

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра вычислительной математики

Аннотация к дипломной работе

**«РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ
МОНИТОРИНГА И АУДИТА АЛГОРИТМОВ МАШИННОГО
ОБУЧЕНИЯ»**

Шпаковский Артём Денисович

Научный руководитель – доцент кафедры вычислительной математики ФПМИ
Мандрик П. А.

Минск, 2025

АННОТАЦИЯ

Дипломная работа, 50 страниц, 1 таблица, 11 иллюстраций, 23 источника.

Ключевые слова: АУДИТ МОДЕЛЕЙ, МАШИННОЕ ОБУЧЕНИЕ, КАЧЕСТВО ДАННЫХ, ИНФОРМАЦИОННАЯ СИСТЕМА, БАНКОВСКИЙ СЕКТОР, ДРИФТ ДАННЫХ, МЕТРИКИ КАЧЕСТВА, АНАЛИЗ ДАННЫХ,

Объект исследования является методы и инструменты аудита качества данных и моделей машинного обучения в контексте их применения в банковском секторе Республики Беларусь.

Предметом исследования являются специализированная информационная система качества данных и генерация регулярных отчетов.

Целью работы является разработка и реализация программного решения для автоматизации выявления артефактов в данных, минимизация рисков использования некорректных моделей машинного обучения.

Методами исследования являются анализ существующих инструментов, проектирование модульной архитектуры на Python, реализация алгоритмов расчета стабильности данных и мультиколлинеарности.

Полученные результаты и их новизна: разработана система, сокращающее время ручного аудита на 40-60% за счет автоматизации проверок качества данных, выявлены и устранены критические ограничения существующих решений к мониторингу дрейфа, включая улучшенные методы расчета показателей качества данных. Предложен новый механизм интеграции с MLOps-инфраструктурой банков, обеспечивающий хранение истории проверок и автоматические рекомендации по устранению проблем.

Достоверность материалов и результатов дипломной работы: использованные материалы и результаты дипломной работы являются достоверными. Работа выполнена самостоятельно.

Областью возможного практического применения является внедрение в банковские организации для повышения надежности моделей машинного

обучения, снижения регуляторных рисков и обеспечения соответствия стандартам внутреннего контроля.

АНАТАЦЫЯ

Дыпломная праца, 50 старонак, 1 табліца, 11 ілюстрацый, 23 крыніцы.

Ключавыя словы: АЎДЫТ МАДЭЛЯЎ, МАШЫНАЕ НАВУЧАННЕ, ЯКАСЦЬ ДАНЫХ, ІНФАРМАЦЫЙНАЯ СІСТЭМА, БАНКАЎСКІ СЕКТАР, ДРЫФТ ДАДЗЕННЫХ, МЕТРЫКІ ЯКАСЦІ, АНАЛІЗ ДАДЗЕННЫХ.

Аб'ектам даследавання з'яўляюцца метады і інструменты аўдыту якасці дадзеных і мадэляў машыннага навучання ў кантэксце іх прымянення ў банкаўскім сектары Рэспублікі Беларусь.

Прадметам даследавання з'яўляюцца спецыялізаваная інфармацыйная сістэма якасці дадзеных і генерацыя рэгуляторных справаздач.

Мэтай даследавання з'яўляецца распрацоўка ПА для аўтаматызацыі выяўлення артэфактаў у дадзеных, мінімізацыі рызык выкарыстання некарэктных мадэляў МА і забеспячэння празрыстасці працэсаў машыннага навучання.

Метадамі даследавання з'яўляюцца аналіз існуючых інструментаў, праектаванне модульнай архітэктурны на Python, рэалізацыя алгарытмаў разліку стабільнасці дадзеных, а таксама візуалізацыя вынікаў.

Атрыманыя вынікі і іх навізна: распрацавана сістэма, які скарачае час ручнога аўдыту на 40-60% за кошт аўтаматызацыі праверак якасці даных, выяўлены і ліквідаваны крытычныя абмежаванні існуючых рашэнняў да маніторынгу дрыфту. Прапанаваны новы механізм інтэграцыі з MLOps-інфраструктурай банкаў, які забяспечвае захоўванне гісторыі праверак і аўтаматычныя рэкамендацыі па ўстараненню праблем.

Даставернасць матэрыялаў і вынікаў дыпломнай працы: выкарыстаныя матэрыялы і вынікі дыпломнай Працы з'яўляюцца даставернымі. Праца выканана самастойна.

Вобласцю магчымага практычнага прымянення з'яўляецца укараненне ў банкаўскія арганізацыі для павышэння надзейнасці мадэляў машыннага навучання, зніжэння рэгуляторных рызык і забеспячэнні адпаведнасці стандартам унутранага кантролю.

ANNOTATION

Diploma work, 50 pages, 1 table, 11 illustrations, 23 sources.

Keywords: MODEL AUDIT, MACHINE LEARNING, DATA QUALITY, INFORMATION SYSTEM, BANKING SECTOR, DATA DRIFT, QUALITY METRICS, DATA ANALYSIS.

The object of the research is methods and tools for auditing data quality and machine learning models in the context of their application in the banking sector of the Republic of Belarus.

The subject of the research is a specialized information system for data quality and generation of regular reports.

The purpose of the research is the development and implementation of a software solution for automating the detection of data artifacts and minimizing risks associated with the use of incorrect machine learning models.

Methods of research are analysis of existing tools, design of a modular Python architecture, and implementation of algorithms for calculating data stability and multicollinearity.

The results of the work and their novelty: the system has been developed that reduces manual audit time by 40-60% through automation of data quality checks. Critical limitations of existing data drift monitoring solutions have been identified and eliminated, including improved methods for calculating data quality metrics. A new integration mechanism with MLOps infrastructure has been proposed, ensuring audit history storage and automated troubleshooting recommendations.

Authenticity of the materials and results of the diploma work: the materials used and the results of the diploma work are authentic. The work has been put through independently.

Recommendations on the usage. Implementation in banking organizations to enhance ML model reliability, reduce regulatory risks, and ensure compliance with internal control standards.