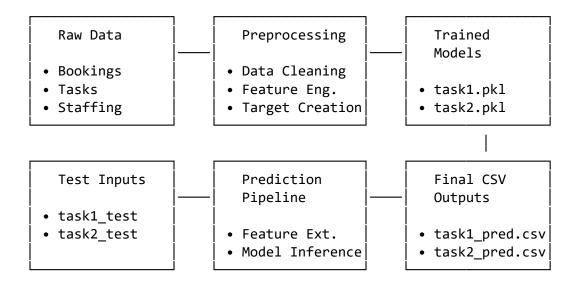
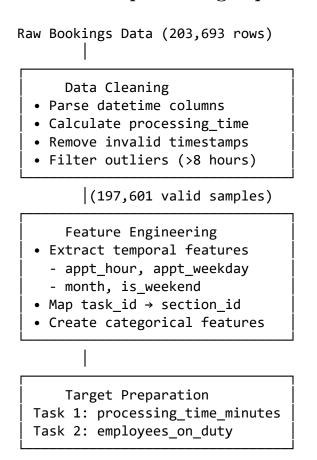
Architecture Diagrams

BitByBit Datathon 2025 - Team BitByBit

1. Overall System Architecture



2. Data Preprocessing Pipeline



3. Model Architecture - Task 1 (Processing Time)

```
Input Features: [task_id, appt_hour, appt_weekday, section_id, ...]
|
```

```
Hierarchical Baseline Model

Level 0: task_id + appt_hour + appt_weekday

▼ (if no match)

Level 1: task_id + appt_hour

▼ (if no match)

Level 2: task_id + appt_weekday

▼ (if no match)

Level 3: task_id only

▼ (if no match)

Level 4: appt_hour + appt_weekday

▼ (if no match)

Level 5: appt_hour only

▼ (if no match)

Level 6: Global Median (42.8 minutes)
```

Output: Processing time in minutes (integer, 0-480 range)

4. Model Architecture - Task 2 (Staffing)

Input Features: [section_id, weekday, month, is_weekend, ...]

```
Hierarchical Baseline Model

Level 0: section_id + weekday

▼ (if no match)

Level 1: section_id only

▼ (if no match)

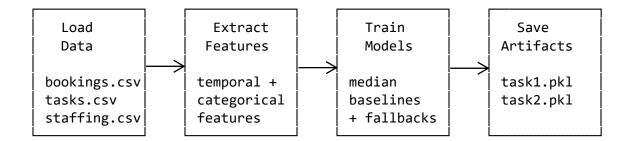
Level 2: weekday only

▼ (if no match)

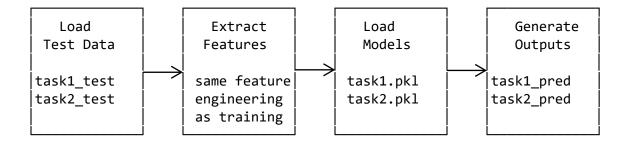
Level 3: Global Median (4 employees)
```

Output: Number of employees needed (integer, 1-50 range)

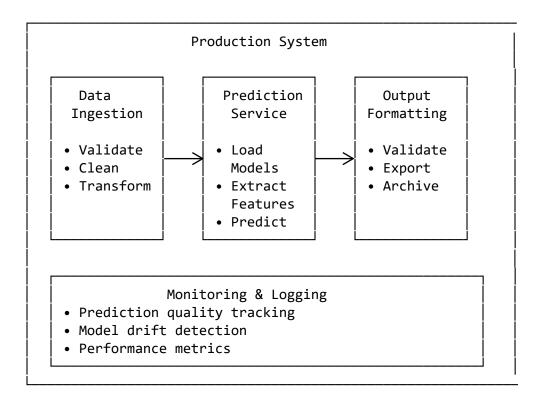
5. Training Pipeline Flow



6. Prediction Pipeline Flow



7. Deployment Architecture (High Level)



8. Technology Stack

Tech Stack

Data Processing: pandas, numpy

Modeling: scikit-learn (optional), custom baselines

Validation: Custom validation functions I/O: CSV, pickle serialization

Logging: Python logging module

Environment: Python 3.8+, cross-platform

9. Key Design Decisions

• Baseline-First Approach: Prioritized robust statistical methods over complex ML

• Hierarchical Fallbacks: Ensured 100% prediction coverage for all test cases

• Modular Architecture: Separated preprocessing, training, and prediction concerns

• Defensive Programming: Comprehensive error handling and data validation

• Business Logic Integration: Applied domain knowledge (working hours, staffing limits)

10. Performance Characteristics

Training Performance:

Task	MAE	RMSE	Samples
Task 1	14.37 min	19.74 min	197,601
Task 2	0.60 emp	0.80 emp	5,802

Prediction Coverage: 100% (no failed predictions)
Model Size: <1MB per model (lightweight .pkl files)

Execution Time: <30 seconds for full pipeline