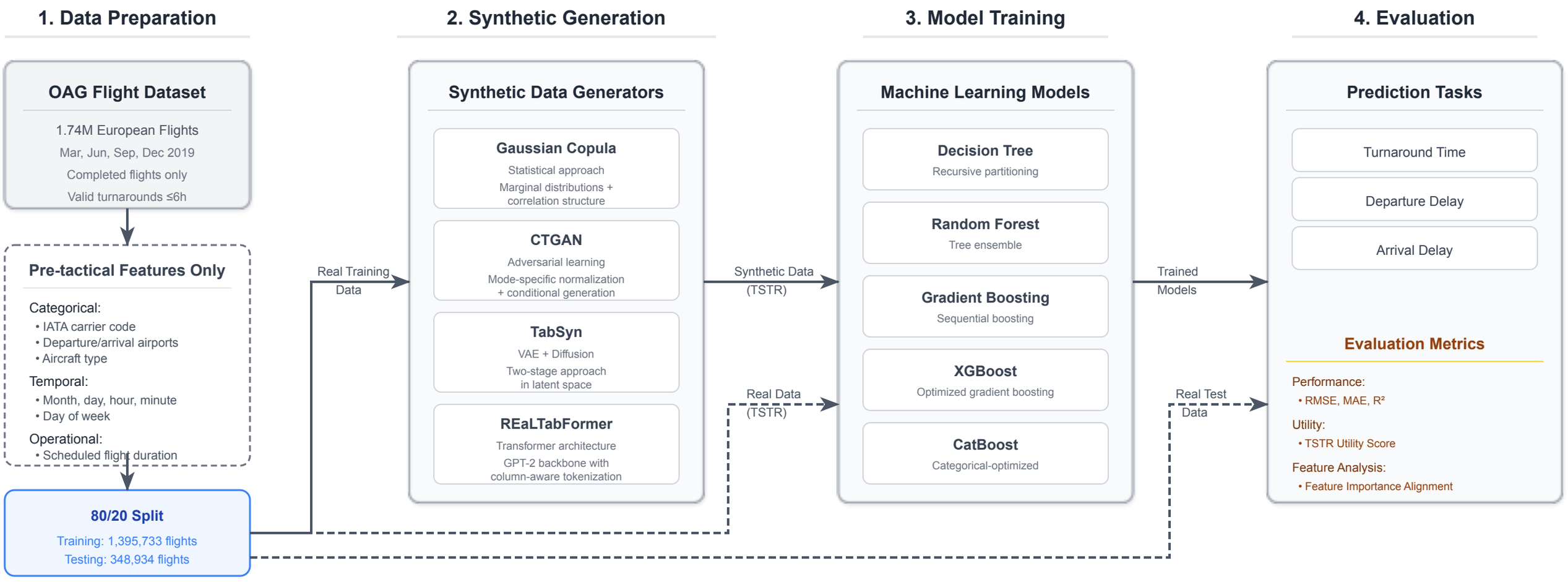
1. Data Preparation
2. Synthetic Generation
3. Model Training
4. Evaluation
OAG Flight Dataset
1.74M European Flights
Mar, Jun, Sep, Dec 2019
Completed flights only
Valid turnarounds ≤6h
Pre-tactical Features Only
Categorical:
• IATA carrier code
• Departure/arrival airports
• Aircraft type
Temporal:
• Month, day, hour, minute
• Day of week
Operational:
• Scheduled flight duration
80/20 Split
Training: 1,395,733 flights
Testing: 348,934 flights
Real Training Data
Synthetic Data Generators
Gaussian Copula
Statistical approach
Marginal distributions + correlation structure
CTGAN
Adversarial learning
Mode-specific normalization + conditional generation
TabSyn
VAE + Diffusion
Two-stage approach in latent space
REaLTabFormer
Transformer architecture
GPT-2 backbone with column-aware tokenization
Synthetic Data (TSTR)
Real Data (TSTR)
Machine Learning Models
Decision Tree
Recursive partitioning
Random Forest
Tree ensemble
Gradient Boosting
Sequential boosting
XGBoost
Optimized gradient boosting
CatBoost
Categorical-optimized
Trained Models
Real Test Data
Prediction Tasks
Turnaround Time
Departure Delay
Arrival Delay
Evaluation Metrics
Performance:
• RMSE, MAE, R²
Utility:
• TSTR Utility Score
Feature Analysis:
• Feature Importance Alignment