLP202 T10561161
SPECI KDEN 210134Z 29024G32KT 10SM FEW040 SCT100 M04/M18 A2998 RMK AO2 PK WND 28036/0123
METAR KDEN 210153Z 30018G28KT 10SM FEW040 SCT100 M05/M17 A3000 RMK AO2 PK WND 28036/0123 SLP202 T10501172
METAR KDEN 210253Z 29018KT 10SM FEW040 BKN100 M05/M18 A3002 RMK AO2 SLP205 T10501178 51027
METAR KDEN 210353Z 33012KT 10SM FEW040 SCT100 M06/M18 A3003 RMK AO2 SLP209 T10611183
METAR KDEN 210453Z 07010KT 10SM FEW040 SCT100 M12/M17 A3007 RMK AO2 SLP236 T11221172
METAR KDEN 210553Z 11011KT 10SM FEW040 SCT100 M14/M18 A3010 RMK AO2 SLP247 T11441183 11044 21144 53023
METAR KDEN 210653Z 13009KT 10SM FEW040 SCT100 M16/M19 A3012 RMK AO2 SLP258 T11561189 410281156
```

METAR logs example

2.2 Radar track dataset inspection

Dataset Information

- **2 years** of data (2014 - 2016)
- Radar track **+** Ground radar
- **Time series**
- Already **preprocessed**:
  - **Unknown data wrangling** and **interpolation** methodologies
- Duplicated callsign problem

Potential variables

Aircraft type

Flight level

Latitude

ICAO category

On runway

Longitude

Speed

2.3 Airport information dataset inspection

Dataset Information

- **2 years** of data (2014-2016)
- In message log format
- After message **aggregation** + **filtering**:

Total Flights	725.187
Total Arrivals	**364.583 (50.27%)**
Total Depatures	360.601 (49.73%)

Potential variables

Departure airport

AABT/AOBT

Arrival airport

Gate | ATA | ETA

Runway | ATOT

2.4 METAR

Dataset Information

- Message log
- Format **not fixed**→ Hard to parse
- Message **Filtering** to capture messages .
- **Main** meteo dataset.

Variables extracted

Ceiling	Wind direction	Wind speed	Temperature	QNH	1º Cloud layer visibility	Dew point
Minimum visibility	Wind variation	Runway status	Phenomenas	2º Cloud layer visibility		

2.5 SNOWTAM

Dataset Information

- Message log
- Only information for R11 and R16.
- But... R11 and R16 are R27 and R34 inverted
- Record **only during the "snow season"**

Potential Variables

Runway	Deposits	Depth	Apron clearance
Friction	Runway clearance		Taxiway clearance

2.5 Other meteo: WMA, Visibility and SODAR

Datasets Information

- **WMA**: Wind per runway. Higher frequency than METAR.
- **Visibility**: 4 cloud layers, higher frequency, more detailed than METAR.
- **SODAR:** Outliers, **high variance - high bias**. **Discarded** due to bad data. quality...

Potential Variables

WMA

Wind speed

Wind direction

Visibility

1º Cloud layer visibility

2º Cloud layer visibility

3º Cloud layer visibility

4º Cloud layer visibility

Minimum visibility

3. Questions and problem definition

Knowing
the **Data** available
and
the **problem**...

- What **research questions** can you raise?
- How well can your data **answer** them?

3. Questions and problem definition

RQ1: Can we predict if the aircraft going to take the procedural exit in R34?

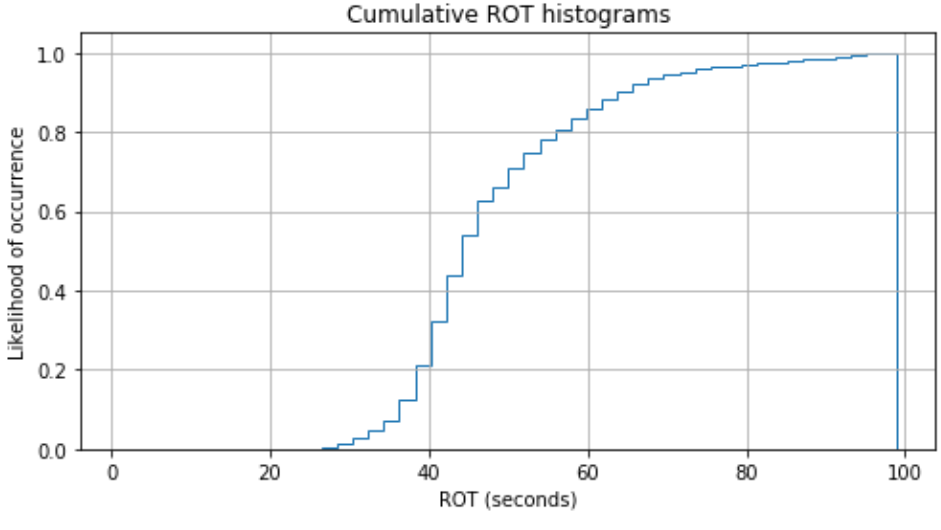
- Definition of the expected/procedural exit from AIP.
- Exit taken absent from the data, but can be extracted from the radar.
- Machine learning problem definition:
  - **Binary classification**
  - Target variable:
    - 1 procedural exit taken
    - 0 procedural exit not taken

Total arrivals R34	**59.369**
Taking the AIP exit	**43.933 (74%)**
Not taking the AIP exit	**15.436 (26%)**

3. Questions and problem definition

RQ2: Can we predict how much time will certain flight occupy the runway?

- **Runway Occupation Time (ROT)**: Time between coordinates of runway threshold and coordinates of exit.
- Ground movements with a second resolution, **bias** is an issue.
- Time error when over-passing threshold is not present in data...
- ML problem definition:
  - **Regression**
  - Target variable
    - **ROT**

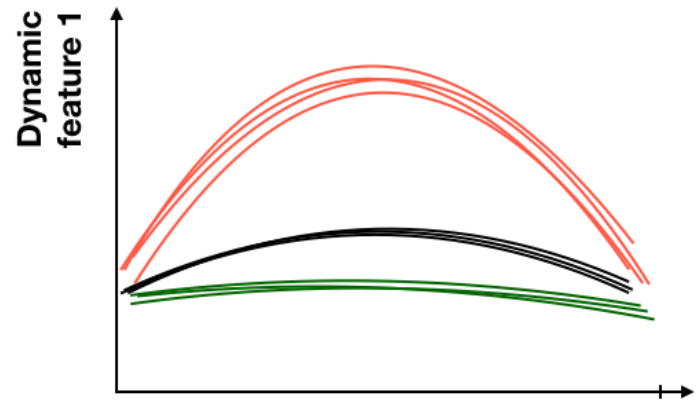


4. Features engineering

The **most important step** in the data pipeline!!

Static information

- **Unique information** per flight.
- Type of aircraft, state of the runway, etc...



Dynamic information

- **Time series** per flight: position, velocity, etc...
- More than 100 observations per flight and variable!!
- We need to **abstract** dynamic series **to single variables!! e.g. aircraft energy at 2NM**

4. Features engineering

Potential Precursors (based on ATCOs experience)

Velocity patterns

Weight / Aircraft type

Airlines protocols

Wind direction and speed

Visibility

Succeeding aircraft distance/speed at threshold

...

Engineered Features

Transversal - parallel wind decomposition

Visibility ranges

Aircraft energy

Velocity at X NM

Velocity slope

Largest break

Wet runway indicator

Delay

...

5. ML algorithm: Introduction to Boosting Frameworks

- Also known as **Gradient Boosting Machines (GBMs).**
- Currently one of the most popular solutions for building predictive models.
- Methodology:

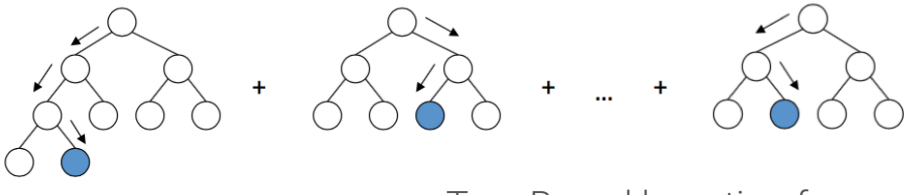
Ensemble ML model:

Additive model of many weak learners (typically **regression trees**)

→ The model learns **iteratively**

→ Learns only from data samples that were **"difficult" to learn** in the previous iteration

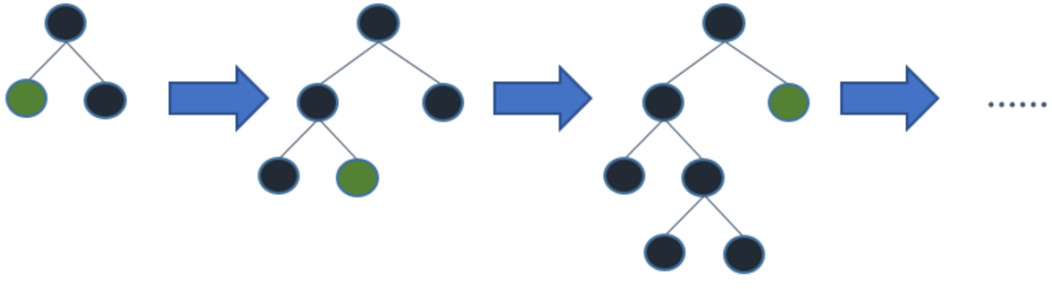
→ Huge **reduction** of **bias** and **variance** while **maximizing accuracy**



Tree-Based boosting framework

5. ML algorithm: LightGBM

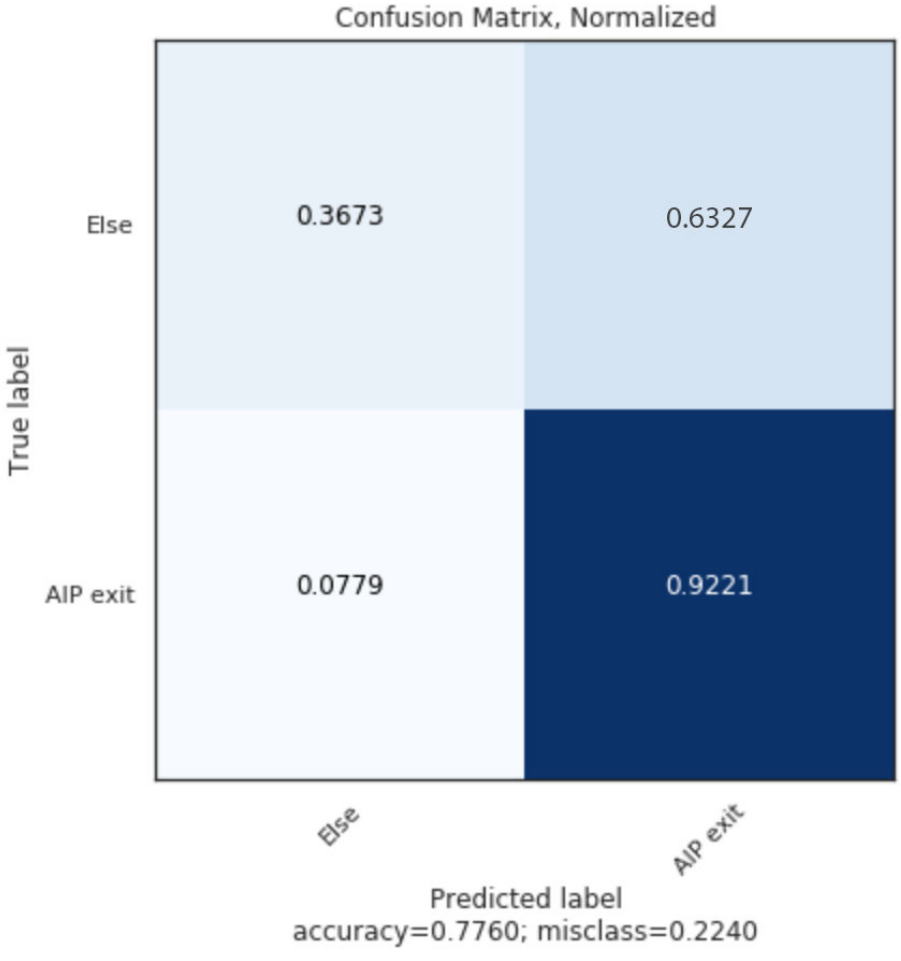
- Light Gradient Boosting Machine **(LightGBM)**
- **Gradient boosting** framework that uses **tree-based** learning algorithms.
- Histogram based algorithm - aggregates continuous features into discrete bins - speeds up training and reduces memory usage.
- Grows the trees **leaf-wise**: **can reduce even more the loss** than a level-wise algorithm (e.g. XGBoost).
- Can be used for both **classification** and **regression**.



Leaf-Wise tree growth

6.1 Results: Understanding the classifier

- **78% accuracy** (AUC)
- **92%** when the flights are taking the designated exit.
- **36%** when the flights are not taking the designated exit.
- The classifier is not good at identifying the **abnormal behaviours**...



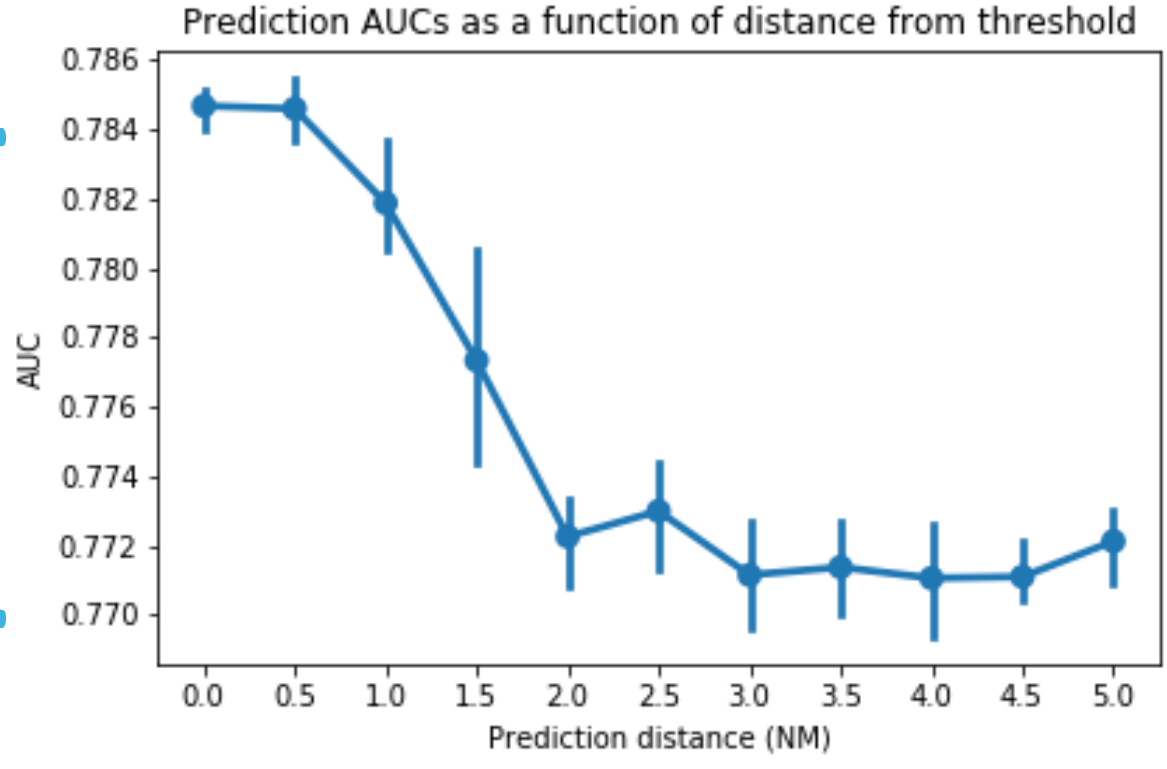
6.1 Results: Understanding the classifier

- Prediction error **slowly decays with the distance** from threshold.
- **Most of the features** relevant for the prediction are **known beforehand.**

Decay

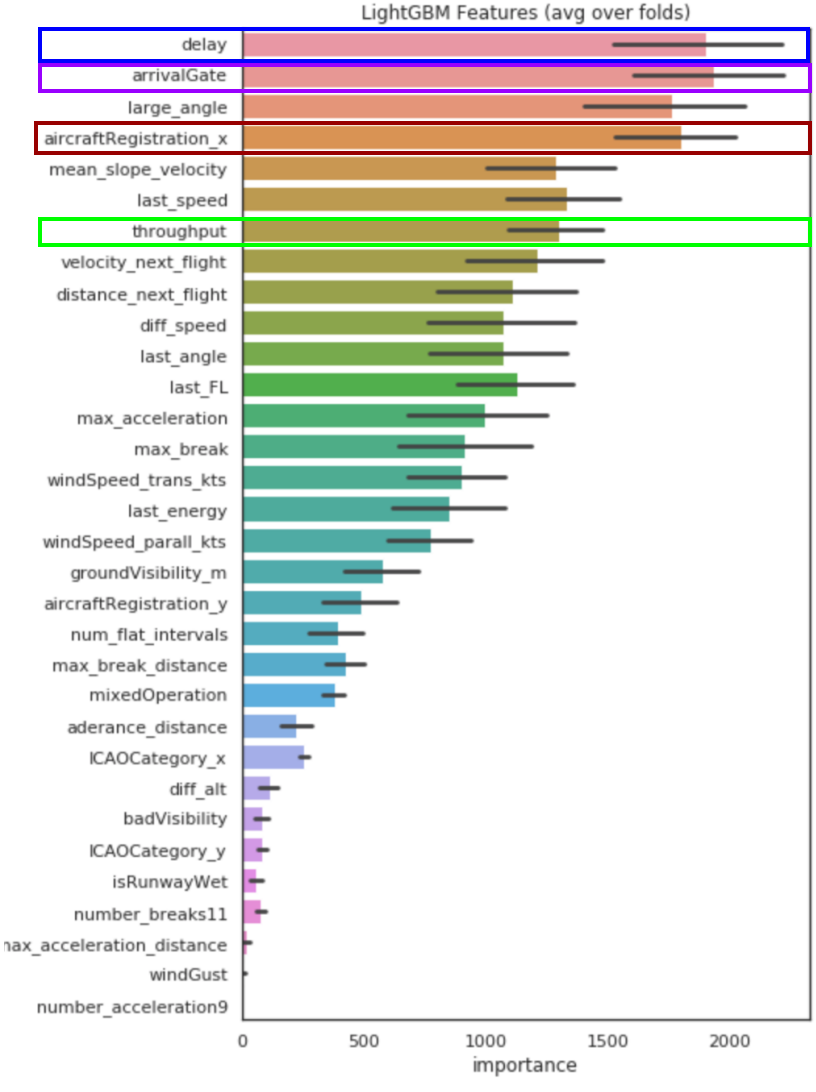
of

1,4%

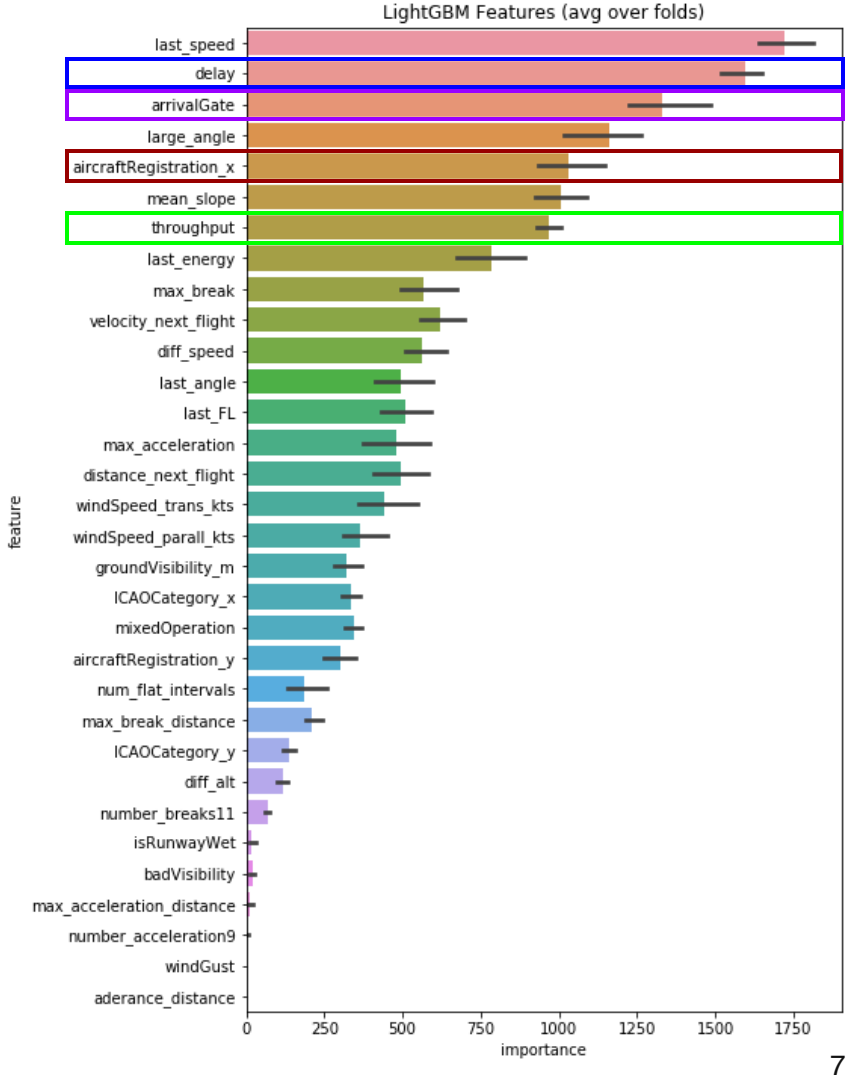


6.2 Features importance / Precursors analysis

Prediction at 2NM

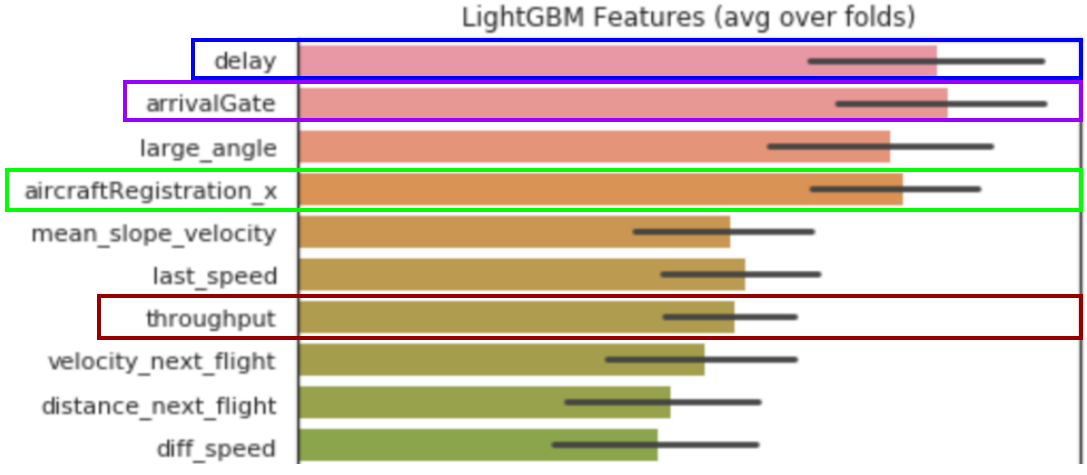


Prediction at threshold

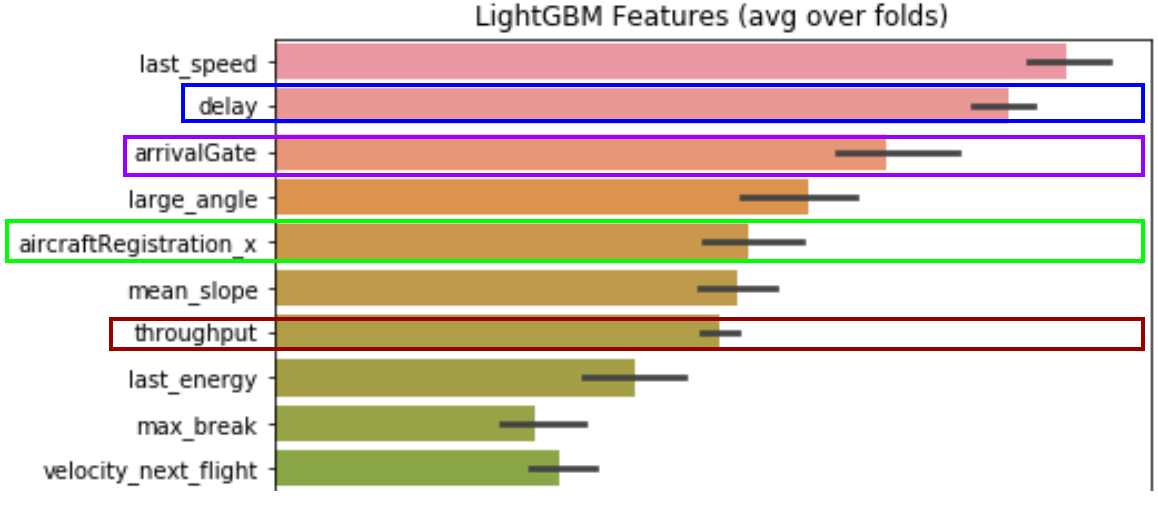


6.2 Features importance / Precursors analysis

Prediction at 2NM

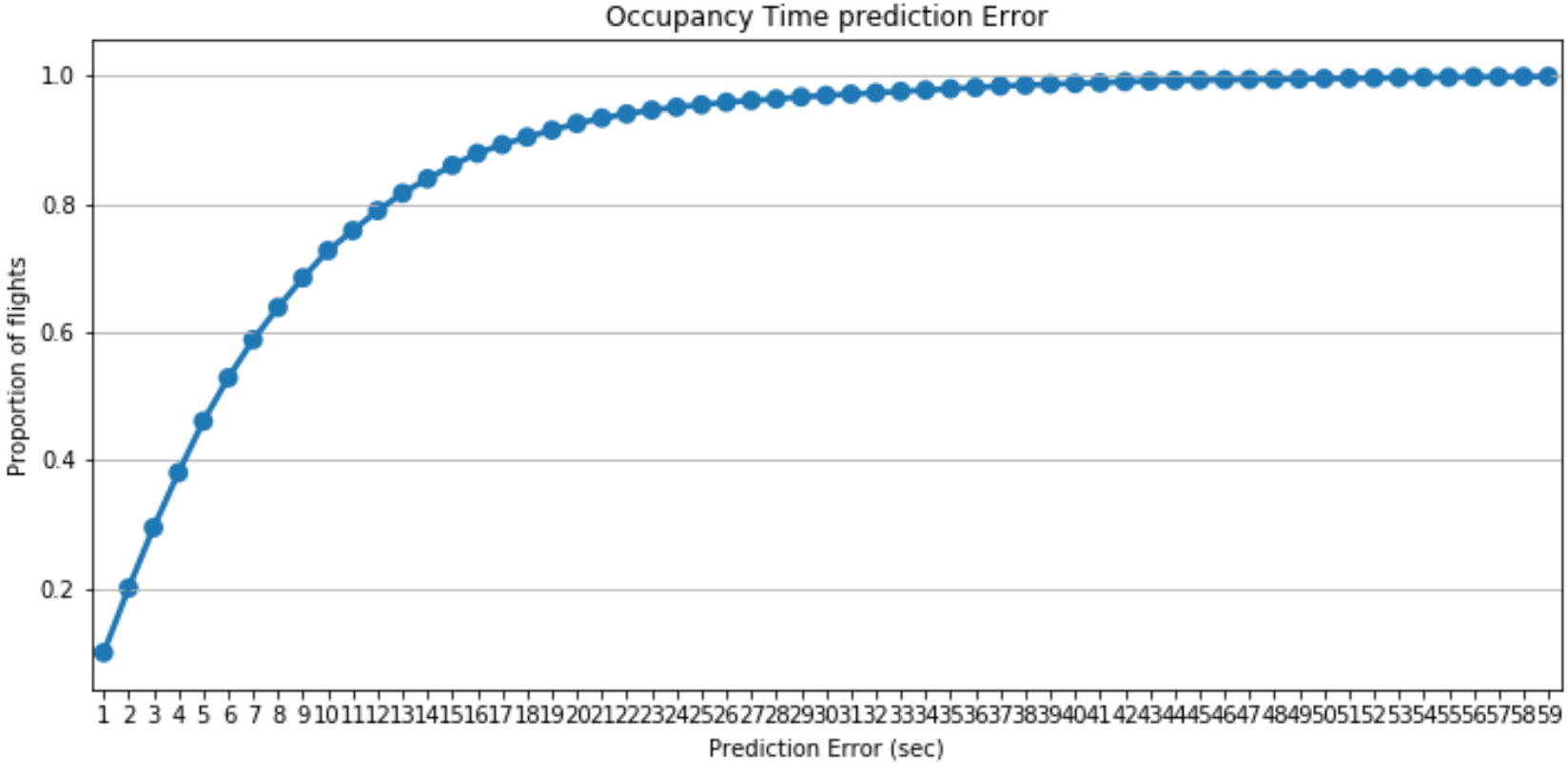


Prediction at threshold



7. Results: Understanding the regressor

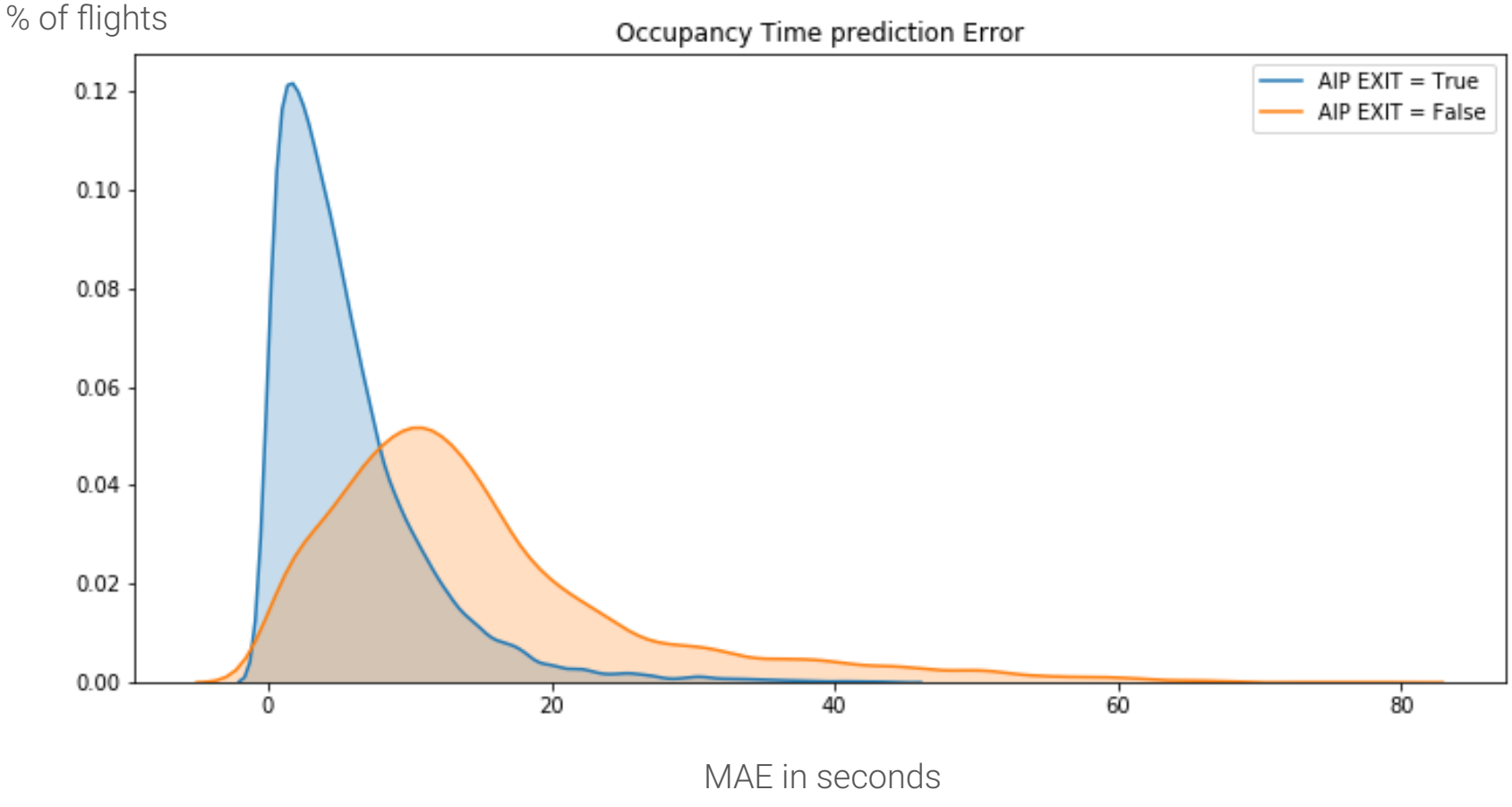
- Mean ROT **49.66 seconds** with a standard deviation of 14.4 seconds.
- **80% of the ROTs** are predicted with **< 14 seconds of error**.



7. Results: Understanding the regressor

The regressor experiments the same limitations as the classifier:

fails at identifying the flights not behaving correctly.



Thank you!

Darío Martínez

dm@innaxis.org

www.linkedin.com/in/dmartr

SafeClouds.eu