

# Technical Documentation: AI-Powered Process Mining for E-commerce

Master Project Documentation;

Academic Year: 2025-2026;

Dataset: <https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-electronics-store/data>

## Summary

Introduction.....	2
1. Project Objectives .....	2
2. Description of the Dataset .....	3
2.1 Overview of E-commerce data.....	3
2.2 Structure and Format (CSV/XES).....	3
2.3 Event Log Statistics .....	3
3. Methodology .....	5
3.1 Data Preprocessing (Cleaning, Filtering e Sessionizzazione) .....	5
3.2 Process Discovery algorithms (Alpha, Heuristic, Inductive).....	5
3.3 Quality Evaluation Metrics.....	5
4. Results and Findings .....	6
4.1 Comparative Analysis of Model Performance .....	6
4.2 Characteristics of Discovered and Variant Processes .....	6
5. Reasoning and Analysis via LLM.....	8
5.1 Generation of Strategic Reports .....	8
5.2 Identification of Anomalies and Bottlenecks .....	9
5.3 Process Optimization Strategies .....	9
6. Predictive Analysis .....	11
6.1 Prediction of Next Activity.....	11
6.2 Behavior Patterns and Transition Probability.....	13
7. Implementation: Interactive Dashboard .....	14
7.1 Technology Stack and System Architecture .....	14
7.2 Chatbot interface and AI assistant.....	14
7.3 Guide to Using and Viewing KPIs.....	15
8. Conclusions and Future Work.....	16

# Introduction

This project aims to apply Process Mining methodologies to analyze, evaluate and optimize purchasing processes within an E-commerce platform. The analysis is not limited to simple Process Discovery, but integrates advanced Artificial Intelligence (LLM) techniques to provide critical and predictive analysis of data.

## 1. Project Objectives

The work was structured to cover the entire Process Mining life cycle, as required by the specifications:

- **Data Analysis:** Knowledge extraction from a real dataset of e-commerce events.
- **Process Modeling:** Application of discovery algorithms (Alpha, Heuristic, Inductive) to generate Petri Net and Process Tree models.
- **Qualitative Evaluation:** Measurement of model performance through standard metrics of Fitness, Precision, Generalization and Simplicity.
- **Reasoning & Intelligence:** Integration of a Large Language Model (Llama 3 via Groq) for automatic report generation, anomaly identification and suggestions for business optimization.
- **Prediction:** Implementation of a predictive engine to anticipate future user activities based on historical behavior.

## 2. Description of the Dataset

For this study, a dataset relating to user sessions on an e-commerce site was selected. The data captures fundamental user interactions, such as viewing products (view), adding to cart (cart) and completing the purchase (purchase), allowing you to map the "Customer Journey" and identify points of abandonment or inefficiency.

### 2.1 Overview of E-commerce data

The dataset used for the study consists of the events.csv file (taken from the link <https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-electronics-store/data>) which collects the event logs generated within an e-commerce platform. The database captures the main interactions of users, including viewing products, placing them in the cart and the final purchase phase, allowing a complete mapping of the customer journey. Each log line records essential information such as the session identifier, the type of activity, the exact time reference and the user ID. Characterized by an initial size of approximately 105MB, the dataset comes from a Kaggle-type data collection and provides the context necessary to identify anomalies and behavioral inefficiencies through process mining methodologies.

### 2.2 Structure and Format (CSV/XES)

Data is managed through two main formats to ensure compatibility and readability throughout the entire analysis cycle. The CSV format contains information structured in four fundamental columns: the user session identifier is stored as Case ID, the type of action performed represents the Activity, the timestamp is recorded as Timestamp and the individual user identifier is mapped as Resource. At the end of the cleaning phase, the log is also exported in XES (eXtensible Event Stream) format, which is the XML standard used to ensure interoperability with professional Process Mining software such as ProM. This dual structure allows you to maintain a data representation that is both accessible for quick tabular manipulations and compliant with international standards for event log exchange.

### 2.3 Event Log Statistics

Event log statistics reflect the impact of filtering and cleaning operations performed to ensure a representative but manageable volume of data. The final log used for the analysis is composed of 5000 unique cases, corresponding to the first sessions extracted from the original dataset to maintain the computational efficiency of the discovery algorithms. Within these sessions, the main activities of the e-commerce flow are monitored, with an average duration of cases calculated in seconds which allows us to evaluate the speed of user interaction in completing the purchasing journey.

The distribution of activities highlights the most frequent paths, defined as variants, where the common sequences indicate standard navigation behavior, while the less recurring variants

signal potential anomalies or bottlenecks in the user experience. This statistical data provides the quantitative basis necessary to feed the LLM reasoning engine and to calculate the quality metrics of the extracted models.

## 3. Methodology

The methodology adopted is divided into three sequential phases which guarantee the transformation of raw data into structured business knowledge.

### 3.1 Data Preprocessing (Cleaning, Filtering e Sessionizzazione)

The process began with the execution of the `preProcessing.py` script, dedicated to cleaning and filtering the original log. In this phase, incomplete sessions were removed and timestamps normalized to ensure temporal consistency of events. A targeted filter was applied to limit the analysis to 5000 unique cases, thus ensuring computational stability during model generation. Finally, the data was mapped according to Process Mining standards, renaming the original columns to Case ID, Activity and Timestamp to ensure compatibility with analysis tools.

### 3.2 Process Discovery algorithms (Alpha, Heuristic, Inductive)

Based on the clean log, three core algorithms were applied for process model extraction: Alpha Miner, Heuristic Miner, and Inductive Miner. The Alpha Miner was used to identify direct dependency relationships between e-commerce activities, while the Heuristic Miner made it possible to manage the intrinsic noise of real data by filtering the least frequent paths. The Inductive Miner was instead used to ensure the generation of correct, structurally sound and deadlock-free models, providing a hierarchical view of user behavior.

### 3.3 Quality Evaluation Metrics

The final phase of the methodology concerned the objective evaluation of the models through the `dashboard.py` script. To determine the reliability of the obtained representations, four essential metrics were calculated: Fitness to verify the capacity of the model to correctly reproduce the log traces, Precision to measure its specificity relative to the observed data, Generalization to evaluate its robustness on future behaviors, and Simplicity to quantify the structural readability of the resulting graph.

## 4. Results and Findings

The execution of the discovery processes and the subsequent analysis phase made it possible to extract fundamental numerical and structural evidence for understanding the analyzed e-commerce flow.

### 4.1 Comparative Analysis of Model Performance

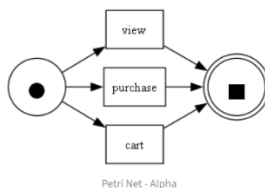
The comparative analysis conducted through the dashboard highlighted substantial differences between the three algorithms applied to the dataset. The Alpha algorithm showed a perfect precision score of 1.00, but a fitness limited to 0.69, highlighting an excessively rigid structure that fails to correctly represent a significant part of the real traces present in the log. On the contrary, the Heuristic Miner significantly increased the reliability of the model, reaching a fitness of 0.99 and maintaining a high precision of 0.92, demonstrating that it is able to effectively manage the noise typical of online interactions. Finally, the Inductive Miner returned the maximum fitness value, equal to 1.00, guaranteeing total reproducibility of all observed cases; however, this result involves a decline in structural simplicity, which reaches the minimum value of 0.0145 due to an extremely articulated and complex Petri net.

#### Algoritmo: Alpha

Fitness  
0.69

Precision  
1.00

Simplicity  
0.0909

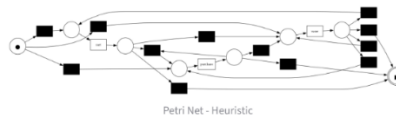


#### Algoritmo: Heuristic

Fitness  
0.99

Precision  
0.92

Simplicity  
0.0185

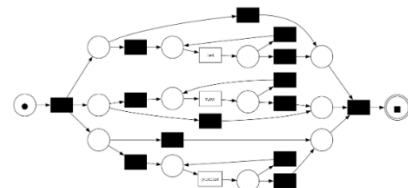


#### Algoritmo: Inductive

Fitness  
1.00

Precision  
0.90

Simplicity  
0.0145



### 4.2 Characteristics of Discovered and Variant Processes

The characteristics of the processes discovered reflect a Customer Journey focused on three key activities: viewing the product (view), adding it to the cart (cart) and finalizing the order (purchase). Through variant extraction, it was possible to identify the linear sequence from interest to purchase as the dominant path, but the models also revealed more complex dynamics, such as recursive loops on product displays that indicate long research phases before the decision. The detection of incomplete variants, where the process stops after the cart without reaching the transaction, made it possible to precisely map the abandonment points of the platform. These behavioral patterns, combined with the structural complexity

revealed by the Heuristic and Inductive models, provide the basis for identifying operational bottlenecks that impact final conversions.

# 5. Reasoning and Analysis via LLM

The integration of the Large Language Model Llama 3 via the Groq API has allowed us to elevate the analysis beyond simple numerical computation, transforming raw metrics into a strategic evaluation of the business process.

## 5.1 Generation of Strategic Reports

The AI system produced a detailed analytical report comparing the performance of the three mining algorithms used in the project. The report highlights how the Alpha algorithm, despite having excellent precision, shows satisfactory but not optimal fitness, suggesting that its rigid structure does not fully capture the variability of real data. In contrast, the Heuristic and Inductive algorithms are described as excellent for log replay capabilities, although the report reports a drastic reduction in simplicity score. This strategic document serves as a basis for management, providing specific recommendations such as optimizing data quality for simpler models and verifying structural completeness for more complex ones, ensuring that every aspect of the e-commerce process is properly monitored.

### AI Analyst - Anomalie e Ottimizzazione

Generami un reoprt

#### Rapporto di Analisi dei Dati di Process Mining

##### Introduzione

Il presente rapporto analizza i dati di Process Mining forniti, con l'obiettivo di identificare colli di bottiglia e anomalie nei processi aziendali. Le metriche fornite sono relative a tre diversi algoritmi di Process Mining: Alpha, Heuristic e Inductive.

##### Metriche correnti

Le metriche fornite sono le seguenti:

- **Alpha:**
  - Fit: 0,69
  - Prec: 1,00
  - Simp: 0,0909
- **Heuristic:**
  - Fit: 0,99
  - Prec: 0,92
  - Simp: 0,0185
- **Inductive:**
  - Fit: 1,00
  - Prec: 0,90
  - Simp: 0,0145

##### Analisi dei risultati

Dai dati forniti, emergono alcune osservazioni interessanti:

- L'algoritmo **Heuristic** sembra essere il più preciso (Prec: 0,92), ma ha un valore di semplicità (Simp) molto basso (0,0185), il che potrebbe indicare che il modello è troppo complesso e difficile da interpretare.
- L'algoritmo **Inductive** ha un valore di adattamento (Fit) perfetto (1,00), ma un valore di precisione (Prec) leggermente più basso (0,90) rispetto all'algoritmo Heuristic.
- L'algoritmo **Alpha** ha un valore di adattamento (Fit) più basso (0,69) rispetto agli altri due algoritmi, ma un valore di precisione (Prec) più alto (1,00).

##### Colli di bottiglia e anomalie

Dai dati forniti, possono essere identificati alcuni colli di bottiglia e anomalie:

- Il valore di semplicità (Simp) molto basso dell'algoritmo Heuristic potrebbe indicare che il modello è troppo complesso e difficile da interpretare, il che potrebbe essere un collo di bottiglia nel processo di analisi.
- Il valore di adattamento (Fit) più basso dell'algoritmo Alpha potrebbe indicare che il modello non è in grado di rappresentare accuratamente il processo, il che potrebbe essere un'anomalia nel processo di analisi.
- Il valore di precisione (Prec) leggermente più basso dell'algoritmo Inductive potrebbe indicare che il modello non è in grado di rappresentare accuratamente il processo, il che potrebbe essere un collo di bottiglia nel processo di analisi.

##### Conclusioni

In base ai dati forniti, emerge che l'algoritmo Heuristic sembra essere il più preciso, ma ha un valore di semplicità molto basso, il che potrebbe indicare che il modello è troppo complesso e difficile da interpretare. L'algoritmo Inductive ha un valore di adattamento perfetto, ma un valore di precisione leggermente più basso rispetto all'algoritmo Heuristic. L'algoritmo Alpha ha un valore di adattamento più basso e un valore di precisione più alto.

Per superare i colli di bottiglia e le anomalie identificate, potrebbe essere necessario:

- Semplificare il modello dell'algoritmo Heuristic per renderlo più interpretabile.
- Ottimizzare il modello dell'algoritmo Inductive per migliorare la precisione.
- Analizzare ulteriormente il modello dell'algoritmo Alpha per capire le cause del valore di adattamento più basso.

Inoltre, potrebbe essere utile analizzare ulteriormente i dati per identificare altre anomalie o colli di bottiglia nel processo di analisi.



## 5.2 Identification of Anomalies and Bottlenecks

Through automated reasoning, significant anomalies in model performance were identified, particularly regarding the discrepancy between fitness and accuracy in the Inductive algorithm. The AI assistant interpreted this gap as a signal that the model, while capable of representing the observed process, may lack accuracy in predicting paths not yet charted. Another critical bottleneck was identified in the low fitness of the Alpha algorithm, interpreted not only as an algorithmic limitation but as a symptom of potential poor quality in the input data. These analyzes revealed common sticking points, including the lack of detailed information in some log steps and the oversimplification of heuristic models, which risk overlooking key steps in the customer journey.



Identifica anomalie e colli di bottiglia



Identificazione di anomalie e colli di bottiglia

Dai dati forniti, emergono alcune anomalie e colli di bottiglia:

### Anomalie

1. **Valore di adattamento (Fit) di Alpha:** 0,69 (più basso rispetto agli altri due algoritmi)
  - Ciò potrebbe indicare che il modello di Alpha non è in grado di rappresentare accuratamente il processo.
2. **Valore di precisione (Prec) di Alpha:** 1,00 (più alto rispetto agli altri due algoritmi)
  - Ciò potrebbe indicare che il modello di Alpha è troppo semplice o che ci sono errori nel processo di analisi.
3. **Valore di semplicità (Simp) di Heuristic:** 0,0185 (molto basso)
  - Ciò potrebbe indicare che il modello di Heuristic è troppo complesso e difficile da interpretare.

### Colli di bottiglia

1. **Complessità del modello di Heuristic:** il valore di semplicità (Simp) basso potrebbe indicare che il modello è troppo complesso e difficile da interpretare.
2. **Errori nel processo di analisi di Alpha:** il valore di adattamento (Fit) basso e il valore di precisione (Prec) alto potrebbero indicare errori nel processo di analisi.
3. **Limitazioni del modello di Inductive:** il valore di precisione (Prec) leggermente più basso potrebbe indicare limitazioni del modello.

### Tipologia di anomalie

1. **Anomalie di adattamento:** il valore di adattamento (Fit) di Alpha è più basso rispetto agli altri due algoritmi.
2. **Anomalie di precisione:** il valore di precisione (Prec) di Alpha è più alto rispetto agli altri due algoritmi.
3. **Anomalie di complessità:** il valore di semplicità (Simp) di Heuristic è molto basso.

### Tipologia di colli di bottiglia

1. **Colli di bottiglia di complessità:** il modello di Heuristic è troppo complesso e difficile da interpretare.
2. **Colli di bottiglia di errore:** il processo di analisi di Alpha potrebbe contenere errori.
3. **Colli di bottiglia di limitazione:** il modello di Inductive potrebbe avere limitazioni.

## 5.3 Process Optimization Strategies

Based on the critical issues identified, an optimization strategy structured into four operational phases distributed over a six-month period was outlined. The first phase involves enhancing data quality collection and analysis, ensuring that datasets are complete and accurate for subsequent phases. The second and third phases focus on developing more complex models

via machine learning techniques and optimizing critical parameters to improve the stability and accuracy of the representations. Finally, the last phase introduces a continuous implementation and monitoring protocol to periodically update the models based on the evolution of purchasing behaviors. The expected results from this strategy include a clear reduction in data errors, an increase in analytical precision and a general improvement in the platform's ability to faithfully represent real operational flows.



Dammi delle strategie di ottimizzazione



Strategie di ottimizzazione

Per ottimizzare i processi di analisi e superare le anomalie e i colli di bottiglia identificati, possono essere utilizzate le seguenti strategie:

#### Strategie di ottimizzazione per Alpha

1. **Riduzione della complessità del modello:** semplificare il modello di Alpha per renderlo più interpretabile e migliorare il valore di adattamento (Fit).
2. **Introduzione di nuovi dati:** aggiungere nuovi dati per migliorare la precisione del modello di Alpha e ridurre il valore di precisione (Prec) esagerato.
3. **Utilizzo di tecniche di regolarizzazione:** utilizzare tecniche di regolarizzazione per ridurre la complessità del modello di Alpha e migliorare il valore di adattamento (Fit).

#### Strategie di ottimizzazione per Heuristic

1. **Semplificazione del modello:** semplificare il modello di Heuristic per renderlo più interpretabile e migliorare il valore di semplicità (Simp).
2. **Introduzione di nuovi parametri:** aggiungere nuovi parametri per migliorare la precisione del modello di Heuristic e ridurre il valore di complessità.
3. **Utilizzo di tecniche di ottimizzazione:** utilizzare tecniche di ottimizzazione per trovare i parametri ottimali per il modello di Heuristic.

#### Strategie di ottimizzazione per Inductive

1. **Miglioramento della precisione:** migliorare la precisione del modello di Inductive per ridurre il valore di precisione (Prec) leggermente più basso.
2. **Introduzione di nuovi dati:** aggiungere nuovi dati per migliorare la precisione del modello di Inductive e ridurre il valore di precisione (Prec) leggermente più basso.
3. **Utilizzo di tecniche di selezione dei dati:** utilizzare tecniche di selezione dei dati per selezionare i dati più rilevanti per il modello di Inductive.

#### Strategie di ottimizzazione generiche

1. **Utilizzo di tecniche di cross-validazione:** utilizzare tecniche di cross-validazione per valutare la precisione dei modelli e ridurre il valore di precisione (Prec) esagerato.
2. **Utilizzo di tecniche di bootstrap:** utilizzare tecniche di bootstrap per valutare la stabilità dei modelli e ridurre il valore di complessità.
3. **Utilizzo di tecniche di ottimizzazione dei parametri:** utilizzare tecniche di ottimizzazione dei parametri per trovare i parametri ottimali per i modelli.

#### Benefici delle strategie di ottimizzazione

1. **Miglioramento della precisione:** le strategie di ottimizzazione possono migliorare la precisione dei modelli e ridurre il valore di precisione (Prec) esagerato.
2. **Riduzione della complessità:** le strategie di ottimizzazione possono ridurre la complessità dei modelli e migliorare il valore di semplicità (Simp).
3. **Miglioramento della stabilità:** le strategie di ottimizzazione possono migliorare la stabilità dei modelli e ridurre il valore di complessità.

## 6. Predictive Analysis

The implementation of a predictive module within the dashboard responds to the need to anticipate subsequent events to optimize the user experience and maximize commercial conversions. By exploiting the structure of the previously processed event log, the system is able to analyze the transition probabilities between the different e-commerce activities, providing a quantitative basis which is then enriched by the qualitative reasoning capacity of artificial intelligence.

### 6.1 Prediction of Next Activity

Through the dedicated functionality in the dashboard, the system allows you to insert individual activities or complex sequences, such as the complete path view -> cart -> purchase, to determine the most probable future action. Regardless of the complexity of the entered path, the statistical results extracted from the log show that the most frequently expected next activity is often a new view phase. This occurrence was interpreted by the AI assistant as a manifestation of the natural tendency of users to explore complementary products, search for additional information to confirm their choices or navigate in a non-linear way between the different categories of the website.

Seleziona il percorso attuale dell'utente:

view x

Esegui Predizione

Prossima attività predetta: view

Analisi del percorso dell'utente e suggerimenti per aumentare la conversione

1. Perché questa predizione è sensata in un contesto E-commerce?

La predizione che l'evento successivo più frequente dopo "view" sia nuovamente "view" è sensata in un contesto E-commerce per diverse ragioni:

- Gli utenti tendono a esplorare diversi prodotti e categorie prima di effettuare un acquisto, quindi è probabile che continuino a visualizzare altre pagine di prodotti.
- La fase di "view" rappresenta una fase di ricerca e valutazione, durante la quale gli utenti stanno ancora decidendo se effettuare un acquisto o meno.
- La navigazione all'interno del sito web è spesso non lineare, quindi gli utenti possono tornare indietro e visualizzare nuovamente pagine che hanno già visitato in precedenza.
- La "view" può anche rappresentare una fase di confronto tra diversi prodotti, quindi gli utenti possono visualizzare più pagine di prodotti per confrontare caratteristiche e prezzi.

2. Quale azione di business suggeriresti per aumentare la conversione in questo punto?

Per aumentare la conversione in questo punto, suggerisco di implementare le seguenti azioni di business:

- **Personalizzazione:** utilizzare i dati di navigazione e di acquisto degli utenti per personalizzare le offerte e i contenuti del sito web. Ad esempio, suggerire prodotti simili a quelli già visualizzati o offrire sconti sui prodotti che l'utente ha già guardato.
- **Facilitazione della ricerca:** semplificare la ricerca dei prodotti e delle categorie per aiutare gli utenti a trovare ciò che stanno cercando più rapidamente. Ciò può essere fatto implementando una funzione di ricerca avanzata o migliorando la navigazione del sito web.
- **Incentivi per l'acquisto:** offrire incentivi per l'acquisto, come ad esempio sconti, gratis spedizione o punti fedeltà, per motivare gli utenti a completare l'acquisto.
- **Call-to-Action (CTA) chiaro e visibile:** assicurare che il CTA per l'acquisto sia chiaro e visibile sulla pagina di prodotto, in modo che gli utenti sappiano cosa fare per completare l'acquisto.
- **Rimozione degli ostacoli:** identificare e rimuovere gli ostacoli che potrebbero impedire agli utenti di completare l'acquisto, come ad esempio un processo di checkout complesso o una mancanza di opzioni di pagamento.

Implementando queste azioni di business, è possibile aumentare la conversione e migliorare l'esperienza dell'utente sul sito web, riducendo la probabilità che gli utenti continuino a visualizzare pagine senza completare l'acquisto.

cart ×

Esegui Predizione

Prossima attività predetta: view

1. Perché questa predizione è sensata in un contesto E-commerce?

La predizione che l'evento successivo più frequente dopo "cart" (aggiungi al carrello) sia "view" (visualizza) è sensata in un contesto E-commerce per diverse ragioni:

- **Conferma della scelta:** Dopo aver aggiunto un articolo al carrello, l'utente potrebbe voler verificare le caratteristiche del prodotto, controllare il prezzo, le dimensioni, i colori disponibili, o leggere le recensioni degli altri clienti per confermare la propria scelta di acquisto. Questo comportamento è comune, poiché gli utenti vogliono essere sicuri di avere fatto la scelta giusta prima di procedere all'acquisto.
- **Ricerca di informazioni aggiuntive:** Gli utenti potrebbero anche voler esplorare altri prodotti simili o accessori al prodotto che hanno appena aggiunto al carrello. La visualizzazione di altri prodotti può aiutare gli utenti a trovare opzioni alternative o complementari, aumentando le possibilità di acquisto aggiuntivo o di sostituzione.
- **Pianificazione dell'acquisto:** In alcuni casi, gli utenti potrebbero aggiungere prodotti al carrello per pianificare futuri acquisti o per confrontare prezzi e caratteristiche tra diversi prodotti. La fase di visualizzazione after-aggiunta al carrello può aiutare gli utenti a prendere decisioni più informate.

2. Quale azione di business suggeriresti per aumentare la conversione in questo punto?

Per aumentare la conversione in questo punto, si potrebbero implementare le seguenti azioni di business:

- **Semplificazione del processo di acquisto:** Assicurarsi che il processo di acquisto sia il più lineare e facile possibile, riducendo gli step necessari per concludere l'acquisto. Questo può includere la possibilità di comprare come ospite, senza la necessità di creare un account.
- **Informazioni chiare e complete:** Fornire informazioni dettagliate e facilmente accessibili sui prodotti, come descrizioni, specifiche tecniche, recensioni degli utenti e immagini ad alta qualità. Questo aiuta gli utenti a prendere decisioni d'acquisto più rapide e riduce le probabilità di abbandono del carrello.
- **Promozioni e incentivi:** Offrire promozioni speciali, sconti o voucher di sconto per gli articoli nel carrello. Questo può incentivare gli utenti a completare l'acquisto e aumentare il valore medio dell'ordine.
- **Raccomandazioni di prodotti:** Fornire suggerimenti di prodotti correlati o complementari a quelli già aggiunti al carrello. Questo può aumentare le vendite incrociate e migliorare l'esperienza dell'utente, rendendo più facile la scoperta di nuovi prodotti.
- **Feedback e assistenza:** Offrire un'eccellente assistenza clienti, tramite canali come il live chat, la posta elettronica o il servizio clienti telefonico. La disponibilità di aiuto immediato può rassicurare gli utenti esitanti e aumentare la fiducia nell'acquisto.

Implementare queste strategie può aiutare a ridurre l'abbandono del carrello e aumentare le tasse di conversione, migliorando nel complesso l'esperienza di acquisto online.

purchase ×

Esegui Predizione

Prossima attività predetta: view

1. Perché questa predizione è sensata in un contesto E-commerce?

La predizione che l'evento successivo più frequente dopo "purchase" (acquisto) sia "view" (visualizzazione) potrebbe sembrare inizialmente controintuitiva, poiché potremmo aspettarci che dopo un acquisto l'utente si trovi in una fase di post-acquisto, come ad esempio la ricezione dell'ordine o il feedback sul prodotto. Tuttavia, ci sono diverse ragioni per cui questa predizione potrebbe essere sensata:

- **Ricerca di prodotti complementari:** dopo aver effettuato un acquisto, un utente potrebbe essere interessato a cercare prodotti complementari o accessori che possano integrarsi con l'articolo acquistato. Ad esempio, se un utente ha acquistato un telefono cellulare, potrebbe poi cercare un caso o un set di cuffie compatibili.
- **Conferma dell'ordine:** gli utenti potrebbero tornare sul sito per verificare lo stato del loro ordine o per visualizzare le informazioni relative all'acquisto appena effettuato.
- **Navigazione post-acquisto:** alcuni utenti potrebbero semplicemente continuare a navigare sul sito, esplorando altre sezioni o categorie di prodotti, anche se non hanno intenzione di effettuare un altro acquisto immediato.

2. Quale azione di business suggeriresti per aumentare la conversione in questo punto?

Per aumentare la conversione in questo punto, suggerirei le seguenti azioni di business:

- **Offerte di prodotti complementari:** sulla base degli acquisti precedenti, offrire prodotti complementari o accessori che possano integrarsi con l'articolo acquistato. Ciò può essere fatto tramite e-mail difollow-up o attraverso raccomandazioni personalizzate sulla pagina di conferma dell'ordine.
- **Promozioni e sconti:** offrire promozioni o sconti su prodotti specifici per incentivare gli utenti a effettuare un altro acquisto. Ciò può essere fatto attraverso notifiche push, e-mail di marketing o banner sul sito.
- **Contenuto di valore:** fornire contenuto di valore agli utenti, come ad esempio guide, tutorial o consigli su come utilizzare i prodotti acquistati. Ciò può aiutare a costruire fiducia e lealtà nei confronti del marchio, aumentando le possibilità di futuri acquisti.
- **Semplificazione del processo di acquisto:** assicurarsi che il processo di acquisto sia semplice e intuitivo, riducendo al minimo i passaggi necessary per completare un ordine. Ciò può aiutare a ridurre la frustrazione e a aumentare la probabilità di conversione.

In sintesi, la predizione statistica che gli utenti che hanno effettuato un acquisto siano più likely di visualizzare altre pagine del sito è sensata in un contesto e-commerce, poiché gli utenti potrebbero cercare prodotti complementari, verificare lo stato del loro ordine o continuare a navigare sul sito. Per aumentare la conversione in questo punto, si possono implementare strategie come offerte di prodotti complementari, promozioni e sconti, contenuto di valore e semplificazione del processo di acquisto.

view × cart × purchase ×

Esegui Predizione

Prossima attività predetta: view

1. Perché questa predizione è sensata in un contesto E-commerce?

Sono felice di aiutarvi!

La predizione che l'evento successivo più frequente sia "view" dopo il percorso "view → cart → purchase" è sensata per diversi motivi:

- **Comportamento dell'utente:** Gli utenti che hanno già completato un acquisto potrebbero essere interessati a cercare altri prodotti simili o a esplorare le offerte del negozio online.
- **Ciclo di acquisto:** Il ciclo di acquisto spesso include una fase di ricerca e valutazione dei prodotti, quindi gli utenti potrebbero tornare alla fase di "view" per confrontare prezzi, caratteristiche o leggere recensioni di altri prodotti.
- **Fattori psicologici:** Gli utenti potrebbero sentire il bisogno di confermare la loro scelta di acquisto o di cercare ulteriori informazioni per ridurre il rimpianto dell'acquisto (post-purchase dissonance).

In un contesto E-commerce, questo comportamento è comune e può essere influenzato da fattori come la qualità del prodotto, la soddisfazione del cliente e la facilità di navigazione del sito web.

2. Quale azione di business suggeriresti per aumentare la conversione in questo punto?

Per aumentare la conversione in questo punto, suggerirei le seguenti azioni di business:

- **Personalizzazione:** Utilizzare tecniche di personalizzazione per mostrare ai clienti prodotti simili a quelli che hanno già acquistato, aumentando le possibilità di una nuova vendita.
- **Recommandazioni:** Implementare un sistema di raccomandazioni di prodotti basato sul comportamento di acquisto e sulle preferenze del cliente.
- **Promozioni e offerte speciali:** Offrire promozioni e offerte speciali ai clienti che hanno già effettuato un acquisto, aumentando la loro lealtà e la probabilità di un nuovo acquisto.
- **E-mail marketing:** Utilizzare e-mail marketing per mantenere i clienti informati sulle nuove offerte e promozioni, aumentando la probabilità di un ritorno sul sito web.
- **Ottimizzazione del sito web:** Ottimizzare il sito web per migliorare la velocità di caricamento, la facilità di navigazione e la chiarezza delle informazioni, riducendo le barriere all'acquisto.

Inoltre, potrebbe essere utile analizzare i dati di abbandono dei carrelli e i motivi di abbandono per identificare aree di miglioramento nel processo di acquisto e aumentare la conversione in questo punto.

## 6.2 Behavior Patterns and Transition Probability

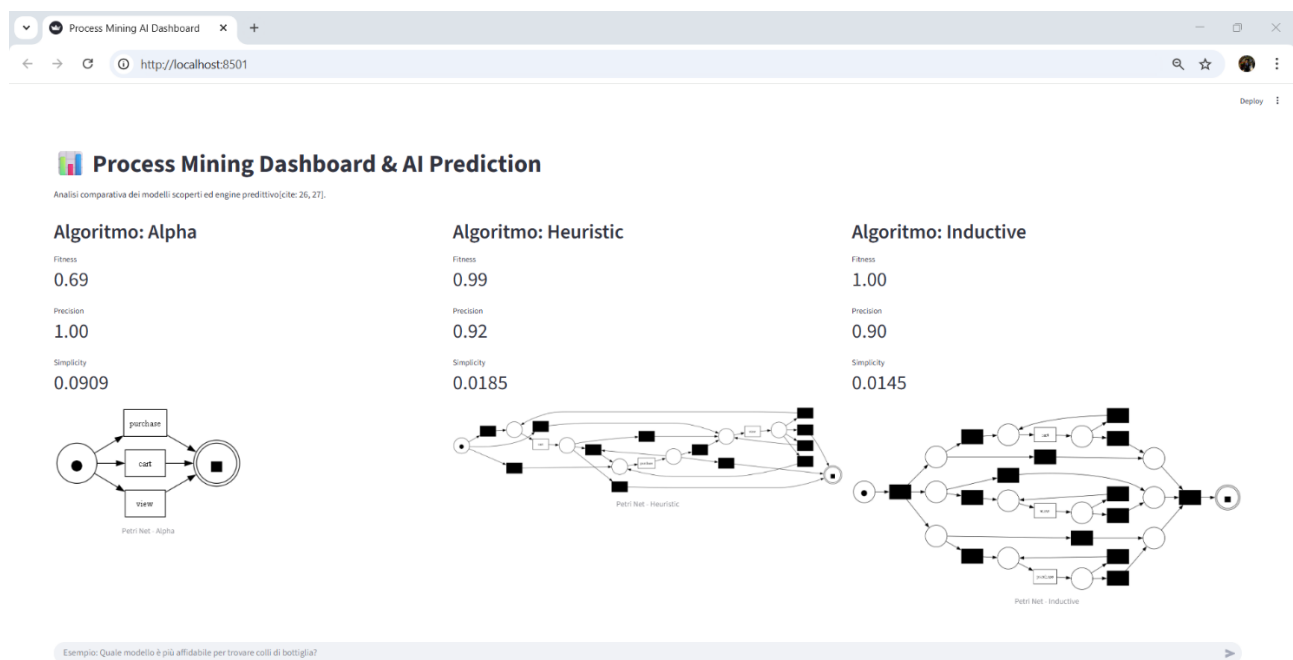
The analysis of behavioral patterns highlighted how transitions within the platform do not always follow a purely linear path towards the transaction, reflecting the exploratory nature of online navigation. For example, after a completed purchase, statistical probability suggests a return to viewing the products, a behavior that the LLM attributes to the search for accessories or to the user's desire to reassure himself about the choice made by consulting the technical specifications or reviews. To capitalize on these transition patterns, the business strategy suggested by artificial intelligence involves the implementation of cross-selling actions and personalized recommendations based on browsing history, transforming potential browsing loops into concrete opportunities to increase the average value of each user session.

## 7. Implementation: Interactive Dashboard

The last section of the documentation describes the operating environment created to make the results of the Process Mining analysis accessible through a centralized and interactive interface.

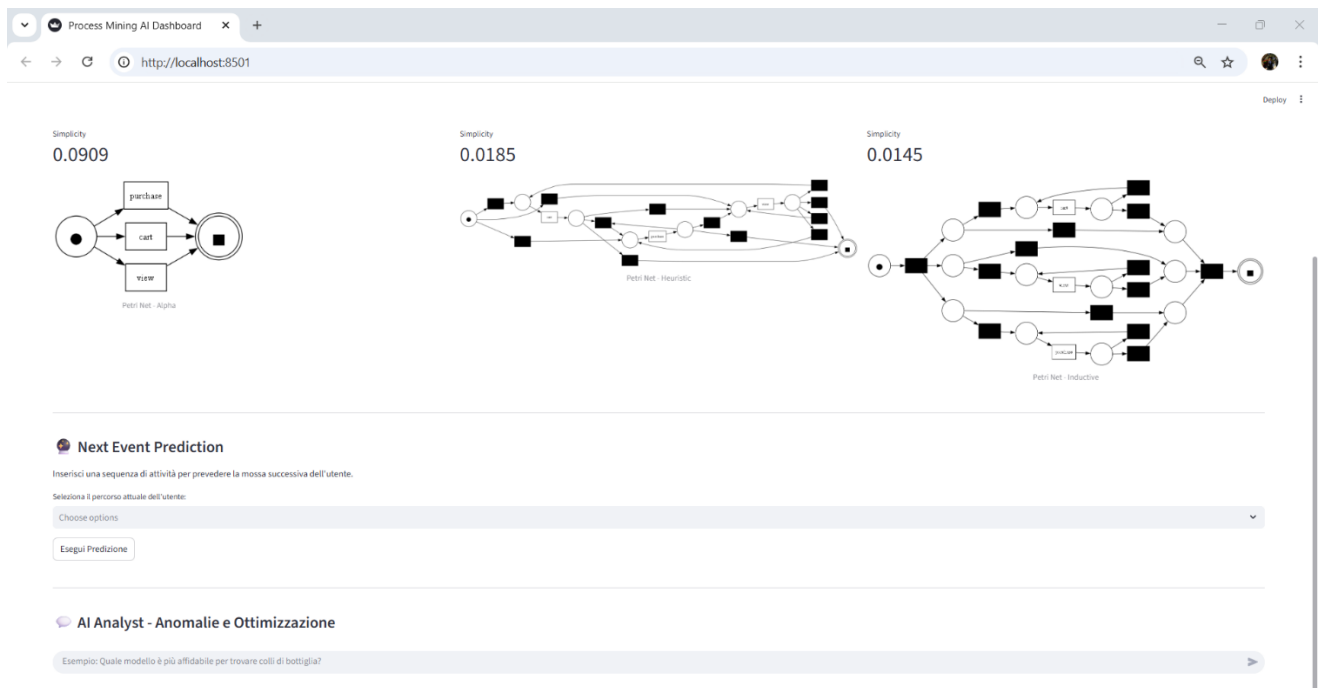
### 7.1 Technology Stack and System Architecture

The system architecture was designed to integrate different stages of the Process Mining lifecycle into a single Streamlit-based web application. The application core is written in Python, using the pm4py library for managing event logs and executing discovery algorithms. For the artificial intelligence component, the Groq API was used to interface with the Llama 3 model, ensuring fast response times for the reasoning module. The structure follows a modular flow: the `preProcessing.py` script prepares the data, `processDiscovery.py` generates the models, and `dashboard.py` acts as an orchestrator for visualization and user interaction.



### 7.2 Chatbot interface and AI assistant

The integration of the AI assistant ("AI Analyst") represents the heart of the reasoning component of the project, as required by the specifications. The chatbot is not a simple textual interface, but a system powered by the dynamic context of metrics extracted from Alpha, Heuristic and Inductive models. Through a prompt engineering system, Llama 3 receives Fitness, Precision and Simplicity data to autonomously identify structural anomalies or operational bottlenecks. Users can query the assistant to obtain strategic reports, explanations on unusual patterns or suggestions for process optimization, making the analysis accessible even to non-technical figures.



## 7.3 Guide to Using and Viewing KPIs

The dashboard interface is structured into distinct sections to facilitate data exploration. The upper part presents an immediate comparative view of the three mining algorithms, where the fundamental KPIs are shown for each: Fitness, Precision, Generalization and Simplicity Score. Below the numerical data, the corresponding Petri Nets are dynamically rendered, allowing a direct visual comparison of the model's complexity. The middle section hosts the "Next Event Prediction" module, where the user can build customized paths (e.g. view -> cart) to obtain a statistical and narrative prediction of the next activity. Finally, the lower part is dedicated to the interactive chatbot for in-depth anomaly analysis.

## 8. Conclusions and Future Work

The project successfully demonstrated the effectiveness of the integration between traditional Process Mining methodologies and the advanced analytical capabilities of Large Language Models for the optimization of an e-commerce process. Through the development of an interactive dashboard based on Streamlit, it was possible not only to discover and visualize real operational flows via algorithms such as Alpha, Heuristic and Inductive Miner, but also to subject them to a rigorous quantitative evaluation based on the four fundamental quality metrics. The analysis highlighted how high-fitness models, such as the Inductive one, are essential for representing the complexity of user interactions, providing a solid basis for the reasoning module. Llama 3's contribution was instrumental in transforming raw data into strategic insights, allowing the identification of bottlenecks and the proposal of targeted business solutions that go beyond simple statistical analysis.

Despite the positive results obtained in mapping and predicting activities, the work paves the way for several future evolutions to make the system even more robust and dynamic. A first relevant development concerns the implementation of incremental learning, which would allow the model to automatically update when new data arrives without having to process the entire log from scratch, thus responding to the concept of continuous review of the process. Furthermore, prediction accuracy could be improved by integrating deep learning algorithms, such as recurrent neural networks, to capture long-term dependencies in user journeys that escape current analyzes based on immediate transition probabilities. Finally, expanding the system towards real-time monitoring would enable immediate proactive interventions, such as sending recovery notifications for abandoned carts or instant personalized offers, maximizing the operational impact of the entire Process Mining architecture.