PREDICTING AND PREVENTING ROTATIONAL DELAYS OF AIRCRAFT

ALISHA RIECKER, LONGTAO LIU, MICHAEL SUHENDRA, LENNART JAYASURIYA, KHADER MOHAMMAD
TUM DATA INNOVATION LAB
MUNICH, 25.02.2022

AGENDA

THE DELAY OF FLIGHTS IS A SEVERE PRODUCTIVITY INHIBITOR THAT COSTS A LOT OF MONEY FOR AIRLINES

Problem & Motivation

- Aircraft are among the most expensive assets for an airline
  → Aircraft productivity should be as high as possible (blockhours/day)

- Delay of one flight influences punctuality of many other flights
  - High number of influencing variables
  - Lufthansa has steadily increased knowledge about these variables through digital ops twin in Celonis EMS

→ **To maximize aircraft productivity, delays should be minimized!**

The DI-Lab project should leverage the gathered data to predict and prevent delays

WE FORMULATED A SMART GOAL

Specific **M**easurable **A**ttainable **R**elevant **T**ime-based

*„We will develop a **machine learning model** that predicts the estimated delay of flights based on several features in the flight process and a **dashboard based on feedback** from future users until February 10th in Celonis EMS.“*

OUR PROJECT PLAN WAS DIVIDED INTO THREE PHASES

Project Timeline

CW 47/21
Celonis Trainings and Lufthansa IT Setup

CW 49/21
Data Preprocessing and definition of model approach

CW 51/21
Milestone 1:
First Model ready

CW 52/21
Iterate model, improve with different features & set up dashboard mock-up

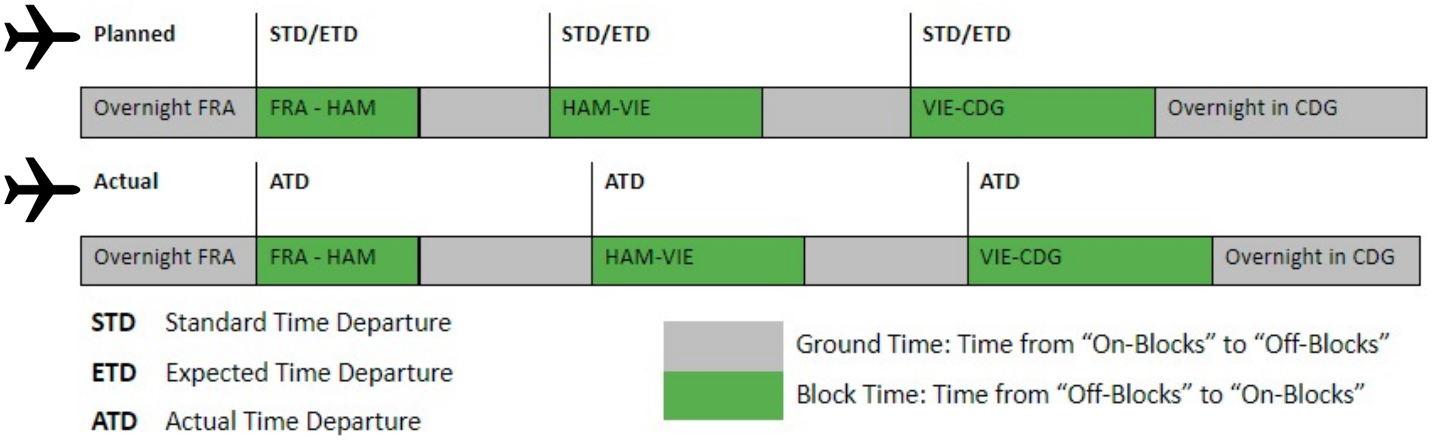
CW 2/22
Milestone 2:
Iterated model with first dashboard draft based on user feedback

CW 5/22
Prepare Handover, Prediction Pipeline Documentation & Project Report

ROTATION PROCESS

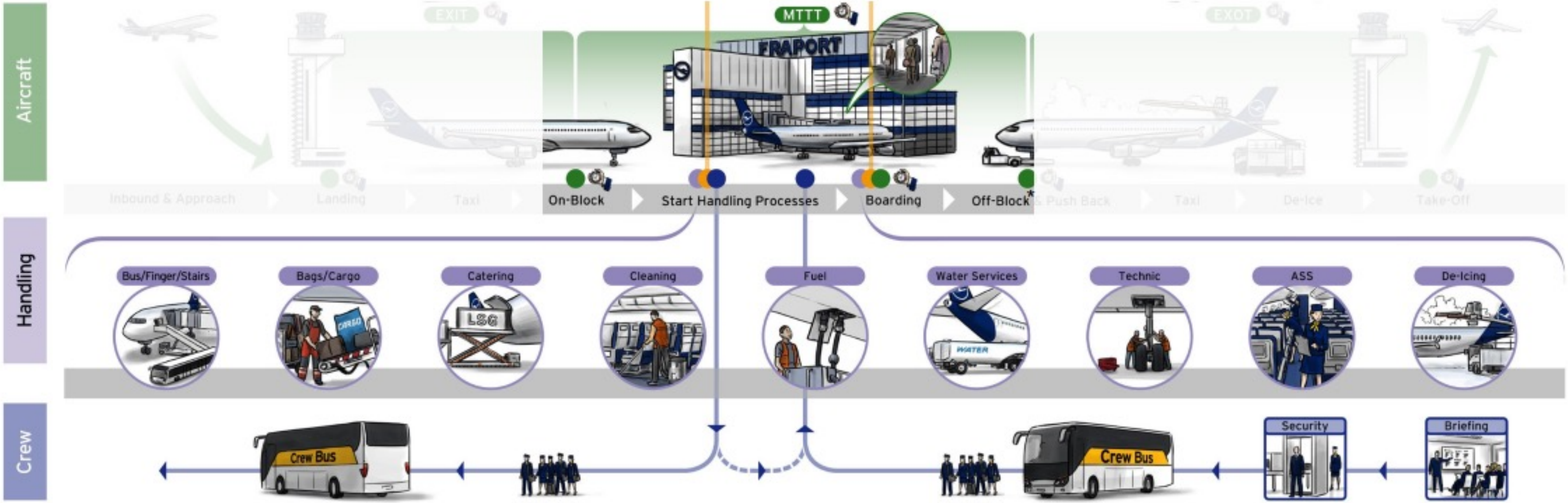
Rotation process overview

- Lufthansa monitors > 300 aircraft with more than 1500 daily flights
- An aircraft flies multiple legs (sectors) a day – rotation
- Operations team steers ground processes with tight schedules



ROTATION PROCESS

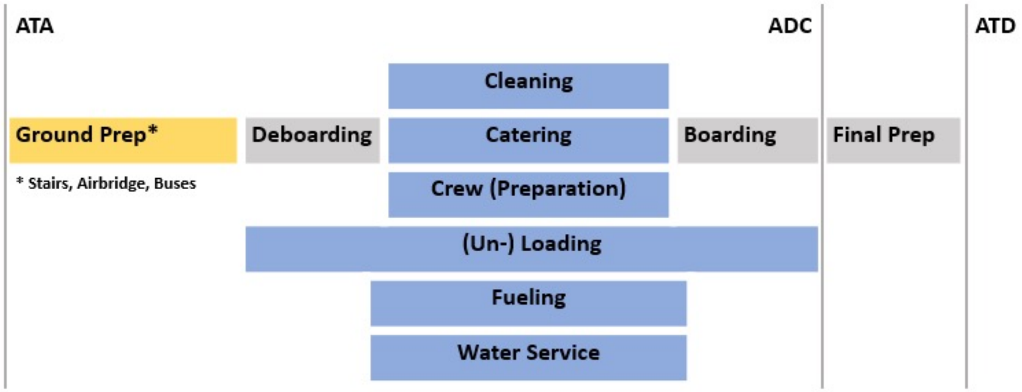
Typical Ground Processes



- Ground processes are planned and monitored by operations teams

ROTATION PROCESS

Critical Path



ATA	Actual Time of Arrival
ADC	All Doors Closed
ATD	Actual Time of Departure

- Not all processes influence the departure time
- Few other factors influcence the departure time (parking position, runway, weather, traffic etc.)
- Detailed ground service timestamps are only available at FRA & MUC hubs.

FEATURE SELECTION

We iteratively refined our Feature Selection

FLIGHT LEG & GROUND OPERATION DATA

- **General flight data:** Flight number, Flight date, Scheduled time of departure (STD), Departure and Arrival Airport, Subfleet
- **Temporal data:** Month, Day and Weekday of the flight
- **Rotational data:** Flight of the day and Rotation Type
- **Operational flight data:** Total number of passengers booked, Departure Runway, Taxi-in and Taxi-out time, Flight Distance and Calculated Flight Time
- **Label features:** Delay flag or Delay delta

WE ADDED FURTHER FEATURES

Feature Engineering

WEATHER DATA

- wind data, visibility and snow from a public database[3][6]

AIRPORT TRAFFIC DATA

- peak times for FRA and MUC
- indicating peak/ off-peak

DATA ON PREVIOUS FLIGHT

- delay or cancellation of the previous flight
- flight time, distance and number of passengers

EVENT RELATED DATA

- process durations
- medians for actual duration, target duration and delta between both

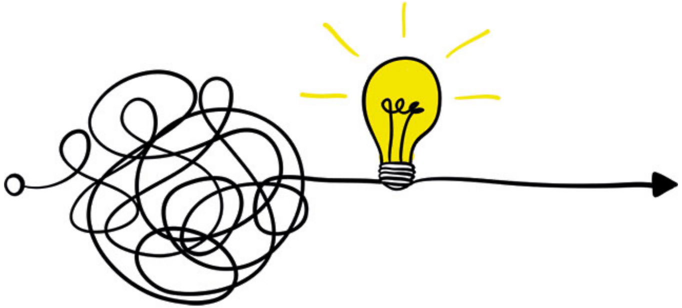
REAL LIFE DATA = CHAOS

Raw Data

FLIGHT LEG & GROUND OPERATION DATA

Filter for Flights with _Case_Key start with “LH” from 1st December 2020 to 1st December 2021, e.g, LH223_2021-11-30

→ approx. 20,000 rows of raw data



Clean data with no missing values, as few errors as possible

DATA MUNGING

Errors or edge cases

- Duplicates of Case Keys -> Remove
  - Triangle Flights
  - Make sure Case Key is unique
- Cancelled Flights -> Remove
  - Where Cancellation Time exists
- Flights with Zero and Negative Passengers -> Remove
  - Ferry Flights i.e. non-revenue-generating flight
  - Input error
- Filtering Flights Departing from MUC & FRA
  - Ground operation data only available for Munich and Frankfurt
  - Around half of flights filtered out

DATA MUNGING

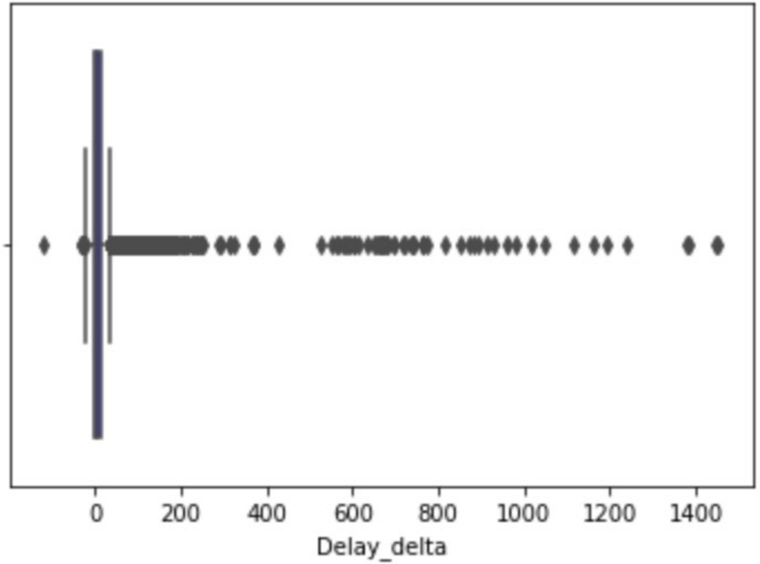
Missing Values

- Missing Independent Variables –> Drop all
  - ML model require target variables (Delay) to train
- Missing Inbound Information –> Drop all
  - Due to removal of triangle flights
- Missing Runway Data -> Imputation
  - Impute through a mapping from wind direction and speed to runway
  - More precise but also require extra engineering step
- Missing Other Dependent Variables -> Imputation
  - Impute based on historical data i.e. median or mode group by flight number
  - E.g. Replace missing Taxi In Time of LH108_2021-05-05 with median Taxi In Time of all LH108 flights

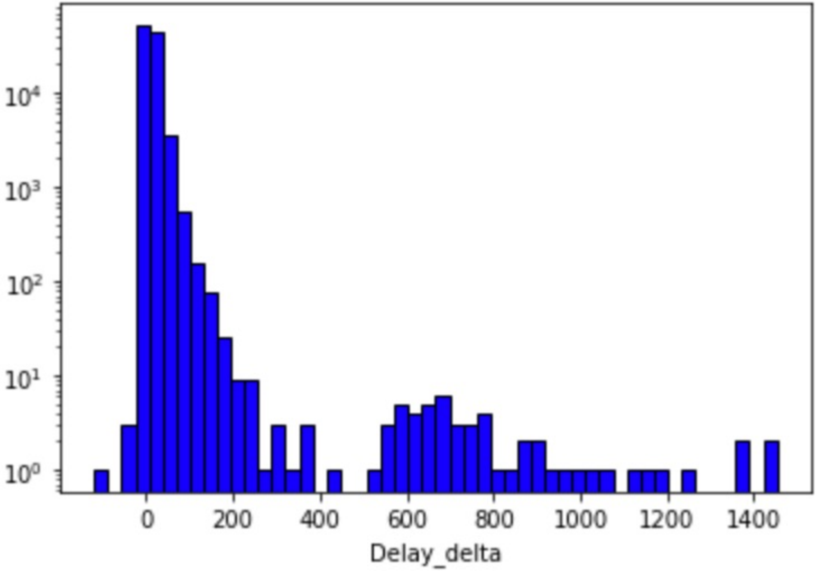
DATA MUNGING

Outliers

- Removing Outliers
  - Flights with Delay Delta > 200 mins



Boxplot of Delay_delta



Histogram of Delay_delta

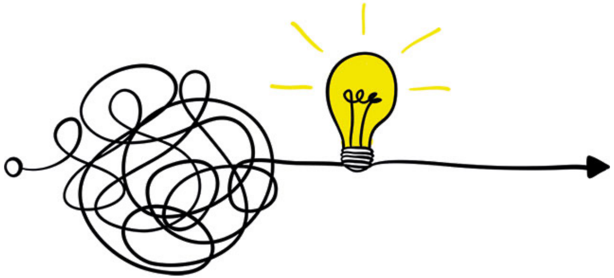
DATA MUNGING - RESULTS

Clean Data

FLIGHT LEG & GROUND OPERATION DATA

Filter for Flights with _Case_Key start with "LH" from 1st December 2020 to 1st December 2021

→ approx. 20,000 rows of raw data



Roughly 10,000 rows of clean data to train and test the ML models.

"Assumption" of flights:

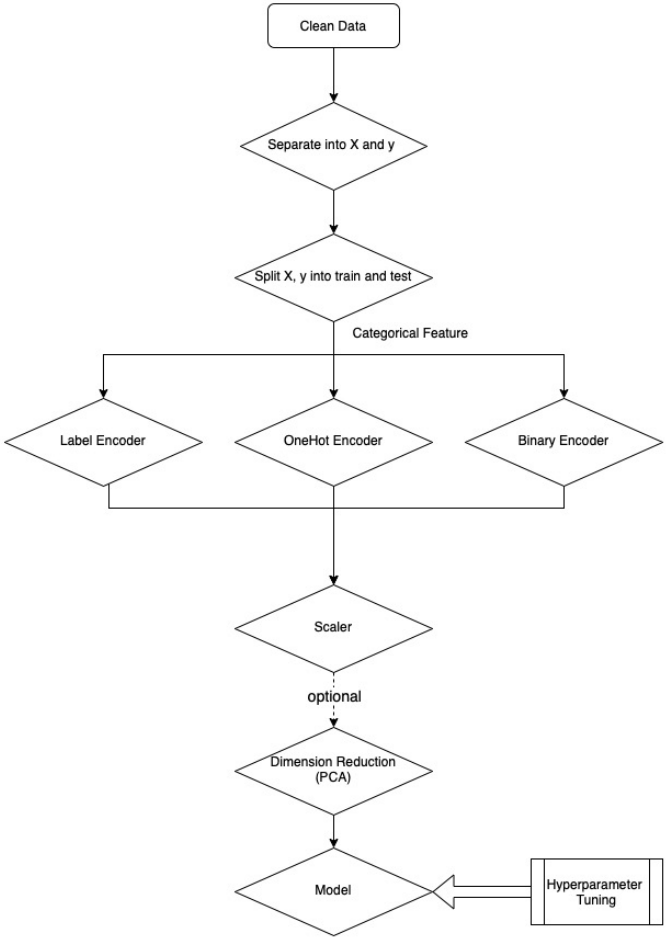
1. Non-triangle passenger flights
2. Flight must have inbound flight
3. Flight departs from MUC or FRA (inbound flight can depart from any airport)

MACHINE LEARNING MODEL - ALGORITHM

Chosen Algorithms

Random Forest and its Variants

- Suitable for both regression or classification task
- Able to handle binary features, categorical features, and numerical features
- Great with high dimensional data
- Quick Training/Prediction Speed



MACHINE LEARNING MODEL - METRICS

Evaluation Metrics

Aim: prediction for Lufthansa to take measures to counter the possible delays

Evaluation metrics:

- Classification:
  - Precision (and recall)

- Regression:
  - R2 score

In the context of the project

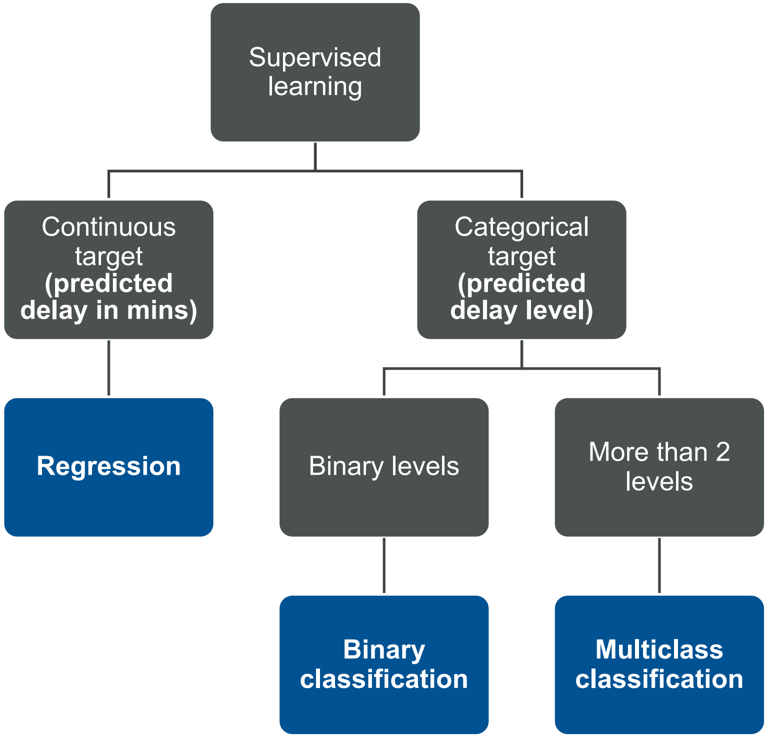
$$\boldsymbol{Precision} = \frac{\textbf{\# correct predicted delays}}{\textbf{\# total predicted delays}}$$

$$\boldsymbol{Recall} = \frac{\textbf{\# correct predicted delays}}{\textbf{\# total true delays}}$$

$$\boldsymbol{R2\ score} = \textbf{the proportion of the variance}$$
$$\textbf{in the delay that is predictable from the features}$$

MODEL – MACHINE LEARNING PROBLEM

What should be shown on the dashboard?



MODEL EXPERIMENTS – BINARY VS. MULTICLASS

How feasible is a fine-grained categorization of delay?

Type	Class	Precision	Recall
Binary	0 (0-14mins)	0.88	0.76
Binary	1 (15-200mins)	0.38	0.60
Multiclass	0 (0-4mins)	0.89	0.70
Multiclass	1 (5-14mins)	0.21	0.40
Multiclass	2 (16-30mins)	0.16	0.30
Multiclass	3 (30-200mins)	0.01	0.01

- Overall better accuracy and recall for binary classification
- Multiclass: Classes 0 and 3 are often confused
- **DECISION: Binary classification**

MODEL EXPERIMENTS – UPSAMPLING (SMOTE)

Since random forest is sensitive to imbalanced classes

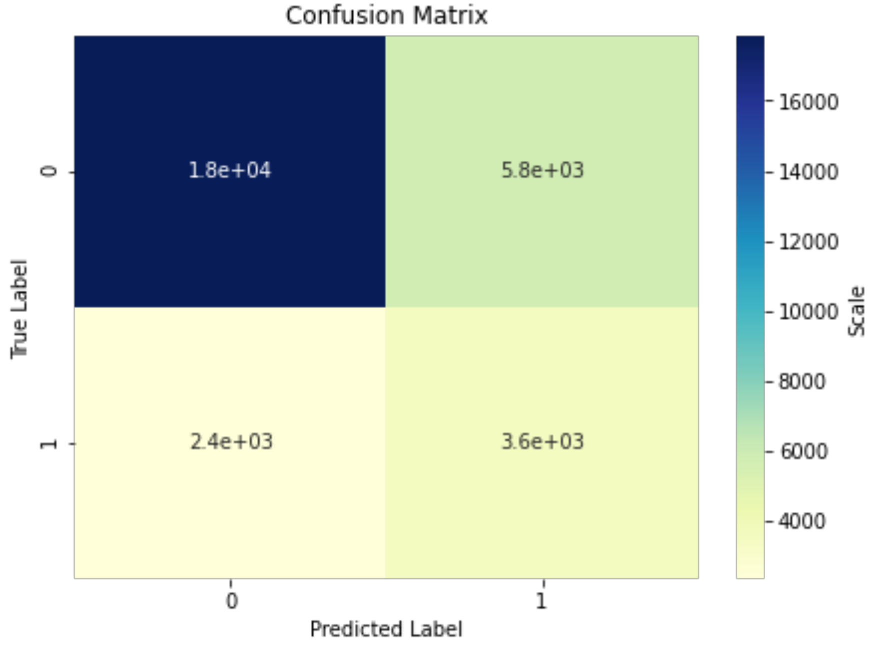
- Upsampling is done with SMOTE-NC: SMOTE for datasets with numerical and categorical features

Type	Class	Precision	Recall
Binary	0 (0-14mins)	0.88	0.76
Binary	1 (15-200mins)	0.38	0.60
Binary with SMOTE-NC	0 (0-14mins)	0.86	0.85
Binary with SMOTE-NC	1 (15-200mins)	0.43	0.46

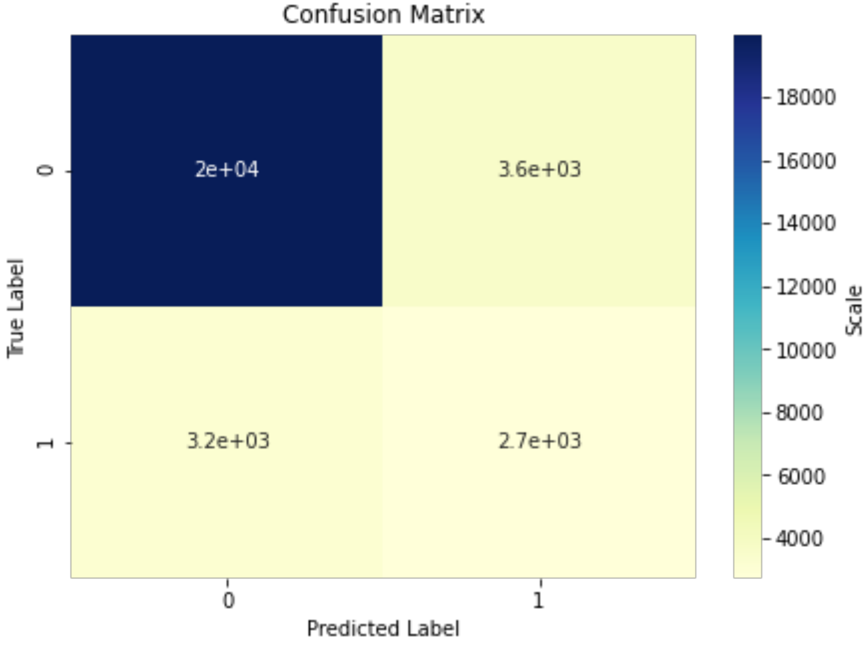
- Better precision with resampling, but less recall for delay = 1
- **DECISION: Upsample**

MODEL EXPERIMENTS - BINARY

Binary classification

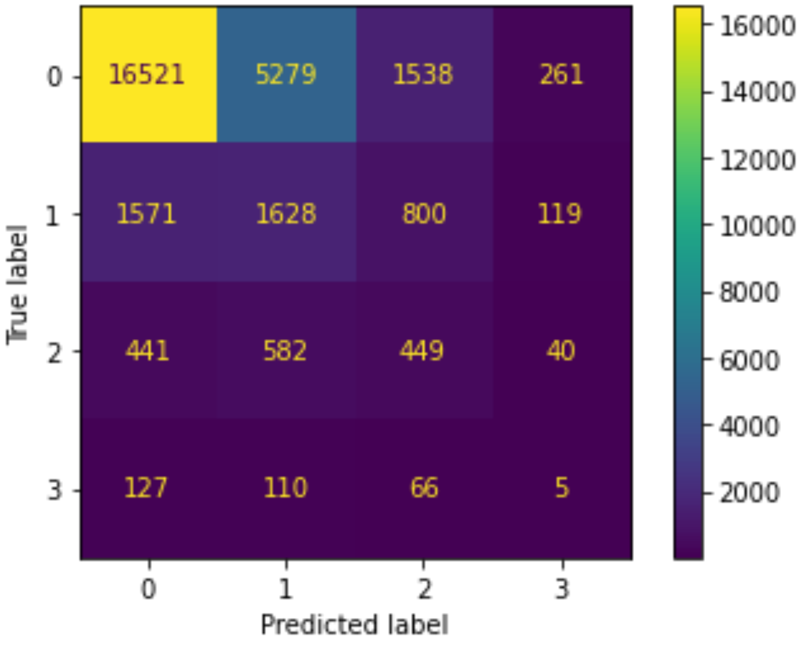


Binary classification + upsampling

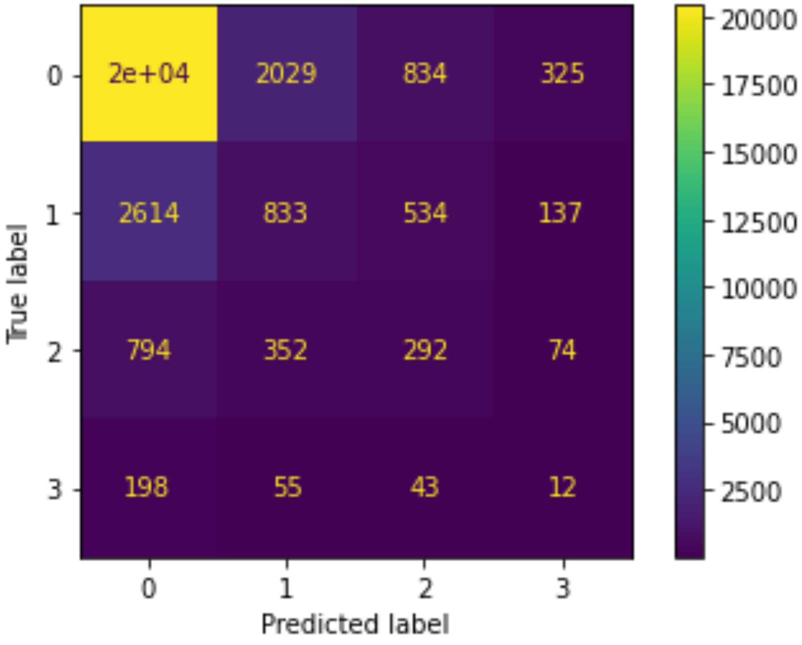


MODEL EXPERIMENTS - MULTICLASS

Multiclass classification



Multiclass classification + upsampling



MODEL EXPERIMENTS – REGRESSION

	R2 on Test Set	R2 on Train Set
Regression	0.10	0.86

- R2 score high on the training set, but low on the test set i.e. Overfitting
- **DECISION: Tend towards classification**

	index	TARGET	PREDICTION	Month	Day	Weekday	Departure_
0	81533	27.0	9.30	9	11	6.0	
1	81534	4.0	20.70	7	3	6.0	
2	81535	22.0	36.35	10	11	1.0	
3	81536	14.0	18.20	3	26	5.0	
4	81537	8.0	4.40	7	2	5.0	
5	81538	34.0	24.55	8	21	6.0	...
6	81539	30.0	30.05	10	15	5.0	
7	81540	4.0	15.15	9	10	5.0	
8	81541	3.0	14.45	11	22	1.0	
9	81542	8.0	11.30	7	22	4.0	

MODEL – HYPERPARAMETER OPTIMIZATION

- After trying out different variation to search for possibility of improvements, we decide to use binary classification with resampling as our base model.

- Hyperparameter tuning with *Optuna* library
  - Model-based optimization
  - Optimize with respect to certain metric(s), e.g. precision, recall or average precision
  - Pre-defined parameter grid

MODEL – RESULTS

Combinations of model, optimized metrics are tested

Model type	Optimized metric(s)	Precision	Recall
RF	Precision	0.61	0.18
RF	Precision, Recall	0.38	0.67
RF	Average precision	0.43	0.52
Balanced RF	Precision	0.38	0.66
Balanced RF	Precision, Recall	0.50	0.35
Balanced RF	**Average precision**	**0.47**	**0.55**
XGBoost	Average precision	0.46	0.51

No model has a sufficiently high precision and recall → **A compromise must be found!**

MODEL – FINAL DECISION

- Prediction of delayed flights are accurate
- A lot of delayed flights are not predicted as delayed (false negative)

Precision

Recall

- A lot of delayed flights are predicted as delayed
- A lot of not delayed flights are predicted as delayed (false positive)

Resampled balancedRF binary classifier optimized for average precision

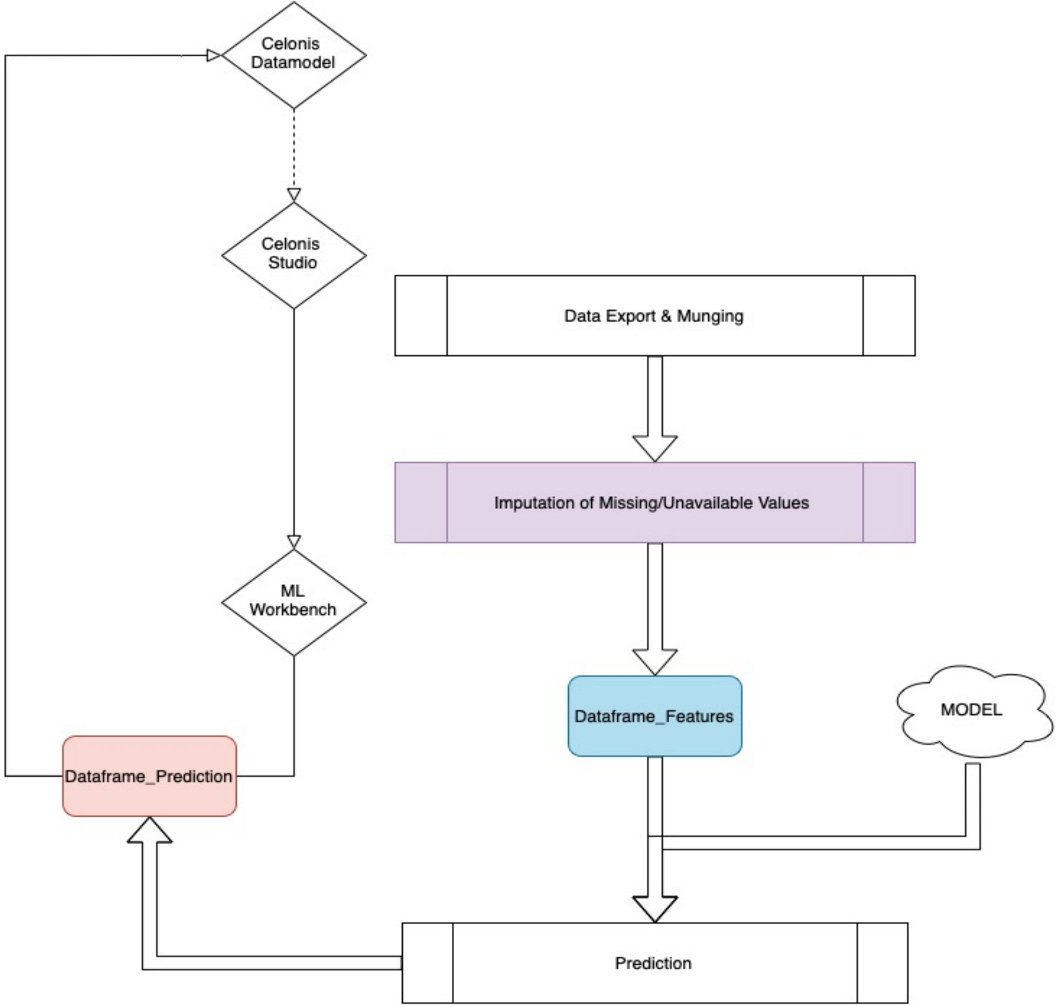
PREDICTION - FEATURE AVAILABILITY

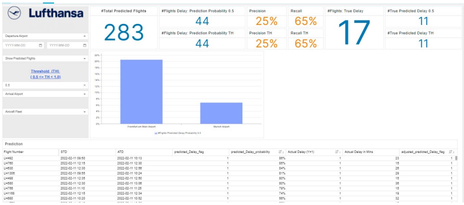
Availability of features for future flights

Issue : Some features of future flights will not be available by the time running prediction (presumably 7 AM everyday)

Solution: **Imputation,** similar to data munging process for training data

PREDICTION
- PIPELINE

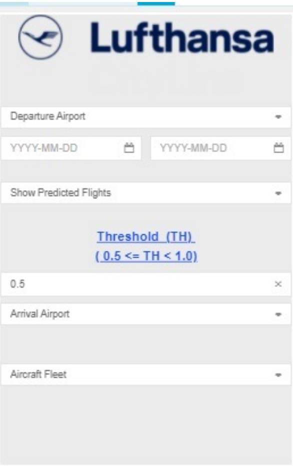




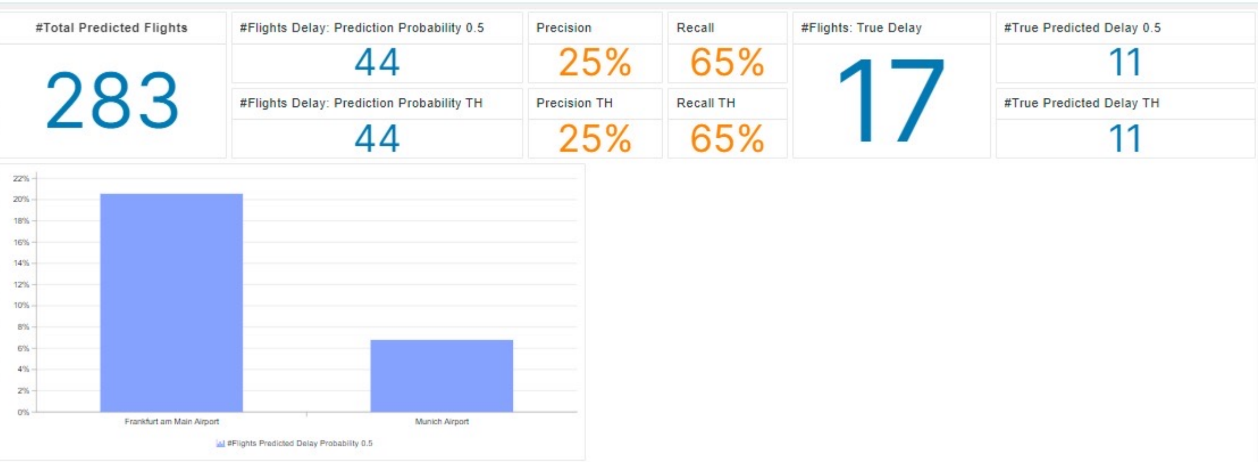
DASHBOARD

Dashboard in Celonis

1. User Inputs (filters)



3. Summary of predictions



Prediction

Flight Number	STD	ATD	predicted_Delay_flag	predicted_Delay_probability	Actual Delay (Y=1)	Actual Delay in Mins	adjusted_predicted_Delay_flag
LH492	2022-02-11 09:50	2022-02-11 10:13	1	86%	1	23	1
LH760	2022-02-11 12:15	2022-02-11 12:30	1	85%	1	15	1
LH630	2022-02-11 12:30	2022-02-11 12:56	1	84%	1	26	1
LH1306	2022-02-11 09:55	2022-02-11 10:24	1	81%	1	29	1
LH498	2022-02-11 12:35	2022-02-11 12:50	1	80%	1	15	1
LH580	2022-02-11 12:30	2022-02-11 13:06	1	80%	1	36	1
LH756	2022-02-11 11:10	2022-02-11 11:25	1	79%	1	15	1
LH1168	2022-02-11 12:15	2022-02-11 12:34	1	74%	1	19	1
LH680	2022-02-11 10:20	2022-02-11 10:52	1	56%	1	32	1

2. Summary table with additional details

CONCLUSION

Final products

- ☑ We preprocessed relevant data and developed a Random Forest Classification model with relatively good precision and recall to yield rotational delays predictions
- ☑ We visualized the prediction results in a Celonis Dashboard
- ☑ The dashboard can be used by the Lufthansa Ops Steering Department as a decision aid to initiate countermeasures for flights with an anticipated delay
- ☑ Thereby we have achieved our SMART goal

LIMITATIONS & OUTLOOK

Basis for future improvements

- Pandemic effects in training data
  - ➢ retrain the model on data resembling the current amount of air traffic
- Imputation of missing values by median or mode
  - ➢ use more accurate methods to approximate missing values (e.g., based on influencing features)
- Random Forest model
  - ➢ apply different predictive algorithms (e.g., Recurrent neural networks, Long Short-Term Memory [7,8,9])

THANK YOU!

WE ARE EXCITED TO HEAR YOUR FEEDBACK AND QUESTIONS.

REFERENCE

- [1] Lufthansa,Lufthansa Group Website, https://www.lufthansagroup.com/en/company.html,Accessed: 02.02.2022
- [2] Celonis SE,Celonis Website, https://www.celonis.com/company/,Accessed: 02.02.2022
- [3] Iowa Environmental Mesonet,Iowa Stat University Database,https://mesonet.agron.iastate.edu/ASOS/, Accessed: 03.02.2022
- [4] Chawla et al. 2002.SMOTE: Synthetic Minority Over-sampling Technique,https://arxiv.org/pdf/1106.1813.pdf, Accessed: 03.02.2022
- [5] Chen et al. 2004.Using Random Forest to Learn Imbalanced Data,https://statistics.berkeley.edu/tech-reports/666, Accessed: 03.02.2022
- [6] TAF Text Data (by station ID),Aviation Weather Center,https://www.aviationweather.gov/taf/data, Accessed: 03.02.2022
- [7] Gui et al., 2019.Flight delay prediction based on aviation big data and machinelearning. IEEE Transactions on Vehicular Technology, 69(1), pp.140-150.
- [8] Kim et al., 2016.A deep learning approach to flight delay prediction. IEEE/AIAA35th Digital Avionics Systems Conference (DASC), pp. 1-6.
- [9] Huang et al., 2020.Modeling train operation as sequences: A study of delay pre-diction with operation and weather data. Transportation Research Part E: Lo-gistics and Transportation Review, Volume 141, 2020, 102022, ISSN 1366-5545,https://doi.org/10.1016/j.tre.2020.102022.
- [10] Celonis SE,PyCelonis documentation, https://celonis.github.io/pycelonis/index.html/,Accessed: 11.02.2022