

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**«Методика сравнения методов машинного обучения в зависимости от
различных параметров задачи»**

Вертинская Антонина Евгеньевна

Научный руководитель – кандидат физико-математических наук
Соболевская Е.П.

Минск, 2020

Реферат

Магистерская диссертация, 41 страница, 18 рисунков, 10 таблиц, 16 источников.

МАШИННОЕ ОБУЧЕНИЕ, АЛГОРИТМЫ КЛАССИФИКАЦИИ, БЕНЧМАРК-ТЕСТ, PYSPARK.ML, SKLEARN, РЕГРЕССИОННЫЕ МОДЕЛИ.

Объект исследования – алгоритмы классификации в машинном обучении, а также проблема выбора подходящего алгоритма в зависимости от параметров задачи.

Цель работы – сравнения методов машинного обучения по таким критериям, как качество предсказаний, загрузка процессора, использование памяти и диска, а также разработка методики выбора подходящего алгоритма согласно требованиям задачи.

Методы исследования – эксперимент, тестирование, анализ, сравнение.

Результаты – методика сравнения и выбора подходящего алгоритма машинного обучения в зависимости от характеристик и требований задачи. Результатом применения методики является рекомендатор алгоритмов классификации библиотек PySpark.ML и sklearn на базе регрессионных моделей, которые предсказывают метрики качества и производительности.

Область применения – сферы, где использование методов машинного обучения может быть полезно, однако такой подход недостаточно развит либо стоит больших затрат.

Abstract

Master thesis, 41 pages, 18 figures, 10 tables, 16 references.

MACHINE LEARNING, CLASSIFICATION ALGORITHMS, BENCHMARK, PYSPARK.ML, SKLEARN, REGRESSION MODELS.

Object of research – machine learning classifications algorithms and the problem of appropriate algorithm selection depending on the parameters.

Objective – classification algorithms comparison depending on the quality metrics, cpu, disk and memory usage and development of the methodology for selecting the right algorithm according to the problem demands.

Methods – experiment, testing, analysis, comparison.

Results – the methodology of selecting an appropriate machine learning algorithm depending on the problem parameters was described and applied in the context of this project. The result of applying this methodology is a classification algorithms' recommender of Pyspark.ML and sklearn libraries based on regression models that predict quality and performance metrics.

Application area – spheres, where machine learning approach can be useful, but isn't well-developed yet or costs a lot.