

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

CTGAN

Adversarial learning
Mode-specific norm.

TabSyn

VAE + Diffusion
Latent space approach

REaLTabFormer

Transformer arch.
GPT-2 tokenization

Synthetic

Fidelity Assessment

- Statistical Similarity (KS, χ^2)
- Correlation Preservation
- Joint Distribution (KL Div.)
- Likelihood Assessment
- Detection Difficulty
- Mixed-Type Metrics

TRTR (baseline)

3

Predictive Models

Machine Learning Models

Decision Tree

Random Forest

Gradient Boosting

XGBoost

CatBoost

Models

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