TAT Estimation – AI/ML Flowchart

Data Sources
- Claims CSV
- Reporting Table CSV
- Rework Status Logs

▼

Data Ingestion Layer
- Pandas CSV Readers
- Schema Validation
- Data Quality Checks
- Define Success Metrics (MAE, R^2)

▼

Data Processing & Cleaning Layer
- Column Normalization
- Missing Value Handling
- String Standardization
- Datetime Conversion
- Outlier Detection

▼

Feature Engineering Layer
- TAT Duration Calculation (Target y)
- ClaimID-Based Merge
- Request-Level Aggregation
- Status Count Pivoting
- Rework Frequency Features
- Priority Encoding

▼

Feature Transformation Layer
- Categorical Encoding
- Numerical Scaling (StandardScaler)
- Feature Readiness Validation

▼

ML Training Pipeline
- Train / Test Split
- Supervised Regression Models
  - RandomForest
  - GradientBoosting
  - ExtraTrees
  - HistGradientBoost
  - XGBoost
  - LightGBM
  - CatBoost
- Neural Network Baseline (MLP)

▼

Hyperparameter Optimization Layer

- GridSearchCV
- Cross Validation (CV = 3)
- Bias-Variance Tradeoff
- Overfitting Prevention

▼

Model Evaluation & Selection

- MAE – Business Error Tolerance
- RMSE – Large Error Penalty
- R² – Variance Explained
- Best Model Selection (XGBoost)

▼

Prediction & Interpretation

- Predict TAT in Seconds
- Convert to HH:MM:SS
- Attach Predictions to Claims

▼

Output & Consumption Layer

- CSV Reports
- Performance Charts
- Dashboards / APIs
- SLA & Ops Insights

▼

Monitoring & Continuous Improvement

- Data Drift Detection
- Model Performance Monitoring
- Periodic Retraining
- Feedback Loop