

ADVANCED PROCESS MINING PROJECT USING PM4Py



1. Project Overview

This project focuses on analyzing business process execution using **Process Mining techniques** implemented through the **PM4Py Python library**.

The objective is to move from:

- Assumed process flow (documentation-based)
- to
- Real executed process (data-driven analysis)

The system analyzes event logs generated by business systems and automatically reconstructs the actual process flow.

2. Business Objective

Organizations face problems such as:

- Unknown process delays
- Process deviations
- Rework and inefficiencies
- Lack of visibility in operations

This project solves these problems by:

- Discovering the real process flow
- Identifying bottlenecks
- Checking process compliance
- Improving operational efficiency

3. Real-World Scenario (Industry Example)

Consider an e-commerce company:

Expected Process

Order Received → Payment Done → Order Shipped

Actual Process (Real Logs)

Order Received → Payment Done → Payment Recheck → Order Shipped

The extra step causes delay.

Without process mining:

 This issue remains hidden.

With process mining:

 Issue becomes visible automatically.

4. Event Log Structure (Core Concept)

Event log is the foundation of process mining.

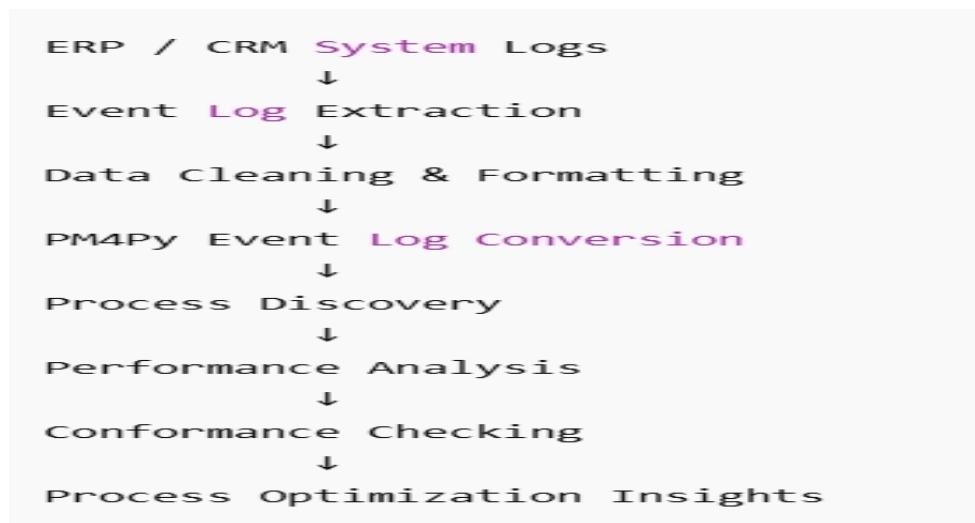
Each event contains:

Field	Description
 Case ID	Unique process instance
 Activity	Step executed
 Timestamp	Execution time

Example:

Case	Activity	Time
101	Order Received	10:00
101	Payment Done	10:05
101	Shipped	12:00

5. Project Architecture (Industry Flow)



6. Methodology

Step 1 — Data Preprocessing

Tasks performed:

- Missing timestamp handling
- Sorting events by time
- Ensuring correct case sequence

Importance:

 Incorrect ordering leads to wrong process discovery.

Step 2 — Event Log Conversion

Data is converted into PM4Py format to enable:

- Process mining algorithms
- Performance analysis
- Visualization

Step 3 — Process Discovery

Algorithm Used:

Inductive Miner

Why used in industry:

-  Produces sound models
-  Handles noise
-  Supports parallel activities

Output:

- Petri Net
- Initial marking
- Final marking

Step 4 — Process Visualization**

Two major models are generated:

Petri Net

Shows:

- Flow logic
- Start and end events
- Activity transitions

Used for:

- Process understanding
- Simulation
- Conformance checking

Directly-Follows Graph (DFG)

Shows:

- Which activity follows another
- Frequency of transitions

Used for:

- Process overview
- Identifying common paths

Step 5 — Performance Analysis

Performance DFG calculates:

-  Average time between activities
-  Waiting time
-  Throughput time

This helps detect:

-  Bottlenecks
-  Delayed activities

7. Bottleneck Detection

A bottleneck is a stage where process speed decreases.

Example:

Payment Done → (3 hours delay) → Shipping

Possible reasons:

- Manual approval

- Resource shortage
- System delay

Business impact:

- Customer dissatisfaction
- Increased operational cost

8. Conformance Checking (Advanced Industry Concept)

Conformance checking compares:

Expected Process Model

VS

Actual Execution Log

Used to detect:

- Rule violations
- Skipped activities
- Unauthorized process paths

Example:

Order shipped without payment confirmation.

9. Variant Analysis

Variant = Different execution paths.

Example:

Variant 1:

Order → Payment → Shipping

Variant 2:

Order → Payment → Recheck → Shipping

Used for:

- Understanding process diversity
- Identifying rare or problematic flows

10. Results & Insights

From analysis:

- Most common process path identified
- Delay between payment and shipping detected
- Additional rework steps discovered
- Process inefficiencies highlighted

11. Industry Applications

Process Mining is actively used in:

-  Banking — Loan approval optimization
-  Healthcare — Patient treatment flow
-  E-commerce — Order lifecycle analysis
-  Manufacturing — Production monitoring
-  Customer Support — Ticket resolution analysis

Industry tools:

- Celonis
- SAP Signavio
- UiPath Process Mining
- PM4Py (Research & Data Science)

12. Interview Important Questions

Q1: What is Process Mining?

Process mining analyzes event logs to discover and improve real business processes.

Q2: Difference between Data Mining and Process Mining?

- Data Mining → Pattern discovery in data
- Process Mining → Process flow discovery from logs

Q3: What is a Bottleneck?

A step causing maximum delay in process execution.

Q4: Why Petri Net?

To formally represent process behavior and flow logic.

13. Conclusion

This project demonstrates how process mining transforms raw system logs into actionable business insights.

Benefits:

- Process transparency
- Performance improvement
- Cost reduction
- Data-driven decision making