

1

Data Preparation

OAG Flight Dataset

1.74M European Flights
Mar, Jun, Sep, Dec 2019
Turnaround ≤ 6 hours

Pre-tactical Features

Categorical: Carrier, Airports, Aircraft

Temporal: Month, Day, Hour, DoW

Operational: Scheduled Duration

80/20 Split

Train: 1,395,733 flights
Test: 348,934 flights

2

Synthetic Generation

Synthetic Data Generators

Gaussian Copula

Statistical modeling
Marginals + correlation structure

CTGAN

Adversarial learning
Mode-specific normalization

TabSyn

VAE + Diffusion models
Two-stage latent space approach

REaLTabFormer

Transformer architecture
GPT-2 with column tokenization

3

Model Training

Machine Learning Models

Decision Tree

Random Forest

Gradient Boosting

XGBoost

CatBoost

4

Evaluation

Prediction Tasks

- Departure Delay
- Arrival Delay
- Turnaround Time

Evaluation Metrics

Performance:

- RMSE, MAE, R^2

Utility:

- TSTR Utility Score

Feature Analysis:

- Importance Alignment

Train Data

Synthetic

Models

TRTR (baseline)

Test Data