

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра компьютерных технологий и систем

Аннотация к дипломной работе
«Алгоритмические и технические аспекты интеграции средств
кластеризации, сопоставление программных реализаций»

Лебедев Артем Владимирович

Научный руководитель – профессор, доктор
физико-математических наук
В.Б. Таранчук

Минск, 2025

АННОТАЦИЯ

Диплом содержит: 51 страницу, 17 иллюстраций, 20 формул, 14 использованных литературных источников.

Ключевые слова: АЛГОРИТМЫ, РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ, КЛАСТЕРИЗАЦИЯ, ACTIVE LEARNING, MLOps, AIRFLOW, CLICKHOUSE, MLFLOW, OPTUNA, SWING, E-COMMERCE.

Объектом исследования алгоритмы и программно-технические средства интеграции средств кластеризации при построении рекомендательных систем маркетплейса.

Целью работы является разработать воспроизводимую, масштабируемую и экономически оправданную экосистему рекомендаций, в которой кластеризация связывает разнородные хранилища данных и различные временные горизонты.

Методами исследования аналитический обзор литературы; программная реализация на Python и SQL (Airflow, ClickHouse, MLflow, Optuna); экспериментальная оценка (A/B-тестирование, статистический анализ).

Результаты работы:

- Выполнен сравнительный анализ классических и графовых алгоритмов кластеризации товаров; предложена стохастическая модификация SWING с параметром β , ускорившая сходимость на $\approx 47\%$.
- Построен гибридный алгоритм рекомендаций (товарная схожесть + кластерный контекст + промо-вес), давший рост CTR блока «С этим покупают» на 10–15 % и снижение доли категории «Прочее» с 5,7 % до 1,9 %.
- Реализован сквозной ML-конвейер *Airflow* $\rightarrow$ *ClickHouse* $\rightarrow$ *MLflow* $\rightarrow$ *Optuna/Neptune* $\rightarrow$ *Redis* с автоматическим откатом при деградации онлайн-метрик.

АНАТАЦЫЯ

Дыпломная праца ўтрымлівае: 51 старонку, 17 малюнкаў, 20 формул, 14 выкарыстаных літаратурных крыніц.

Ключавыя словы: АЛГАРЫТМЫ, РЭКАМЕНДАЦЫЙНЫЯ СІСТЭМЫ, КЛАСТАРЫЗАЦЫЯ, ACTIVE LEARNING, MLOps, AIRFLOW, CLICKHOUSE, MLFLOW, OPTUNA, SWING, E-COMMERCE.

Аб'ектам даследавання алгарытмы і праграмныя сродкі інтэграцыі кластарных метадаў пры пабудове рэкамендацыйных сервераў маркетплэйса

Мэтай даследавання: стварыць рэпрадукаваную, маштабаваную і эканамічна абгрунтаваную экасістэму рэкамендацый, дзе кластарызацыя з'яўляецца звязном паміж разнароднымі сховішчамі даных і рознымі часавымі гарызонтамі.

Вынікі:

- Прадстаўлена стохастычная мадыфікацыя SWING, што паскорыла канвергенцыю на $\approx 47\%$.
- Гібрыдны алгарытм «Basket Hybrid» павялічыў CTR на 10–15 % і скараціў долю катэгорыі «Іншае» да 1,9 %.
- Рэалізаваны скразны ML-канвеер з аўтаматычным адкатам пры пагаршэнні ан-лайн-метрык.

ANNOTATION

The diploma work contains: 51 pages, 17 illustrations, 20 formulas, 14 sources.

Keywords: ALGORITHMS, RECOMMENDER SYSTEMS, CLUSTERING, ACTIVE LEARNING, MLOps, AIRFLOW, CLICKHOUSE, MLFLOW, OPTUNA, SWING, E-COMMERCE.

The object of the research is algorithms and software tools for integrating clustering techniques in marketplace recommender systems.

The goal is to design a reproducible, scalable and economically viable recommendation ecosystem in which clustering bridges heterogeneous data stores and multiple temporal horizons.

Research methods are literature survey; software engineering in Python and SQL (Airflow, ClickHouse, MLflow, Optuna); experimental validation through A/B testing and statistical analysis.

Results:

- A stochastic SWING-based clustering algorithm with annealing parameter β accelerated convergence by ≈ 47 %.
- A hybrid “Basket Hybrid” recommender (item similarity + cluster context + promo weighting) increased widget CTR by 10–15 % and reduced the “Other” category share from 5.7 % to 1.9 %.
- An end-to-end MLOps pipeline (*Airflow* $\rightarrow$ *ClickHouse* $\rightarrow$ *MLflow* $\rightarrow$ *Optuna/Neptune* $\rightarrow$ *Redis*) with automatic rollback was implemented.