

DOI: 10.37791/2687-0649-2026-21-2-27-41

EDN: VCLNLX

Научная статья

ВАК 5.2.2

ГРНТИ 06.39.27, 06.39.21

УДК 005.5:004.62

# Процессная аналитика бизнес-процессов на основе статистической фильтрации и адаптивного синтеза модели в условиях неполноты и зашумленности данных

*А. И. Дереза<sup>1,2</sup>, Н. Д. Лукьянов<sup>2\*</sup>*

<sup>1</sup>ПАО Сбербанк России, Иркутск, Россия

<sup>2</sup>Иркутский национальный исследовательский технический университет, Иркутск, Россия

\*lukyanovnd@istu.edu

**Аннотация.** Существующие подходы технологии при глубоком анализе бизнес-процессов часто демонстрируют низкую устойчивость к распространенным проблемам реальных данных – наличию шумовых записей и неполноте журналов событий. Целью работы является разработка и верификация комплексного трехэтапного подхода, объединяющего статистическую фильтрацию аномалий, вероятностно-временное восстановление пропущенных событий и адаптивный синтез модели процесса. В рамках исследования решались следующие задачи: критический анализ классических алгоритмов Discovery, формализация методов фильтрации на основе Isolation Forest и восстановления событий с использованием вероятностных и временных метрик, а также предложен адаптивный механизм выбора порога шума на основе нормированной энтропии вариативности подпроцесса. Подход реализован в виде программного модуля на Python с использованием библиотек `rt4py`, `scikit-learn` и `numpy`. Эксперименты на синтетических данных, сгенерированных с варьируемым уровнем шума и долей пропущенных событий, подтвердили устойчивость метода. Результаты оценивались по метрикам Fitness, Generalization, Simplicity и F1-мера в сравнении с алгоритмами Alpha Miner, Heuristics Miner и Inductive Miner. Предложенный подход показал статистически значимое улучшение качества результирующих моделей в условиях высокой зашумленности и неполноты исходных данных. Результаты работы формируют основу для создания устойчивых систем анализа бизнес-процессов, способных работать с данными из реальных информационных систем.

**Ключевые слова:** процессная аналитика, наука о данных, управление бизнес-процессами, интеллектуальный анализ процессов, событийные потоки, журналы событий, восстановление процессов, обнаружение аномалий, шум в данных

**Для цитирования:** Дереза А.И., Лукьянов Н.Д. Процессная аналитика бизнес-процессов на основе статистической фильтрации и адаптивного синтеза модели в условиях неполноты и зашумленности данных // Прикладная информатика. 2026. Т. 21. № 2. С. 27–41. DOI: 10.37791/2687-0649-2026-21-2-27-41.

© Дереза А.И., Лукьянов Н.Д., 2026.

# A process mining approach based on statistical filtering and adaptive process model synthesis under incomplete and noisy event logs

A. Dereza<sup>1,2</sup>, N. Lukyanov<sup>2\*</sup>

<sup>1</sup>Sberbank Public Joint Stock Company, Irkutsk, Russia

<sup>2</sup>Irkutsk National Research Technical University, Irkutsk, Russia

\*lukyanovnd@istu.edu

**Abstract.** Existing Process Mining techniques often exhibit low robustness to common real-world data issues, such as noisy traces and incomplete event logs. This paper presents the development and validation of an integrated three-stage approach that combines statistical anomaly filtering, probabilistic and temporal reconstruction of missing events, and adaptive process model synthesis. The study addresses the following tasks: a critical review of classical process discovery algorithms; formalization of filtering methods based on Isolation Forest and event reconstruction using probabilistic and temporal metrics; and the design of an adaptive mechanism for selecting the noise threshold based on the normalized entropy of subprocess variability. The approach is implemented as a Python software module using the pm4py, scikit-learn, and NumPy libraries. Experiments on synthetic datasets generated with varying noise levels and proportions of missing events confirm the robustness of the proposed method. The results are evaluated using the Fitness, Generalization, Simplicity, and F1-score metrics and compared against the Alpha Miner, Heuristics Miner, and Inductive Miner algorithms. The proposed approach yields a statistically significant improvement in the quality of the resulting process models under high noise and log incompleteness, providing a basis for robust business process analytics systems capable of operating on data from real information systems.

**Keywords:** Process Mining, Data Science, Business Process Management, process intelligence, event streams, event logs, process reconstruction, anomaly detection, data noise

**For citation:** Dereza, A., & Lukyanov, N. (2026). A process mining approach based on statistical filtering and adaptive process model synthesis under incomplete and noisy event logs. *Journal of Applied Informatics*, 21(2), 27–41. <https://doi.org/10.37791/2687-0649-2026-21-2-27-41>

© Dereza A., Lukyanov N., 2026.

## Введение

Современные корпоративные информационные системы регистрируют большие объемы данных о выполнении бизнес-процессов в виде журналов событий. Эти журналы содержат

цифровые следы операций, представляющие собой источник для анализа и оптимизации деятельности организаций. Глубокий анализ бизнес-процессов, или процессная аналитика (Process Mining, PM) – это направление на пересечении науки о данных (Data Science) и управле-