AI for Weather Forecasting

Development of BI and AI on METAR and TAF data

Riccardo Patriarca, PhD
Sapienza University of Rome - ITALY
Dept. of Mechanical and Aerospace Engineering

FLY AI Forum
Brussels, 20-21 April 2023

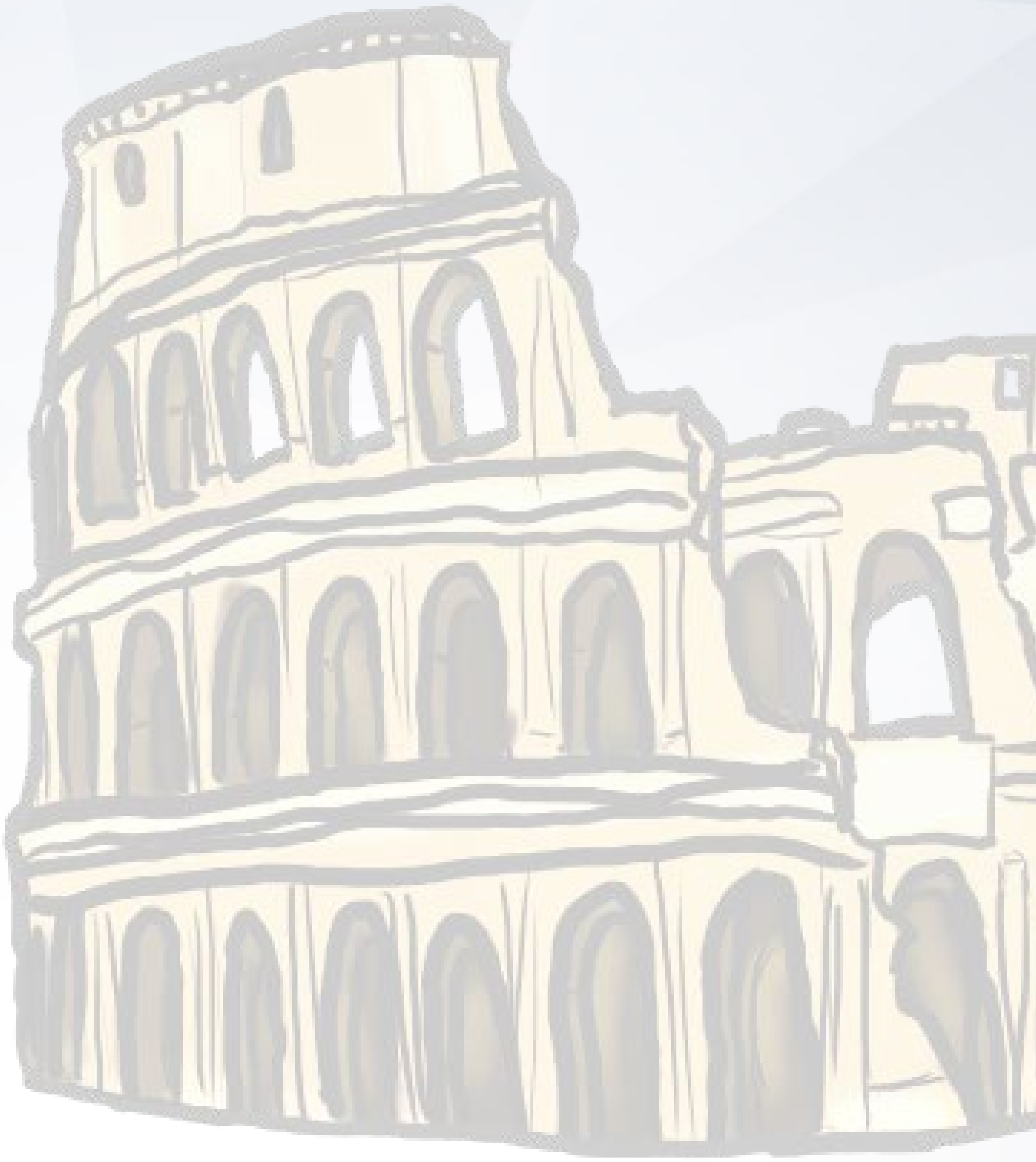
Project Introduction

Sapienza is the oldest university in Rome (founded **1303**), and the largest in Europe: **120,000** students, **3,500** professors.

The **Department of Mechanical and Aerospace Engineering** includes research groups specialized in data science applied to diverse engineering problems.

AGIC aiComply is the Department's technological spin-off, specialized in data intelligence, cloud technologies risk and governance management solutions

Today presentation:
Using advanced Business Intelligence and Artificial Intelligence solutions for <u>weather forecasting management</u>



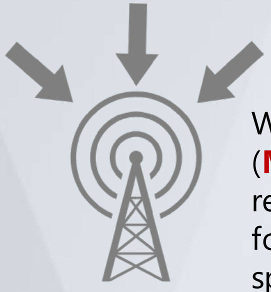
Disclaimer

The following slides are based on publicly available content, as described in the following papers:

- *Patriarca, R., Simone, F., Di Gravio, G. Supporting weather forecasting performance management at aerodromes through anomaly detection and hierarchical clustering (2023) Expert Systems with Applications, 213, art. no. 119210. DOI: 10.1016/j.eswa.2022.119210*
- *Simone, F., Di Gravio, G., Patriarca, R. Performance-based Analysis of Aerodrome Weather Forecasts (2022) New Trends in Civil Aviation, 2022-October, pp. 27-33. DOI: 10.23919/NTCA55899.2022.9934004*

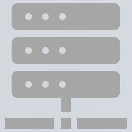
Project Overview

BACKGROUND



Weather bulletins (**METAR**) are emitted at regular frequency + forecasts (**TAF**) and special bulletins **SPECI**

Dedicated **KPIs** defined to capture forecast accuracy



Big-data on text strings of bulletins to be **decoded** first, and then **analysed** systematically

NEED

600'000 METAR/SPECI
60'000 TAF
(generic yearly figure)

KPI to be developed for ***retrospective*** *analyses and to generate* ***proactive*** *indicators*

SOLUTION

User-friendly **systemic dashboards** for the WSP (top management and operational)

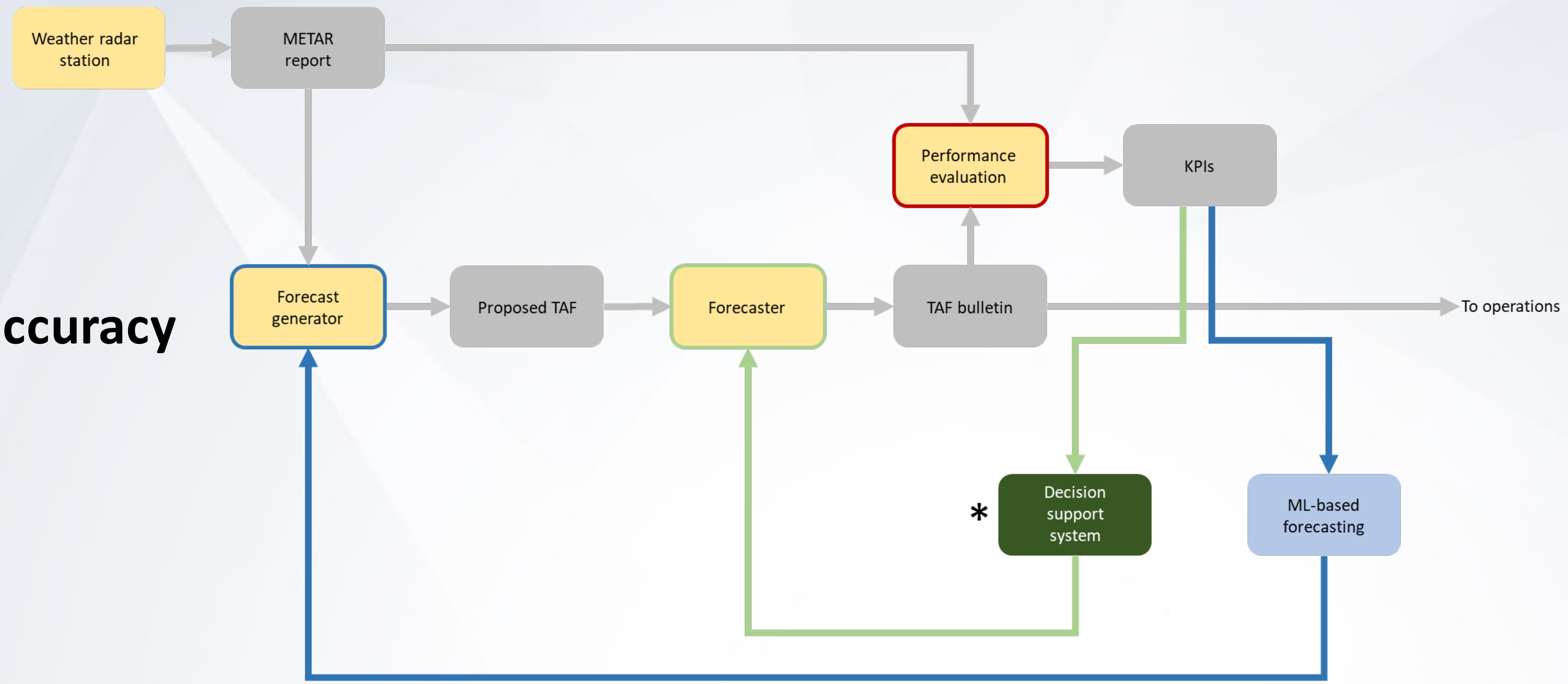
ML analysis on KPIs

Customized periodic **automatic reporting**

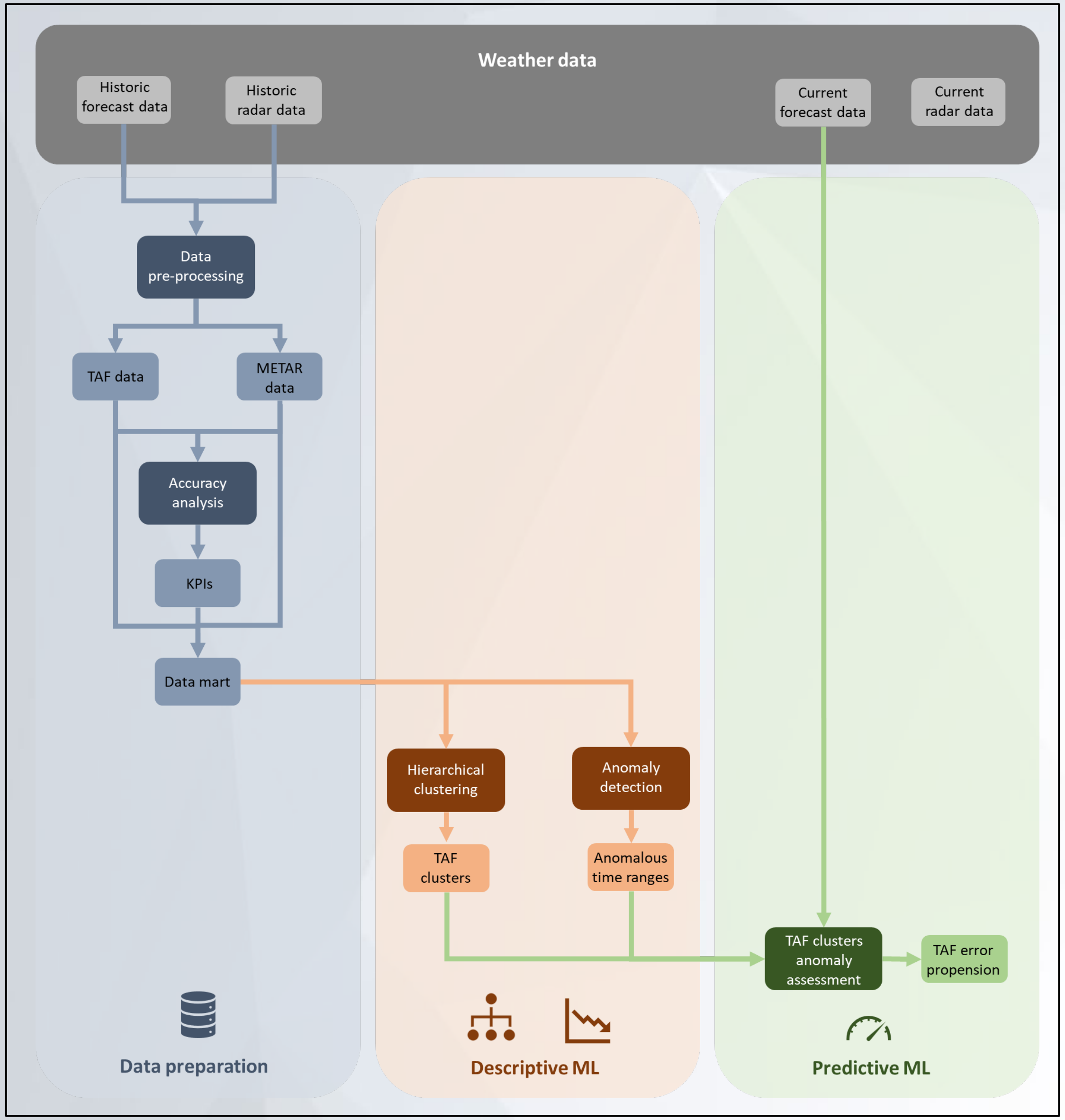
Project Aim

The project aims to:

- ~~Forecast weather~~
- **Manage forecast accuracy**



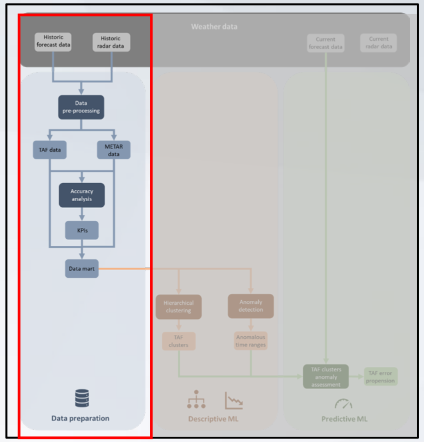
AI Pipeline



AI pipeline passes through **three** stages:

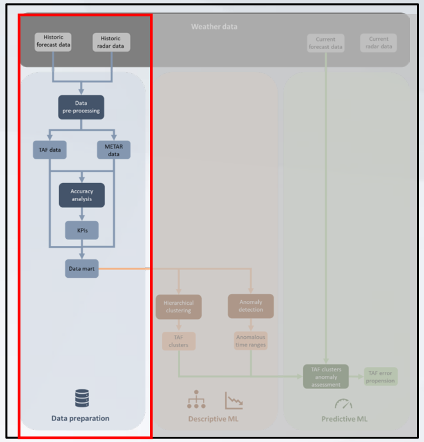
- **Data Preparation**
- **Descriptive ML**
- **Predictive ML**

Data Preparation: Decoding Weather Data



METAR	UTC DATE/TIME
LTFM 241550Z 02022KT 0800 R35R/0800D R17L/P1500N R34L/P1500N R36/P1500N SHSN BLSN SCT002 FEW017CB BKN022 M00/M00 Q1025 RETSSN...	2022-01-24 15:50:00
LTFM 241520Z 36027G37KT 0800 R35R/0500D R17L/P1500D R34L/0650D R36/P1500N +TSSN BLSN SCT002 SCT017CB BKN022 M00/M00 Q1025 RETS...	2022-01-24 15:20:00
[illegible] 0 Q1024 RESHSN BE...	2022-01-24 14:50:00
[illegible] 22 M01/M01 Q1024 ...	2022-01-24 13:50:00
[illegible] 1/M01 Q1024 BECM...	2022-01-24 13:20:00
LTFM 241250Z 36023KT 0300 R35R/0 [illegible] 0350D R36/0300D +SHSN BLSN SCT002 SCT017CB BKN022 M00/M00 Q1024 RETSSN BE...	2022-01-24 12:50:00
LTFM 241220Z 35022G3 [illegible] 5N R36/0600D +TSSN BLSN SCT002 SCT017CB BKN022 M00/M00 Q1024 RESHS...	2022-01-24 12:20:00
LTFM 241150Z [illegible] 4L/0225D R36/1400U SHSN BLSN SCT002 SCT017CB BKN022 M01/M01 Q1024 ...	2022-01-24 11:50:00
LTFM [illegible] 4L/0225N R36/0200N +TSSN BLSN SCT002 SCT017CB BKN022 M01/M02 Q1024 RESHS...	2022-01-24 11:20:00
[illegible] 0175N R36/0225N +SHSN BLSN SCT002 SCT017CB BKN022 M01/M02 Q1024 TEMP...	2022-01-24 10:50:00
[illegible] L/0325N R36/0300N +SHSN BLSN SCT003 SCT017CB BKN022 M01/M02 Q1024 BECM...	2022-01-24 10:20:00
[illegible] 0275N R34L/0375N R36/0375N +SHSN SCT003 SCT017CB BKN022 M01/M02 Q1024 RETSSN BEC...	2022-01-24 09:50:00
[illegible] 150 [illegible] /P1500U R34L/1400U R36/P1500N -SHSN FEW005 FEW018CB BKN025 M01/M01 Q1024 RES...	2022-01-24 09:20:00
[illegible] R [illegible] Q1024 TEMPO TL1000 2000 -TSSN	2022-01-24 08:50:00
[illegible] N025 M00/M02 Q1024 RESHSN TEMPO TL1000 2000 -TSSN	2022-01-24 08:20:00

Data Preparation: Decoding Weather Data



AERODROME FORECAST – TAF DECODE

IDENTIFICATION GROUPS

TAF or TAF AMD or TAF COR — Code names for aerodrome forecast, amended aerodrome forecast and corrected aerodrome forecast, respectively

CCCC — ICAO four-letter location indicator

YYGGggZ — Date and time of issue of forecast in UTC; Indicator of UTC

NIL — Indicator of missing forecast

$Y_1Y_1G_1G_1/Y_2Y_2G_2G_2$ — Period of validity, beginning on Y_1Y_1 day of month at G_1G_1 (UTC) and ending on Y_2Y_2 day of month at G_2G_2 (UTC)

CNL — Indicator of cancelled forecast

FORECAST SURFACE WIND

$dddffGf_mf_m$ KMH or KT or MPS — Mean wind direction in degrees true rounded to nearest 10° (VRB = VaRiaBle when ff < 3 kt); Mean wind speed; Indicator of Gust; Maximum wind speed (gust); Wind speed units used

00000 = calm

P199 KMH (P99 KT, P49 MPS) mean f_mf_m = 200 KMH (100 KT, 50 MPS) or more

FORECAST PREVAILING VISIBILITY

VVVV — Prevailing visibility in metres; 9999 = 10 km or more

FORECAST SIGNIFICANT WEATHER

w'w' — Forecast significant weather (see table w'w' for METAR/SPECI decode)

Replaced when significant weather phenomenon forecast to end by: NSW — Nil Significant Weather

FORECAST CLOUD AMOUNT AND HEIGHT[a]

$N_sN_sN_sh_sh_sh_s$(cc) — Cloud amount: FEW – FEW (1–2 oktas); SCT – SCaTtered (3–4 oktas); BKN – BroKeN (5–7 oktas); OVC – OVerCast (8 oktas). Height of base of cloud in units of 30 m (100 ft). Cloud type – only CB (cumulonimbus) is indicated

Replaced when sky is expected to be obscured and information on vertical visibility is available by: $VVh_sh_sh_s$ — Indicator of Vertical Visibility; Vertical visibility in units of 30 m (100 ft)

Replaced when no CB and no cloud below 1 500 m (5 000 ft) or below the highest minimum sector altitude, whichever is greater, is forecast and CAVOK is not appropriate by: NSC — Nil Significant Cloud

CAVOK

Ceiling And Visibility OK. Replaces visibility, weather and cloud if: (1) Visibility is forecast to be 10 km or more (2) No cumulonimbus cloud and no other cloud forecast below 1 500 m (5 000 ft) or below the highest minimum sector altitude, whichever is greater, and (3) No significant weather forecast (see table overleaf)

SIGNIFICANT CHANGES IN FORECAST CONDITIONS INDICATED BY:

PROBABILITY — $PROBC_2C_2$ — PROBability; Only 30 or 40 used, indicating 30% or 40%

DATE AND TIME — $YYGG/Y_eY_eG_eG_e$ — Beginning day and time (UTC) YYGG and end day and time (UTC) $Y_eY_eG_eG_e$ of forecast period

Probability is used to indicate the probability of occurence of: (a) an alternative element or elements (b) temporary fluctuations

CHANGE — TTTTT — Type of significant change: BECMG – BECoMinG, used where changes are expected to reach or pass through specified values at a regular or irregular rate; TEMPO – TEMPOrary fluctuations of less than 1 h and in aggregate less than half the period indicated by $YYGG/Y_eY_eG_eG_e$

DATE AND TIME — $YYGG/Y_eY_eG_eG_e$ — Beginning day and time (UTC) YYGG and end day and time (UTC) $Y_eY_eG_eG_e$ of forecast period

OR

If one set of weather conditions is expected to change more or less completely to a different set of conditions, thus indicating the beginning of another self-contained part of the forecast, by:

TTYYGGgg

This takes the form FMYYGGgg where FM is the abbreviation for FroM and YYGGgg is the date and time in hours and minutes UTC. All forecast conditions before this group are superseded by conditions indicated after the group

BY REGIONAL AGREEMENT

FORECAST TEMPERATURE

$(TXT_FT_F/Y_FY_FG_FG_FZ$ $TNT_FT_F/Y_FY_FG_FG_FZ)$

TX, TN Indicators of maximum and minimum forecast temperatures, respectively

T_FT_F Date and forecast temperature at G_FG_F. Temperatures below 0°C preceded by M

$Y_FY_FG_FG_F$ Date and time UTC to which forecast temperature refers

Z Indicator of UTC

November 2008

[a] Clouds of operational significance (i.e. below 1 500 m (5 000 ft) or below highest minimum sector altitude, whichever is greater, and CB or TCU)

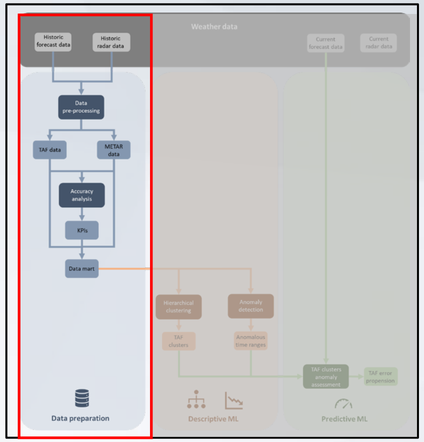
Note: Meteorological offices to issue 30-h TAFs are determined by regional air navigation agreement.

Abbreviated decode of TAF. For details of codes, see WMO Manual on Codes (WMO-No. 306)

WORLD METEOROLOGICAL ORGANIZATION
Weather • Climate • Water

WMO decoding instructions
+
Local decoding practices

Data Preparation: Defining accuracy KPIs



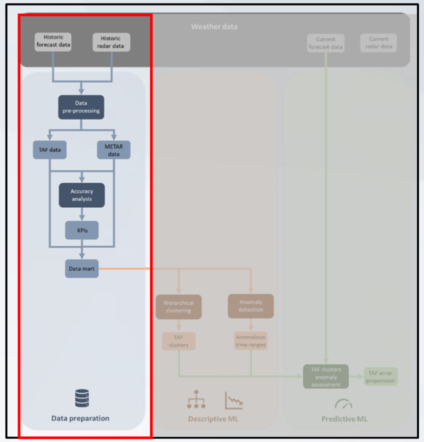
Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

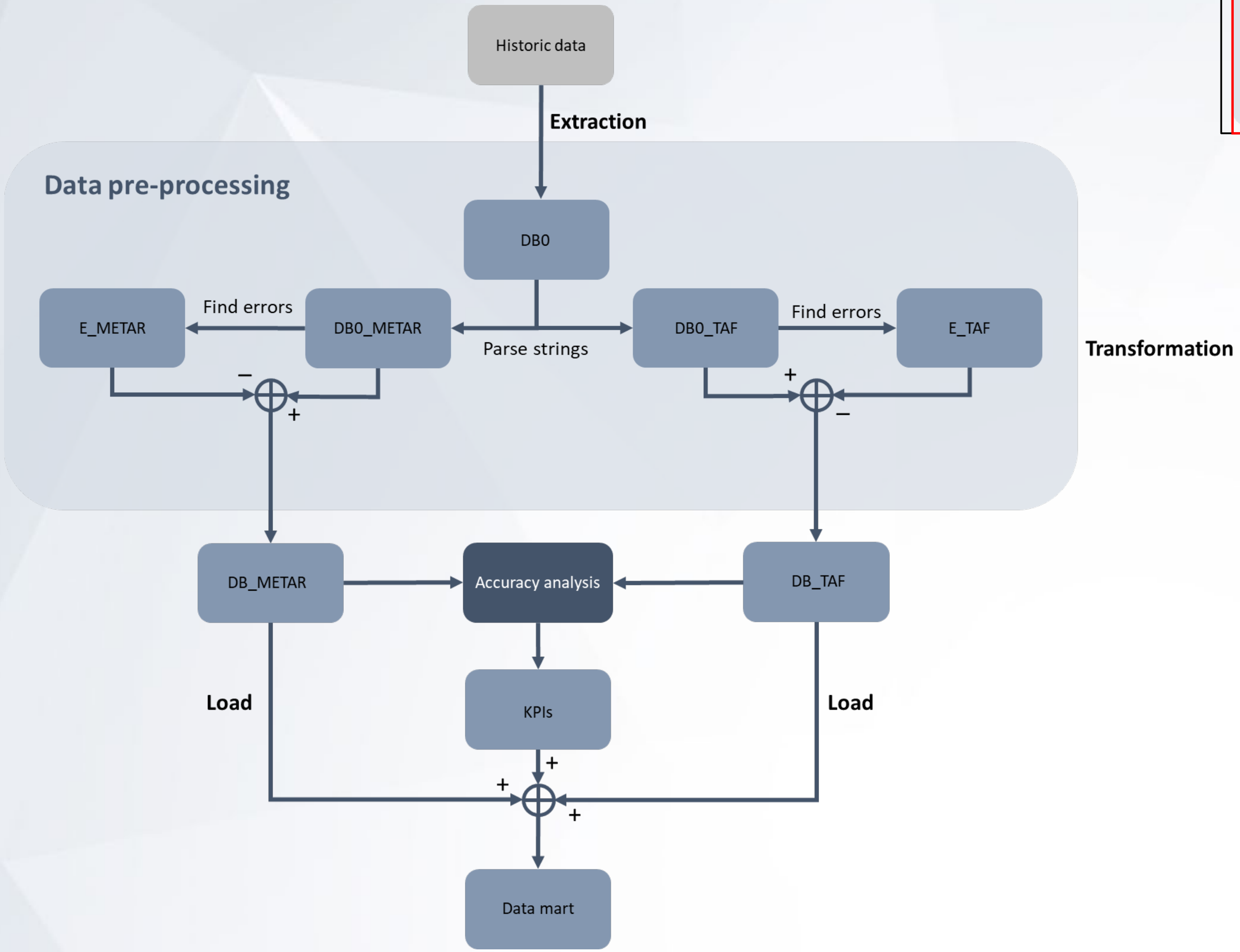
Contingency matrixes for binary parameters, which can be generalized for multi-variate parameters.

KPI	Acronym	Analytical expression
Frequency Bias Index	FBI	$\mathrm{FBI} = \frac{a+b}{a+c}$ (1)
Proportion Correct	PC	$\mathrm{PC} = \frac{a+d}{a+b+c+d}$ (2)
Critical Success Index	CSI	$\mathrm{CSI} = \frac{a}{a+b+c}$ (3)
Probability Of Detection	POD	$\mathrm{POD} = \frac{a}{a+c}$ (4)
False Alarm Ratio	FAR	$\mathrm{FAR} = \frac{b}{a+b}$ (5)

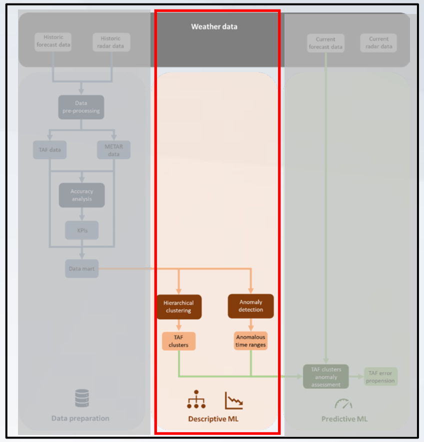
Key Performance Indicators (KPIs) can be calculated accordingly.

Data Preparation: Developing the Data Mart





Descriptive ML: Clustering similar forecasts (TAFs)



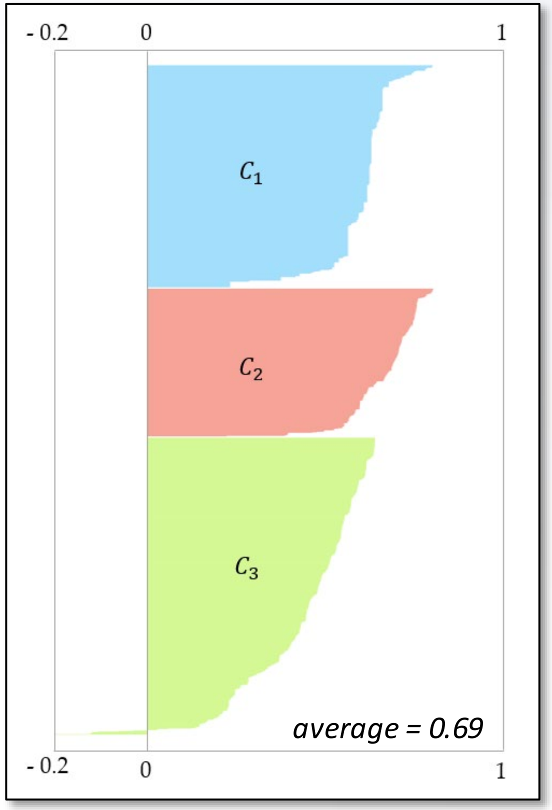
Hierarchical Clustering to spot TAFs characterized by "common" aspects

The hierarchy dendogram is obtained via the cosine distance in the M-dimensional space

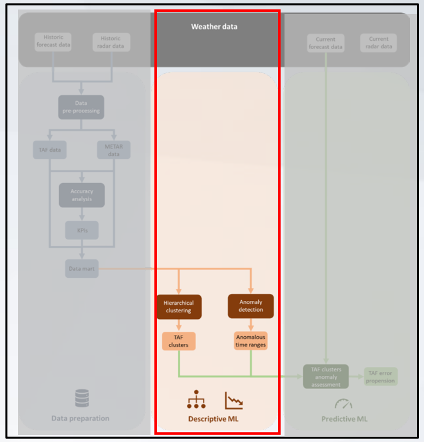
$$d_{o_i o_j} = 1 - \cos\left(\theta_{o_i o_j}\right) = 1 - \frac{o_i \cdot o_j}{\|o_i\| \, \|o_j\|} = 1 - \frac{\sum_{m=1}^{M} o_{im}\, o_{jm}}{\sqrt{\sum_{m=1}^{M} {o_{im}}^2 \cdot \sum_{m=1}^{M} {o_{jm}}^2}}$$

and the Ward linkage criterion (via the Lance-Williams recursive algorithm)

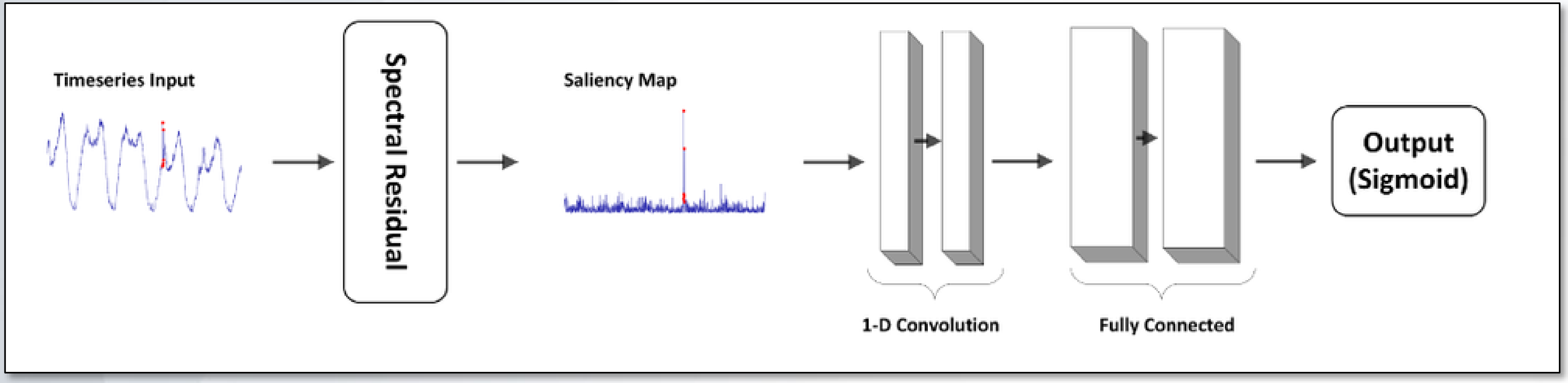
$$d_{(C_I \cup C_J) C_K} = \alpha_{C_I} d_{C_I C_K} + \alpha_{C_J} d_{C_J C_K} + \beta d_{C_I C_J} + \gamma \left| d_{C_I C_K} - d_{C_J C_K} \right|$$



Descriptive ML: Identifying anomalies



Anomaly detection to spot outliers in performance for KPIs

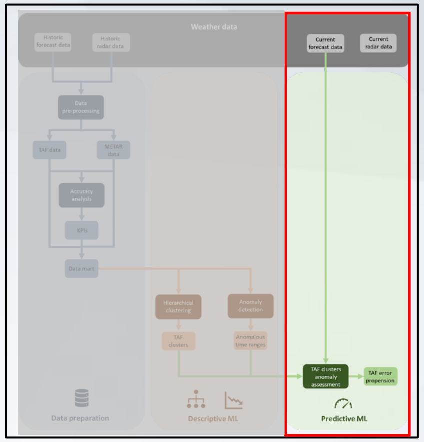


A Spectral Residual, Convolutional Neural Network is applied

$$S(x) = \left\|\mathcal{F}^{-1}(\exp(SR(f) + \sqrt{-1} \cdot P(f)))\right\|$$

$$\mathcal{A} = \{ TAF_u : u = u_{t*}, t^* \mid V_u^e \leq t^* \leq V_u^s \}$$

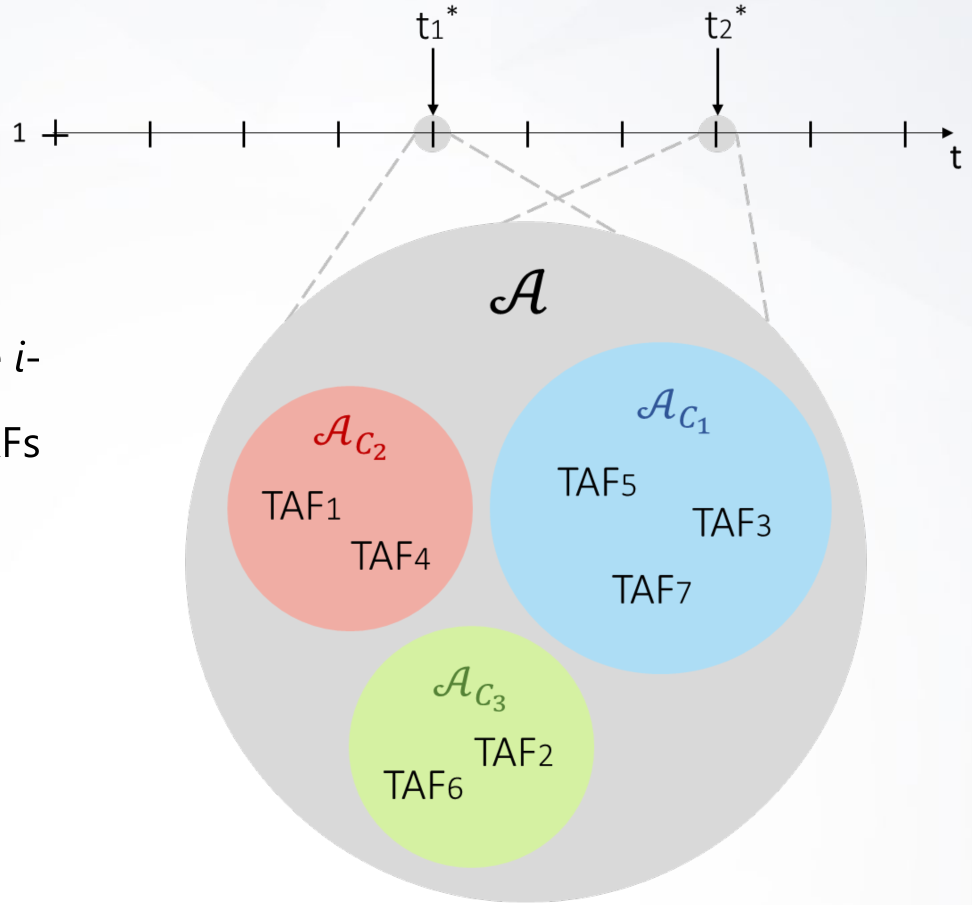
Predictive ML: Assign forecast error propensity



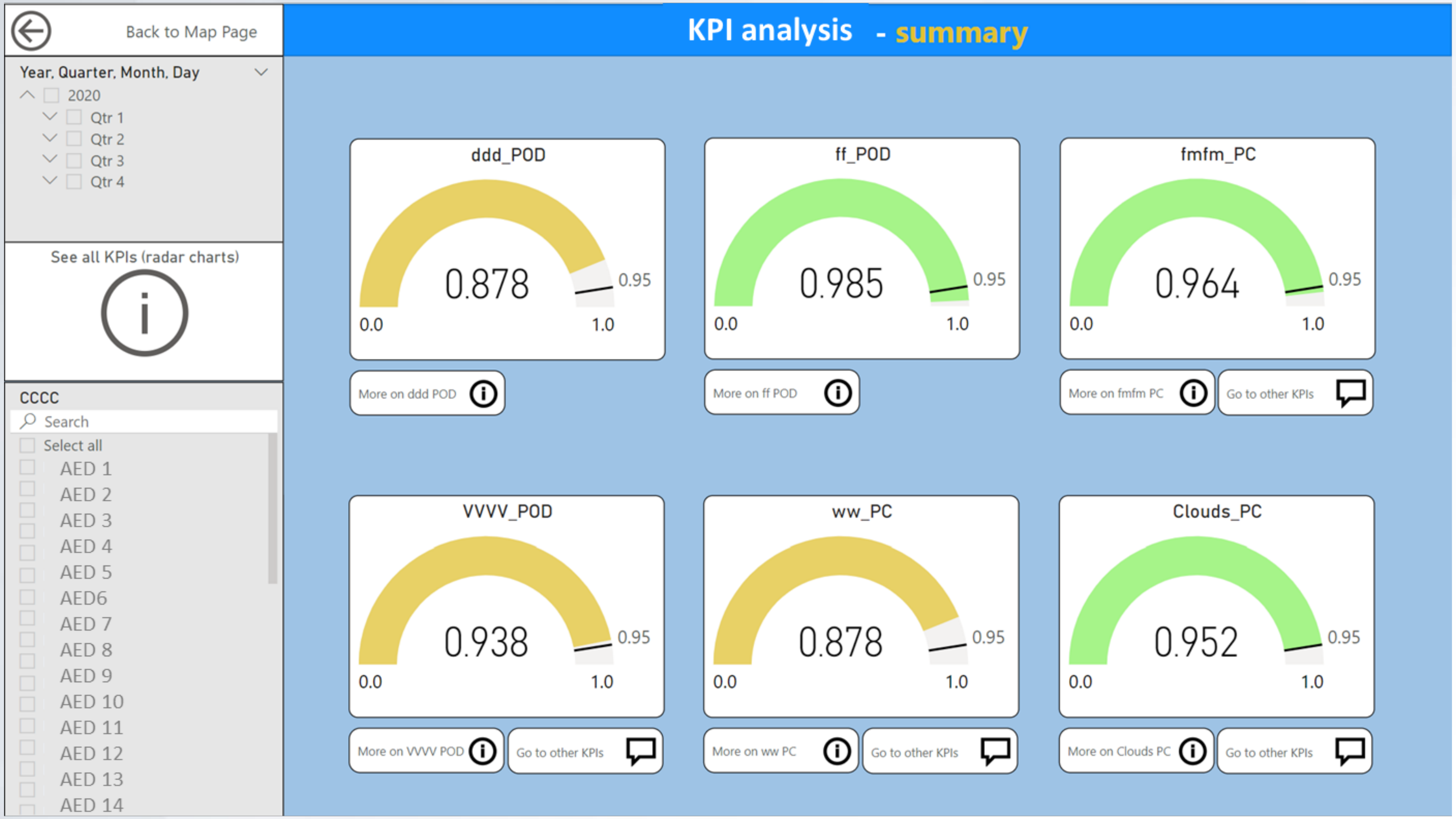
$$\mathcal{A}_{C_i} = \{\, TAF_u \mid TAF_u \in \mathcal{A} \ \wedge \ TAF_u \ is\ classified\ in\ C_i \,\} \qquad i = 1, \dots, Nc$$

$$\eta_{C_i} = \frac{|\mathcal{A}_{C_i}|}{s_{C_i}} \qquad i = 1, \dots, Nc$$

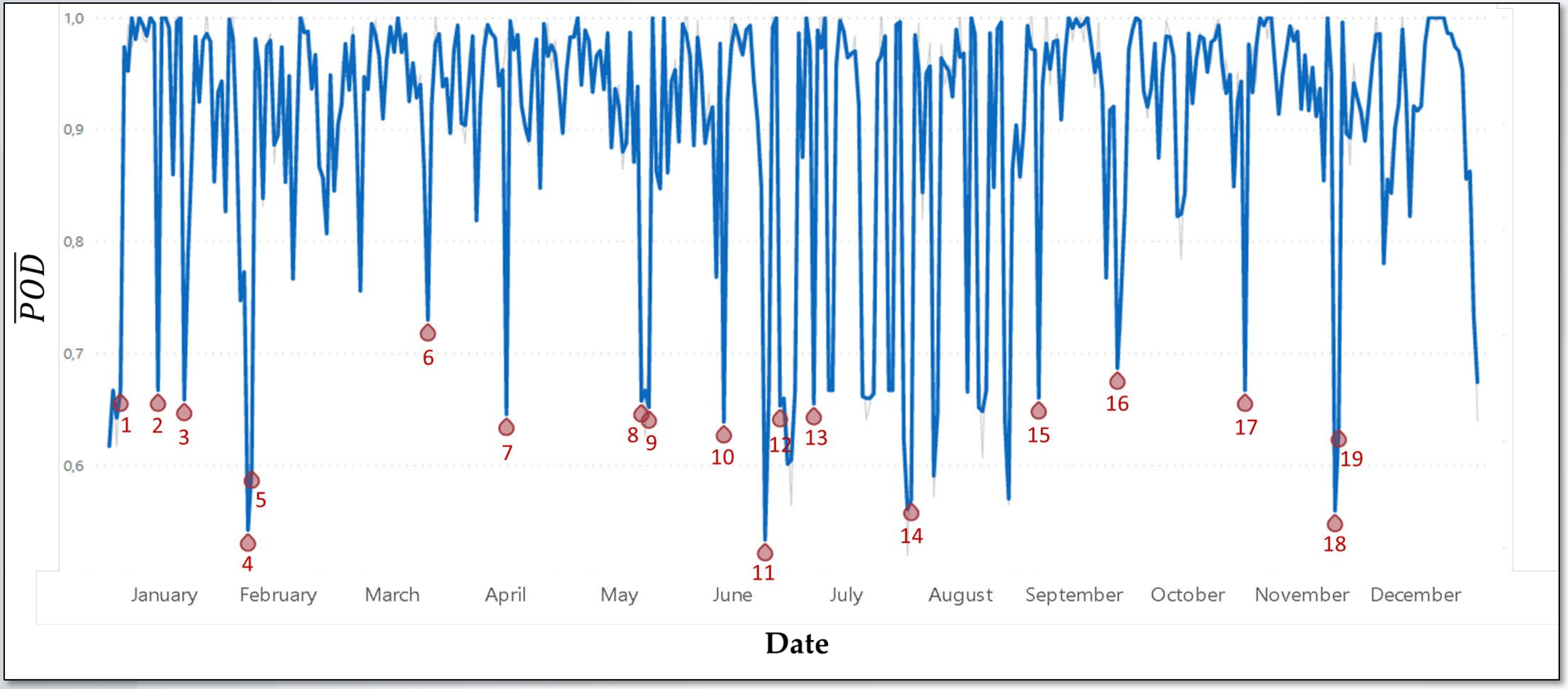
where $|\mathcal{A}_{C_i}|$ is the cardinality of C (i.e., number of anomalous TAFs in the i-th cluster); and s_{C_i} is the size of C_i based on the whole set of historic TAFs obtained from the clustering algorithm.



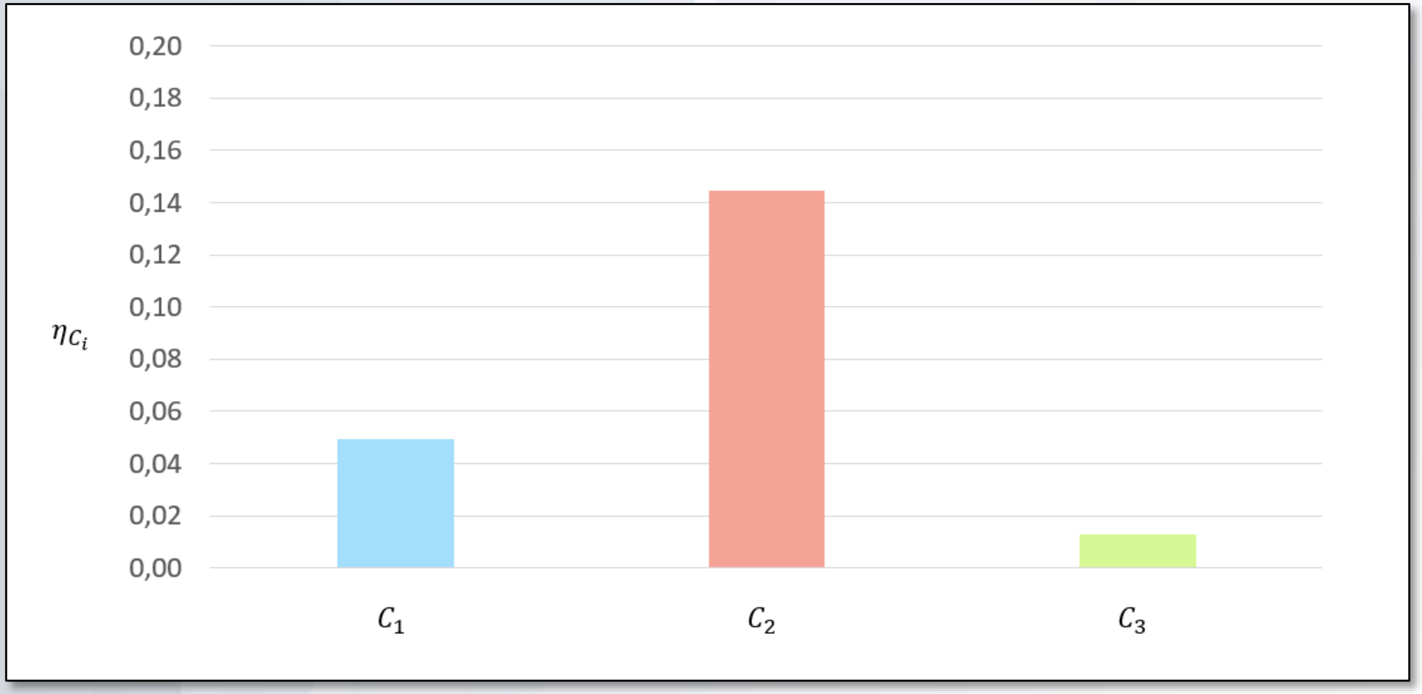
Use case: Airport X

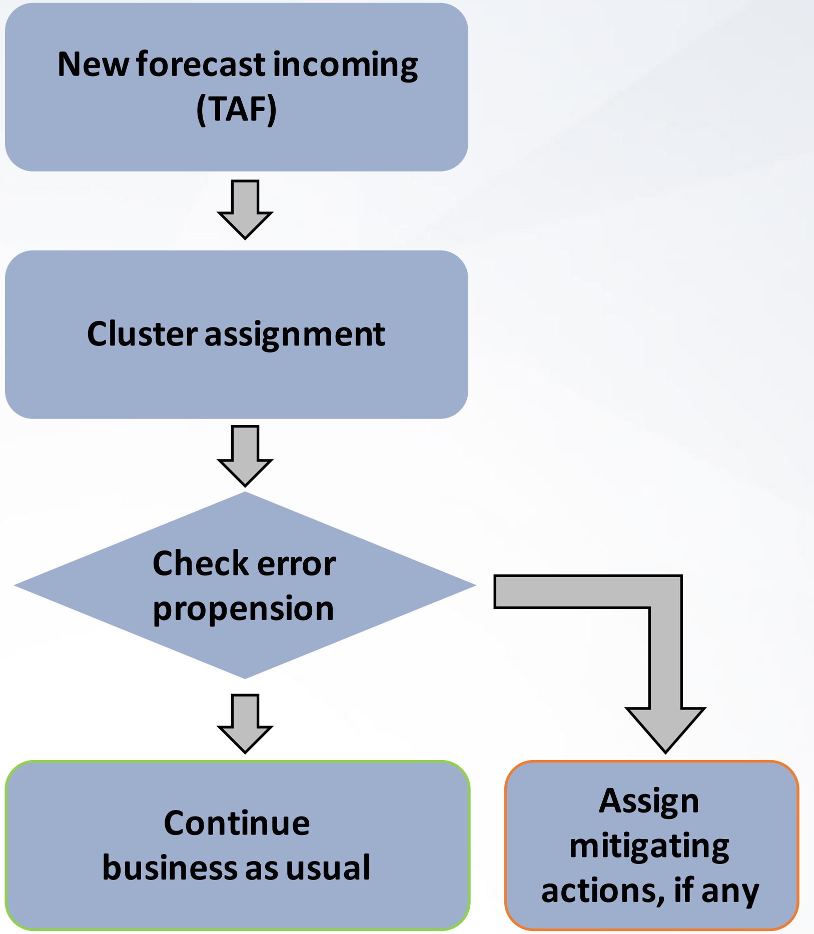


Use case: Airport X

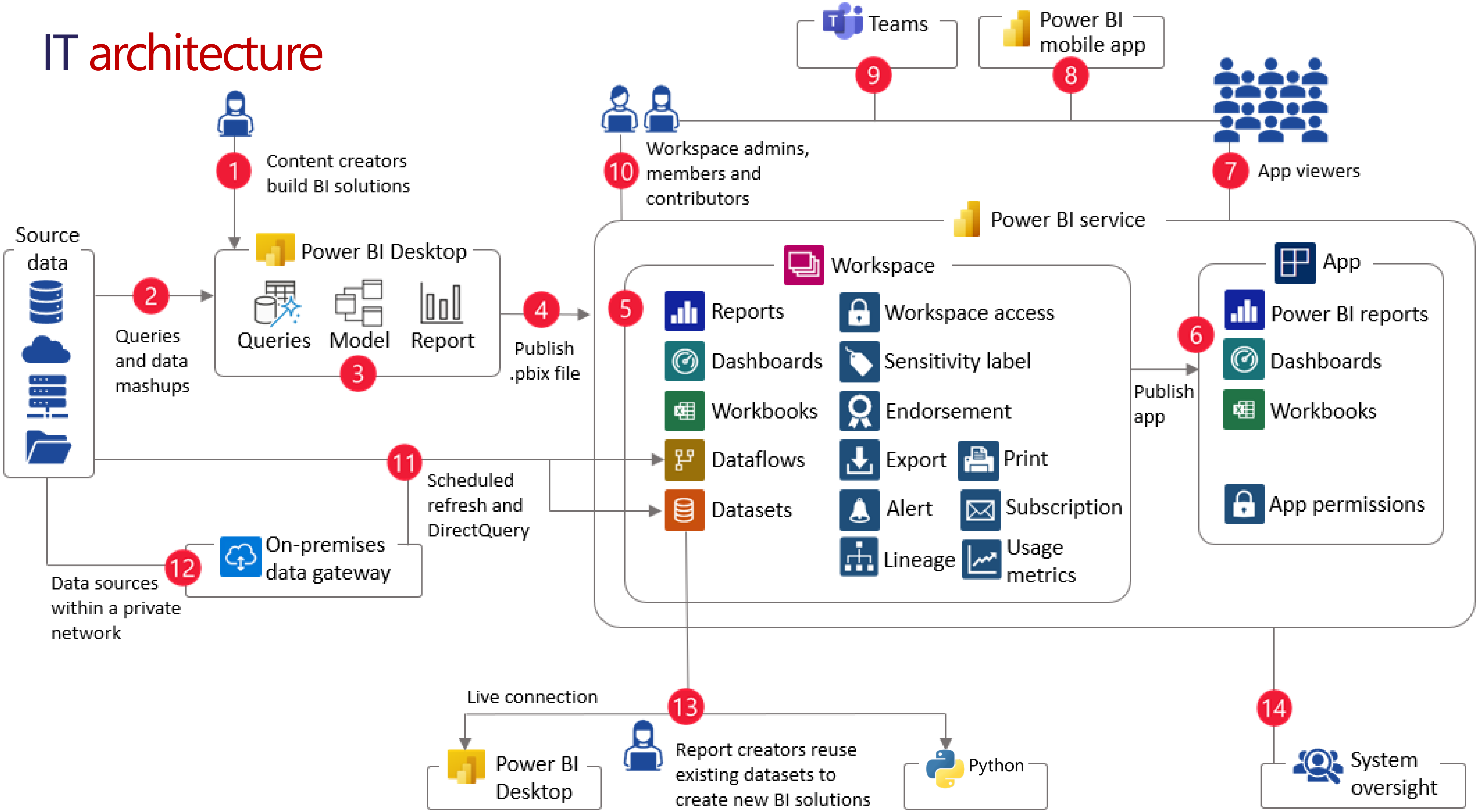


Use case: Airport X





IT architecture



Advantages

01

Integration in enterprise database

02

Row-level security for each data field (each user can see only their own data), using a PBI link to be provided

03

Users can navigate data **live**, to do **self-service reporting*, export** in PDF/PPT, etc.

AI for Weather Forecasting

Development of BI and AI on METAR and TAF data

Riccardo Patriarca, PhD
Sapienza University of Rome - ITALY
Dept. of Mechanical and Aerospace Engineering